Universidad Miguel Hernández de Elche

Los genes *ANGULATA7* y *ORBICULATA1* participan en la biogénesis del cloroplasto y la organogénesis foliar en Arabidopsis

Tamara Muñoz Nortes

Elche, 2017

UNIVERSITAS
Miguel
Hernández

Los genes *ANGULATA7* y *ORBICULATA1* participan en la biogénesis del cloroplasto y la organogénesis foliar en Arabidopsis

Trabajo realizado por la Licenciada Tamara Muñoz Nortes, en la Unidad de Genética del Instituto de Bioingeniería de la Universidad Miguel Hernández de Elche, para optar al grado de Doctora.

Elche, 31 de mayo de 2017

UNIVERSITAS
Miguel
Hernández

JOSÉ LUIS MICOL MOLINA, Catedrático de Genética de la Universidad Miguel Hernández de Elche, y

HÉCTOR CANDELA ANTÓN, Profesor Contratado Doctor de Genética de la Universidad Miguel Hernández de Elche,

HACEMOS CONSTAR:

Que el presente trabajo ha sido realizado bajo nuestra dirección y recoge fielmente la labor desarrollada por la Licenciada Tamara Muñoz Nortes para optar al grado de Doctora. Las investigaciones reflejadas en esta Tesis se han desarrollado íntegramente en la Unidad de Genética del Instituto de Bioingeniería de la Universidad Miguel Hernández de Elche.

José Luis Micol Molina Héctor Candela Antón

Elche, 31 de mayo de 2017

UNIVERSITAS
Miguel
Hernández

Instituto de Bioingeniería
Universidad Miguel Hernández

Avenida de la Universidad s/n
03202 ELCHE (Alicante)
Telf: 96 591 8817 - Fax: 96 522 2033
e-mail: bioingenieria@umh.es

A quien corresponda:

Eugenio Vilanova Gisbert, Catedrático de Toxicología y Director del Instituto de Bioingeniería,

HACE CONSTAR:

Que da su conformidad a la lectura de la Tesis Doctoral presentada por Doña **Tamara Muñoz Nortes**, titulada **"Los genes *ANGULATA7* y *ORBICULATA1* participan en la biogénesis del cloroplasto y la organogénesis foliar en Arabidopsis"**, que se ha desarrollado dentro del Programa de Doctorado en Bioingeniería de este Instituto, bajo la dirección de los profesores Dr. José Luis Micol Molina y Dr. Héctor Candela Antón.

Lo que firmo en Elche, a instancias de la interesada y a los efectos oportunos, a treinta y uno de mayo de dos mil diecisiete.

Eugenio Vilanova Gisbert
Catedrático de Toxicología
Director del Instituto de Bioingeniería

UNIVERSITAS
Miguel
Hernández

A mis padres.

A Patricia.

UNIVERSITAS
Miguel
Hernández

"*La gota horada la roca, no por su fuerza, sino por su constancia.*"

Ovidio

"*In the end, everything will be okay. If it's not okay, it's not the end.*"

Fernando Sabino

"*No sabíamos que era imposible y lo hicimos a lo grande.*"

Etxániz

UNIVERSITAS
Miguel
Hernández

ÍNDICE DE MATERIAS

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

I.- PREFACIO

UNIVERSITAS
Miguel
Hernández

I.- PREFACIO

Este documento se ha elaborado siguiendo la normativa de la Universidad Miguel Hernández de Elche para la "Presentación de Tesis Doctorales con un conjunto de publicaciones" y se ha dividido en las partes siguientes:

1.- Un apartado de *Resumen y conclusiones*.

2.- Una *Introducción*, en la que se presenta el tema de la Tesis, el organismo experimental elegido y los antecedentes y objetivos del trabajo realizado.

3.- Una *Bibliografía de la introducción*.

4.- Un apartado de *Publicaciones,* que contiene las 3 siguientes (se indica en su caso el factor de impacto [FI]).

Muñoz-Nortes, T., Pérez-Pérez, J.M., Ponce, M.R., Candela, H., y Micol, J.L. (2017). The *ANGULATA7* gene encodes a DnaJ-like zinc-finger-domain protein involved in chloroplast function and leaf development in Arabidopsis. *Plant Journal* **89**, 870-884 (FI de 2015: 5,478).

Muñoz-Nortes, T., Pérez-Pérez, J.M., Sarmiento-Mañús, R., Candela, H., y Micol, J.L. (2017). Deficient glutamate biosynthesis triggers a concerted upregulation of ribosomal protein genes in Arabidopsis. *Scientific Reports*, en prensa (FI de 2015: 5,228).

Muñoz-Nortes, T., Candela, H., y Micol, J.L. Suitability of two distinct approaches for the high-throughput study of the post-embryonic effects of embryo-lethal mutations in Arabidopsis. Pendiente de aceptación.

5.- Un anexo que incorpora 12 comunicaciones a congresos: 6 nacionales y 6 internacionales.

Durante mi periodo predoctoral también he publicado dos artículos que no se incluyen en esta Tesis:

Muñoz-Nortes, T., Wilson-Sánchez, D., Candela, H., y Micol, J.L. (2014). Symmetry, asymmetry and the cell cycle in plants: known knowns and some known unknowns. *Journal of Experimental Botany* **65**, 2645-2655 (FI de 2014: 5,677).

Szakonyi, D., Van Landeghem, S., Bärenfaller, K., Baeyens, L., Blomme, J., Casanova-Sáez, R., De Bodt, S., Esteve-Bruna, D., Fiorani, F., Gonzalez, N., Grønlund, J., Immink, R.G.H., Jover-Gil, S., Kuwabara, A., Muñoz-Nortes, T., Van Dijk, A.-J., Wilson-Sánchez, D., Buchanan-Wollaston, V., Angenent, G.C., Van de Peer, Y., Inzé, D., Micol, J.L., Gruissem, W., Walsh, S., y Hilson, P. (2015). The KnownLeaf literature curation system captures knowledge about *Arabidopsis* leaf growth and development and facilitates integrated data mining. *Current Plant Biology* **2**, 1-11 (aún sin FI).

La Introducción de esta Tesis no incluye un apartado de Materiales y métodos, que están descritos en las publicaciones. Este documento incorpora varias bibliografías: las de cada uno de los tres artículos y la de la introducción. Todas las citas que se intercalan en la introducción de esta memoria se corresponden con referencias que aparecen en la bibliografía del mismo apartado; algunas de dichas citas se repiten en las bibliografías de uno o varios de los artículos.

Las tablas S2 de “The *ANGULATA7* gene encodes a DnaJ-like zinc-finger-domain protein involved in chloroplast function and leaf development in Arabidopsis” y S2 y S3 de “Deficient glutamate biosynthesis triggers a concerted upregulation of ribosomal protein genes” no se han incluido en esta memoria de Tesis por su gran tamaño. Las correspondientes hojas de cálculo se han remitido a los miembros del tribunal en formato electrónico.

II.- RESUMEN Y CONCLUSIONES

UNIVERSITAS
Miguel
Hernández

II.- RESUMEN Y CONCLUSIONES

Con el fin de identificar y caracterizar genes implicados en el desarrollo foliar de Arabidopsis, hemos estudiado en esta Tesis los mutantes *angulata7-1* (*anu7-1*), *orbiculata1-1* (*orb1-1*), *orb1-2* y *orb1-3*, obtenidos en el laboratorio de J.L. Micol mediante tratamiento de la estirpe silvestre Landsberg *erecta* (L*er*) con metanosulfonato de etilo.

Las hojas vegetativas de *anu7-1* son pálidas y dentadas. El tamaño de la roseta y la estatura de las plantas adultas de este mutante son menores que los de L*er*. Sus niveles de clorofilas y carotenoides son inferiores a los silvestres. El mesófilo en empalizada de *anu7-1* presenta células de tamaños irregulares y grandes espacios intercelulares. Las membranas tilacoidales de sus cloroplastos están parcialmente desapiladas.

Hemos identificado en el mutante *anu7-1*, mediante clonación posicional, una mutación puntual en At5g53860, un gen nuclear de copia única que codifica una proteína de los nucleoides de los cloroplastos. La proteína ANU7 presenta dos de los cuatro motivos C-X-X-C-X-G-X-G que caracterizan al dominio rico en cisteína (CR) de las proteínas DnaJ, que está conservado en la mayoría de las plantas terrestres, excepto en las gimnospermas y las gramíneas. Hemos confirmado que At5g53860 es *ANU7* mediante ensayos de alelismo entre *anu7-1* y dos líneas insercionales. También hemos comprobado que el transgén *$35S_{pro}$:At5g53860* restablece completamente el fenotipo silvestre en las plantas *anu7-1*. La visualización de la expresión de un transgén *$At5g53860_{pro}$:GUS* indica que *ANU7* se expresa en todos los tejidos y estados de desarrollo estudiados.

Hemos analizado el transcriptoma de *anu7-1* en micromatrices. Los genes nucleares sobreexpresados en *anu7-1* incluyen 7 de los que codifican componentes del complejo plastid transcriptionally active chromosome (pTAC) del cloroplasto. La sobreexpresión de estos genes puede ser una respuesta del núcleo a la señal retrógrada emitida por el cloroplasto como consecuencia de la disfunción de este orgánulo causada por la mutación *anu7-1*. En este mutante también se sobreexpresan algunos genes del genoma del cloroplasto transcritos por la polimerasa de ARN codificada por el núcleo (NEP, de nuclear-encoded RNA polymerase) y ninguno de los que dependen solamente de la polimerasa de ARN codificada por el cloroplasto (PEP, de plastid-encoded RNA polymerase), lo que sugiere que ANU7 participa en el ensamblaje de la maquinaria transcripcional del cloroplasto.

Hemos estudiado las interacciones genéticas entre *anu7-1* y alelos mutantes de *GENOMES UNCOUPLED 1* (*GUN1*), que está implicado en la señalización retrógrada del cloloroplasto al núcleo. El fenotipo morfológico del doble mutante *anu7-1 gun1-1* es

sinérgico: el margen de sus hojas es entero y en su mesófilo en empalizada se normaliza el tamaño de las células y no se observan espacios intercelulares. Las hojas *anu7-1 gun1-1* son variegadas, con sectores de diferentes grados de pigmentación, lo que sugiere que la ausencia de ANU7 y GUN1 potencia ciertas actividades de los cloroplastos, restableciéndose la morfología normal de la hoja pero alterándose localmente la función de este orgánulo y la distribución espacial de las clorofilas.

Las mutaciones *orb1* también reducen el tamaño de la roseta y las hojas vegetativas, que son pálidas, con niveles de clorofilas y carotenoides inferiores a los de L*er*. Las células del mesófilo en empalizada de los mutantes *orb1* son menores que las silvestres, pero el cociente entre su tamaño y el de la superficie de la hoja es similar al de L*er*. Hemos establecido que *ORB1* es At5g04140 combinando el análisis iterativo del ligamiento a marcadores moleculares y la secuenciación masiva. At5g04140 es un gen previamente descrito, que codifica una glutamato sintasa dependiente de ferredoxina, localizada en los cloroplastos: esta enzima sintetiza glutamato y 2-oxoglutarato a partir de glutamina y amonio. Hemos confirmado la identidad de *ORB1* mediante un ensayo de alelismo con un mutante insercional y hemos generado una construcción *At5g04140$_{pro}$:GUS* para visualizar el patrón de expresión de *ORB1*, que es generalizado.

Hemos secuenciado masivamente el ARN de *orb1-3*. Varios de los genes que hemos encontrado desregulados en las rosetas de este mutante están relacionados con el metabolismo del nitrógeno y la biosíntesis de aminoácidos, lo que sugiere que su expresión depende de los niveles de glutamato y/o glutamina. También se sobreexpresan de manera concertada muchos genes que codifican proteínas ribosómicas. Esta observación sugiere una respuesta compensatoria de la célula ante el déficit de glutamato, a fin de incrementar la síntesis de proteínas.

Hemos generado herramientas para el estudio de las funciones postembrionarias de genes letales embrionarios (*EMB*, de embryo defective) mediante análisis clonal. Hemos empleado las líneas CAUT (de cell autonomy) para obtener líneas transgénicas con el genotipo apropiado para el estudio de 24 genes *EMB*. Estas líneas pueden ser irradiadas con rayos X y examinadas posteriormente para encontrar sectores mutantes que manifiesten los efectos de la ausencia de la función de los genes a estudio en tejidos postembrionarios. También hemos adaptado el vector pCB1 a la tecnología Gateway, generando construcciones para el estudio de las funciones postembrionarias de 20 genes letales embrionarios mediante la escisión, mediada por la recombinasa Cre, de una copia *EMB* silvestre inserta en el genoma de plantas homocigóticas para los correspondientes alelos nulos *emb*.

III.- INTRODUCCIÓN

UNIVERSITAS
Miguel
Hernández

III.- INTRODUCCIÓN

III.1.- Arabidopsis, la planta modelo

La crucífera *Arabidopsis thaliana* (en adelante, Arabidopsis) se convirtió en el organismo de elección preferente para el estudio de la biología experimental de las plantas a finales de la penúltima década del siglo XX (Meyerowitz, 2001). Alcanzó así la condición de sistema modelo, denominación que suele darse a las especies que se estudian con el propósito de obtener conclusiones presuntamente extensibles a otras ramas del árbol filogenético (Bolker, 1995). Los motivos de la adopción de Arabidopsis como sistema modelo son los que la hacen fácilmente manejable en un laboratorio: es pequeña, autógama, prolífica y de ciclo de vida corto (Meyerowitz, 1987; Somerville y Koornneef, 2002). El reducido tamaño del genoma de Arabidopsis y la facilidad con que se transforma por infección con la bacteria *Agrobacterium tumefaciens* (Koornneef y Meinke, 2010) contribuyeron a su popularización como sistema modelo y a la rápida generación de un amplísimo inventario de herramientas genéticas. El uso de Arabidopsis ha contribuido muy notablemente a la comprensión de diversos aspectos de la biología vegetal, entre los que destacan los relacionados con la genética del desarrollo de las plantas. En efecto, Arabidopsis ha resultado especialmente útil para llegar a conclusiones acerca de la organogénesis de la raíz (Scheres y Wolkenfelt, 1998), los órganos florales (Coen y Meyerowitz, 1991) o la hoja (Eshed *et al.*, 2004; Ha *et al.*, 2007; Micol, 2009; Bilsborough *et al.*, 2011).

III.2.- Estructura y organogénesis foliar en Arabidopsis

III.2.1.- Importancia del estudio del desarrollo de las hojas de las plantas

Las hojas de las plantas fijan la energía solar, producen el oxígeno que respiramos y constituyen la fuente directa o indirecta de casi todos los alimentos que consumimos. El interés del estudio y la eventual manipulación del desarrollo de las hojas radica en que son el órgano fotosintético básico, en torno al cual gravita la vida en nuestro planeta (Micol, 2009).

Las hojas fueron consideradas cruciales para la taxonomía por Linneo (1751), y las unidades básicas de la arquitectura corporal de las plantas por Goethe (1816). Además, la hoja es el órgano vegetal de mayor relación superficie/volumen y por tanto el que más contribuye a la productividad primaria. En consecuencia, la comprensión de los mecanismos genéticos que regulan su desarrollo resulta fundamental, en la medida en que puede propiciar la mejora del rendimiento de los cultivos.

III.2.2.- Anatomía foliar en Arabidopsis

Las hojas de las plantas vasculares son órganos laterales con crecimiento determinado y estructura laminar. Arabidopsis presenta dos tipos de hojas, las vegetativas (en la roseta) y las caulinares (en los nudos basales de cada rama de la inflorescencia). Los ejes de polaridad de las hojas de Arabidopsis son tres (Figura 1A): el proximodistal (de la base al ápice), el mediolateral (de la vena primaria a los márgenes) y el dorsoventral (del haz al envés). El eje proximodistal de las hojas vegetativas define el peciolo y el limbo, el mediolateral establece dos mitades relativamente simétricas a ambos lados de la vena primaria, y el dorsoventral distingue dos superficies, una dorsal o adaxial, el haz, y otra ventral o abaxial, el envés (Byrne *et al.*, 2001).

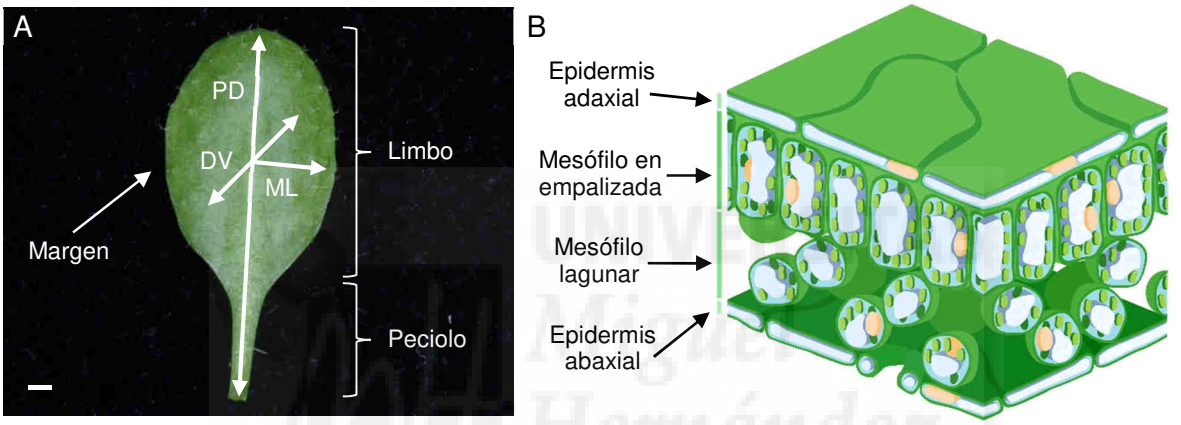


Figura 1.- Morfología e histología de la hoja vegetativa de Arabidopsis. (A) Morfología externa. Se indican las principales partes en que se divide la hoja, así como sus ejes. PD: eje proximodistal. ML: eje mediolateral. DV: eje dorsoventral. La barra de escala representa 1 mm. (B) Capas celulares en una hoja de Arabidopsis. Modificado a partir de https://commons.wikimedia.org/wiki/File: Leaf_tissue_structure_flat.svg?uselang=es

La histología de las hojas de Arabidopsis es simple (Figura 1B). La epidermis está formada por dos monocapas de células, una adaxial y otra abaxial, que aíslan a la planta de su entorno y regulan el intercambio gaseoso y de otras sustancias con el medio externo (Becraft, 1999). La epidermis está constituida mayoritariamente por células pavimentosas, que son aplanadas, interdigitadas, de forma irregular y sin capacidad de fotosíntesis. También forman parte de la epidermis algunas células especializadas, como las oclusivas de los estomas y los tricomas (Johnson, 1975; Zhao y Sack, 1999). En Arabidopsis, las dos monocapas epidérmicas foliares envuelven a los tejidos internos: unas cinco capas de mesófilo. El mesófilo en empalizada se yuxtapone a la epidermis adaxial, e incluye una o dos capas de células fotosintéticas alargadas en el sentido dorsoventral y densamente

empaquetadas. Entre el mesófilo en empalizada y la epidermis abaxial existen otras cuatro capas de mesófilo esponjoso o lagunar, compuesto por células más pequeñas, de formas muy irregulares y separadas por amplios espacios intercelulares, que facilitan la difusión de los gases. El sistema vascular de Arabidopsis, al igual que el de otras plantas, transporta sustancias a la vez que proporciona soporte mecánico a la hoja (Braybrook y Kuhlemeier, 2010). Las venas de Arabidopsis contienen dos tipos de tejidos conductores: el xilema, que transporta agua y sales minerales, y el floema, que conduce los productos de la fotosíntesis (Turner y Sieburth, 2002). Los haces vasculares atraviesan el mesófilo y cursan paralelos a la epidermis, ocupando el xilema su parte más adaxial y el floema, la más abaxial.

III.3.- El cloroplasto

El orgánulo más característico de una célula vegetal es el cloroplasto, que sustenta la vida en la Tierra, ya que convierte la energía solar en carbohidratos y oxígeno mediante la fotosíntesis. Los cloroplastos también juegan un papel vital en otros procesos fundamentales de la biología vegetal, como la asimilación del nitrógeno y la biosíntesis de aminoácidos, nucleótidos, ácidos grasos, hormonas vegetales, vitaminas y otros metabolitos (Daniell *et al.*, 2016).

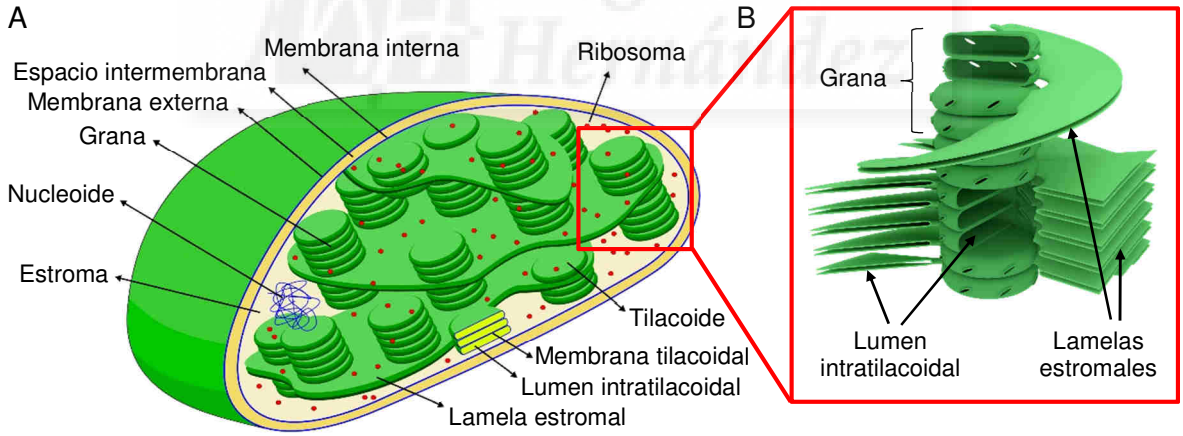


Figura 2.- Estructura del cloroplasto y distribución espacial de los tilacoides. (A) Esquema de un cloroplasto con indicación de sus principales componentes. (B) Detalle de la organización espacial de las membranas tilacoidales de un cloroplasto. Las imágenes se han modificado a partir de (A) https://commons.wikimedia.org/wiki/File:Scheme_Chloroplast-ca.svg y (B) https://commons.wikimedia.org/wiki/File:Helical_granum.png.

Cada célula del mesófilo de una hoja vegetativa de Arabidopsis contiene unos 100 cloroplastos lentiformes, de 5-10 µm de longitud y 3-4 µm de grosor (Figura 2A; Lopez-Juez y Pyke, 2004). La envuelta de estos cloroplastos es una doble membrana lipídica que

alberga un medio acuoso denominado estroma, que contiene los tilacoides. Estos últimos forman una compleja red tridimensional de membranas con un lumen intratilacoidal común y tienden a formar agregados de 10-20 capas denominados grana. Las membranas tilacoidales que interconectan los grana se denominan lamelas estromales (Figura 2B). Las membranas tilacoidales contienen los complejos proteicos que ejecutan la fotosíntesis, como la ATP sintasa, el citocromo b_6/f, los fotosistemas I y II (PSI y PSII) y los complejos antena asociados a los fotosistemas (LHCI y LHCII, de light-harvesting complex associated with photosystems I/II; Dekker y Boekema, 2005; Buchanan *et al.*, 2015).

III.3.1.- Origen y estructura del plastoma

Se cree que los cloroplastos se originaron hace unos 1.200-1.500 millones de años, como consecuencia de la fagocitosis de una cianobacteria fotosintética de vida libre por una célula eucariótica, tras la que establecieron una relación endosimbiótica (Dyall *et al.*, 2004). El endosimbionte fue perdiendo progresivamente autonomía como consecuencia de la transferencia de parte de sus genes al núcleo del hospedador. Los genes del endosimbionte adquirieron en el núcleo las secuencias reguladoras necesarias para su expresión, así como las que codifican el péptido de tránsito necesario para la incorporación de la correspondiente proteína al cloroplasto (Kleine *et al.*, 2009).

El cloroplasto es un orgánulo semiautónomo que contiene su propio material genético: una molécula de ADN bicatenario y circular, habitualmente denominada cromosoma del cloroplasto o plastoma (Sato *et al.*, 2003). El primer genoma cloroplástico completamente secuenciado fue el de *Nicotiana tabacum* (Shinozaki *et al.*, 1986). Se han depositado más de 800 genomas completos de cloroplastos de distintas especies vegetales en la base de datos de genomas de orgánulos del National Center for Biotechnology Information (NCBI). La disponibilidad de la secuencia del genoma del cloroplasto de especies de distintos clados ha llevado a la reconsideración de ciertas relaciones filogenéticas y ha contribuido a un conocimiento más profundo de la biodiversidad vegetal (Daniell *et al.*, 2016).

El genoma del cloroplasto de las plantas terrestres contiene dos repeticiones invertidas (IRa e IRb, de inverted repeats), separadas por dos regiones no repetidas, una de ellas grande (LSC, de large single copy), y la otra, pequeña (SSC, de small single copy). El genoma del cloroplasto de Arabidopsis tiene 0,15 Mb y contiene 110-120 genes, la mayoría de ellos relacionados con la fotosíntesis y la regulación de la expresión génica del propio plastoma (Rochaix, 1997; Sato *et al.*, 1999).

Cada cloroplasto contiene cientos de copias de su cromosoma, organizadas en agregados denominados nucleoides, por su analogía con los bacterianos (Melonek *et al.*, 2012). Los nucleoides de los cloroplastos se encuentran en las membranas tilacoidales y cada uno de ellos contiene 10-20 copias del cromosoma del cloroplasto, además de ARN y proteínas (Yagi y Shiina, 2014). Cada cloroplasto alberga unos 20 nucleoides, en los que tienen lugar procesos esenciales como la transcripción y el ensamblaje de los ribosomas (Qiao *et al.*, 2013).

III.3.2.- Maquinaria transcripcional del cloroplasto

La transcripción del genoma del cloroplasto en las plantas superiores es llevada a cabo por tres polimerasas de ARN, dos de ellas codificadas por genes nucleares y denominadas NEP (de nuclear-encoded RNA polymerase) y la tercera codificada por genes del cloroplasto y denominada PEP (de plastid-encoded RNA polymerase). En Arabidopsis son dos las NEP que transcriben el genoma del cloroplasto: RpoTp, exclusiva de los cloroplastos (Hricová *et al.*, 2006), y RpoTpm, presente tanto en las mitocondrias como en los cloroplastos (Baba *et al.*, 2004). Se desconoce si los helechos y las gimnospermas contienen genes que codifiquen alguna RpoTp. Sin embargo, se ha identificado una proteína RpoTp en la angiosperma basal *Nuphar advena*, lo que sugiere que esta polimerasa apareció con las plantas con flores. Por otro lado, ni *Nuphar advena* ni las monocotiledóneas poseen genes que codifiquen una RpoTmp, lo que sugiere que se trata de una adquisición posterior en la evolución de las dicotiledóneas (Yin *et al.*, 2010; Liere *et al.*, 2011).

Los genes del cloroplasto son transcritos por las enzimas NEP y PEP según la secuencia de sus promotores. Aunque muchos genes del cloroplasto contienen promotores que pueden ser reconocidos por las NEP y PEP, las primeras suelen transcribir genes domésticos o relacionados con la maquinaria transcripcional del cloroplasto, como *acetyl-coA carboxylase carboxyl transferase subunit β* (*accD*), *hypothetical conserved frame 2* (*ycf2*) o *RNA polymerase subunit A*, *B*, *C1* y *C2* (*rpoA*, *rpoB*, *rpoC1* y *rpoC2*), que codifican las subunidades α, β, β′ y β″ de la PEP, respectivamente. Esto implica que el correcto funcionamiento de la PEP requiere la acción previa de las NEP (Pfannschmidt *et al.*, 2015). La PEP, por su parte, transcribe preferentemente genes relacionados con la fotosíntesis, como los que codifican los componentes de los fotosistemas I y II (*photosystem I subunit y photosystem II subunit*; *psa* y *psb*, respectivamente), del citocromo b_6/f (*photosynthetic electron transfer*; *pet*) y la subunidad grande de la RuBisCO (*ribulose bisphosphate carboxylase large chain*; *rbcL*; Hajdukiewicz *et al.*, 1997; Hess y Börner,

1999; Qiao *et al.*, 2013; Börner *et al.*, 2015). Según las polimerasas que los transcriben, los genes del cloroplasto pueden dividirse en tres clases, compuestas por los que son transcritos preferentemente por la PEP (clase I), las NEP y PEP (II), y las NEP (III) (Pfalz *et al.*, 2006).

Las NEP son similares a las polimerasas de ARN de ciertos bacteriófagos, como el T3 y el T7, que son monoméricas (Hess y Börner, 1999). En Arabidopsis, las mutaciones en los genes que codifican la RpoTp y la RpoTpm alteran la morfogénesis foliar y retrasan la biogénesis de los cloroplastos y el crecimiento de la planta. Los dobles mutantes *rpotp rpotmp* son deficitarios en clorofila y su desarrollo se interrumpe poco después de la germinación (Baba *et al.*, 2004; Hricová *et al.*, 2006; Börner *et al.*, 2015).

La PEP, por su parte, es una polimerasa multimérica de tipo bacteriano con un núcleo catalítico formado por dos subunidades α y las subunidades β, β′ y β″. Este núcleo catalítico está asociado a proteínas codificadas por genes nucleares, como los factores sigma (σ) y las asociadas a la PEP (PAP, de PEP-associated protein; Steiner *et al.*, 2011; Börner *et al.*, 2015). Los factores sigma son necesarios para que la PEP reconozca los promotores de los genes que transcribe. Existen en Arabidopsis seis factores sigma distintos y se ha comprobado que SIGMA6 (SIG6) y SIG2 desempeñan una función importante en la transcripción de genes relacionados con la fotosíntesis (Qiao *et al.*, 2013; Börner *et al.*, 2015; Kindgren y Strand, 2015). Las PAP también juegan un papel fundamental en la transcripción mediada por la PEP y, por extensión, en la correcta expresión de los genes del cloroplasto (Qiao *et al.*, 2013). De hecho, las mutaciones que alteran genes que codifican algunas PAP suelen causar albinismo, interrupción del desarrollo de los cloroplastos, reducción de la actividad PEP e incremento en la actividad NEP, el mismo fenotipo que causan las mutaciones en los genes *rpo* (Santis-Maciossek *et al.*, 1999; Pfalz *et al.*, 2006; Myouga *et al.*, 2008; Gao *et al.*, 2011; Steiner *et al.*, 2011; Jeon *et al.*, 2012; Yagi *et al.*, 2012; Pfalz y Pfannschmidt, 2013; Yu *et al.*, 2013; Pfannschmidt *et al.*, 2015). Las mutaciones en los genes que codifican las PAP son letales en plántula cuando se cultivan en maceta: estos mutantes albinos pueden cultivarse en medio suplementado con sacarosa (Schröter *et al.*, 2010; Steiner *et al.*, 2011).

Se han realizado varios intentos de aislar y caracterizar la maquinaria transcripcional asociada a la PEP. Debe tenerse en cuenta que los nucleoides no son entidades celulares estables como los ribosomas, sino agregados de ADN, ARN y un conjunto variable de proteínas, lo que constituye una dificultad para su aislamiento. No obstante, en Arabidopsis y la mostaza blanca (*Sinapis alba*) se han conseguido separar dos fracciones de los nucleoides que contienen subunidades del núcleo catalítico de la

PEP (Pfalz *et al.*, 2006). Una de estas fracciones es soluble (sRNAP, de soluble RNA polymerase) y no contiene ADN, por lo que se le debe añadir para que muestre actividad transcripcional. La otra es insoluble (pTAC, de plastid transcriptionally active chromosome), está anclada a la membrana de los tilacoides y contiene ADN, por lo que es capaz de iniciar la transcripción por sí misma (Pfalz y Pfannschmidt, 2013). La fracción que contiene al complejo pTAC está formada por ADN, ARN y unas 50 proteínas, la mayoría de las cuales son productos de genes nucleares. Al menos 12 proteínas del complejo pTAC se consideran también PAP (Tabla 1; Wildwater *et al.*, 2005; Steiner *et al.*, 2011; Pfalz y Pfannschmidt, 2013; Yagi y Shiina, 2014; Yu *et al.*, 2014), mientras que las restantes no parecen estar tan estrechamente relacionadas con la maquinaria transcripcional. En efecto, el fenotipo causado por las mutaciones en los genes que codifican estas proteínas es menos severo que el de los mutantes afectados en PAP, llegando a ser indistinguibles del tipo silvestre en algunos casos (Pfalz *et al.*, 2006; Kindgren *et al.*, 2012; Pfalz y Pfannschmidt, 2013).

Tabla 1.- Proteínas asociadas a la PEP en Arabidopsis

Gen	Proteína	Dominio/Actividad	Referencia
AT3G04260	PAP1 (pTAC3)	Dominio SAP de unión a ADN	Yagi *et al.* (2012)
AT1G74850	PAP2 (pTAC2)	Repeticiones del motivo pentatricopéptido (PPR)	Pfalz *et al.* (2006)
AT3G48500	PAP3 (pTAC10)	Dominio S1 de unión a ARN	Chang *et al.* (2017)
AT5G23310	PAP4 (FSD3)	Actividad superóxido dismutasa	Myouga *et al.* (2008)
AT2G34640	PAP5 (pTAC12)	Dominio similar a RAD23	Pfalz *et al.* (2006)
AT3G54090	PAP6 (FLN1)	Carbohidrato quinasa de la familia PfkB	Arsova *et al.* (2010)
AT4G20130	PAP7 (pTAC14)	Dominio SET	Gao *et al.* (2011; 2012)
AT1G21600	PAP8 (pTAC6)	Dominio PHB	Pfalz *et al.* (2006)
AT5G51100	PAP9 (FSD2)	Actividad superóxido dismutasa	Myouga *et al.* (2008)
AT3G06730	PAP10 (TrxZ)	Dominio tiorredoxina, actividad disulfuro reductasa	Arsova *et al.* (2010)
AT1G63680	PAP11 (AtMurE)	Ligasa de la familia Mur	Garcia *et al.* (2008)
AT5G24314	PAP12 (pTAC7)	Desconocida	Yu *et al.* (2013)

Se clasifica a las PAP en tres grupos, según sus funciones (Steiner *et al.*, 2011; Yu *et al.*, 2014; Chang *et al.*, 2017). Las proteínas del primer grupo contienen dominios de unión a ácidos nucleicos, como PAP1, PAP2, PAP3, PAP5, PAP7 y PAP12 (también llamadas pTAC3, pTAC2, pTAC10, pTAC12, pTAC14 y pTAC7, respectivamente), y están probablemente implicadas en el reconocimiento de promotores, así como en la iniciación y

elongación de la transcripción mediada por PEP (Kindgren y Strand, 2015; Pfannschmidt *et al.*, 2015). Las proteínas del segundo grupo, como PAP10 (también llamada THIOREDOXIN Z; TrxZ) y PAP6 (también llamada FRUCTOKINASE-LIKE 1; FLN1) regulan la actividad transcripcional de la PEP dependiendo del estado redox del cloroplasto (Gilkerson *et al.*, 2012; Yu *et al.*, 2014; Pfannschmidt *et al.*, 2015). Las proteínas del tercer grupo, como PAP4 (también llamada FE SUPEROXIDE DISMUTASE 3; FSD3) y PAP9 (FSD2), protegen al complejo PEP de las especies reactivas de oxígeno (Myouga *et al.*, 2008). Se desconoce la función de PAP8 (también llamada pTAC6) y PAP11 (también llamada AtMurE; Yu *et al.*, 2014; Chang *et al.*, 2017).

III.3.3.- Comunicación entre el núcleo y el cloroplasto

A pesar del reducido número de genes del genoma del cloroplasto, su proteoma incluye unas 3.000 proteínas, en su mayoría codificadas por genes nucleares. Esta es la principal consecuencia de la migración de los genes del endosimbionte ancestral al núcleo de su hospedador (Sato *et al.*, 1999; Jung y Chory, 2010).

Mediante la denominada señalización o regulación anterógrada, el núcleo controla la mayoría de los procesos que tienen lugar en los cloroplastos, incluida la expresión de los genes del plastoma (Singh *et al.*, 2015). La transcripción del plastoma es un claro ejemplo de regulación anterógrada, pues las NEP están codificadas por el genoma nuclear y transcriben los genes del cloroplasto que codifican las subunidades del núcleo catalítico de la PEP (Pfannschmidt *et al.*, 2015). Las otras proteínas que constituyen el complejo pTAC y que participan activamente en la transcripción mediada por la PEP, como los factores sigma y las PAP, también son productos de genes nucleares. Del mismo modo, otros complejos multiméricos presentes en los cloroplastos están compuestos por subunidades codificadas tanto por el genoma nuclear como por el del orgánulo. Son ejemplos de ello los complejos TOC (TRANSLOCON OF THE OUTER CHLOROPLAST ENVELOPE), TIC (TRANSLOCON OF THE INNER CHLOROPLAST ENVELOPE; Reumann y Keegstra, 1999; Dyall *et al.*, 2004), los fotosistemas, el citocromo b_6/f, la ATP sintasa y los ribosomas del cloroplasto (Leister, 2003). Parece pues necesaria la coordinación entre la síntesis de las proteínas codificadas por los dos genomas, con el fin de mantener una estequiometría adecuada de las subunidades de dichos complejos proteicos, que asegure su correcto funcionamiento (Jung y Chory, 2010).

Los cloroplastos también pueden enviar señales al núcleo en un proceso denominado señalización retrógrada, observado por primera vez hace cuatro décadas (Bradbeer *et al.*, 1979; Koussevitzky *et al.*, 2007). La señalización retrógrada regula la

expresión de genes nucleares en respuesta al estado metabólico y de desarrollo del cloroplasto (Singh *et al*., 2015). En las plantas cuyos cloroplastos presentan alteraciones en su biogénesis o el metabolismo fotosintético se reprime la expresión de genes nucleares relacionados con la fotosíntesis, como los que codifican las subunidades del fotosistema II (LIGHT HARVESTING COMPLEX PHOTOSYSTEM II, Lhcb) y la RuBisCO (Susek *et al*., 1993). La señalización retrógrada también controla la expresión y activación de enzimas clave para el ciclo celular, como las ciclinas dependientes de quinasas, y es un factor determinante para la transición entre las fases de proliferación y expansión celular en el mesófilo foliar (Andriankaja *et al*., 2012). La señalización retrógrada también se ha asociado al control de la especificación de los dominios adaxial y abaxial de la hoja (Tameshige *et al*., 2013).

Las moléculas candidatas a mediar la señalización retrógrada incluyen algunos intermediarios de la ruta de los tetrapirroles (como la Mg-protoporfirina IX), especies reactivas del oxígeno (como el peróxido de hidrógeno y el oxígeno singlete), hormonas (como el ácido abscísico) y otros metabolitos (como la 3′-fosfoadenosina 5′-fosfato y el β-ciclocitral) (Leister, 2012; Barajas-López *et al*., 2013; Terry y Smith, 2013; Chan *et al*., 2016).

En la primera búsqueda realizada para identificar genes implicados en la señalización retrógrada se aislaron los mutantes *genomes uncoupled* (*gun*; Susek *et al*., 1993). Los productos de los genes *GUN2*, *GUN3*, *GUN4* y *GUN5* participan en la ruta de biosíntesis de los tetrapirroles, modulando los niveles de Mg-protoporfirina IX (Vinti *et al*., 2000; Mochizuki *et al*., 2001; Larkin *et al*., 2003), cuya acumulación reprime en Arabidopsis la expresión de cientos de genes nucleares que codifican proteínas del cloroplasto relacionadas con la fotosíntesis (Strand *et al*., 2003). GUN1 es una proteína con repeticiones del motivo pentatricopéptido (PPR), que se localiza en el nucleoide del cloroplasto junto a otras pTAC como pTAC5 y pTAC2, asemejándose a esta última (Huang *et al*., 2013; Tadini *et al*., 2016). GUN1 integra señales asociadas a las perturbaciones en la expresión del plastoma, en la síntesis de tetrapirroles y en el estado redox del cloroplasto, y emite una señal retrógrada que reprime la expresión de ciertos genes nucleares que controlan las funciones del orgánulo (Koussevitzky *et al*., 2007; Tadini *et al*., 2016).

El mecanismo molecular de la transmisión al núcleo de las señales generadas en el cloroplasto no se comprende completamente. En efecto, se han identificado muy pocas moléculas citoplásmicas y nucleares responsables de la modulación de la expresión génica en respuesta a dichas señales. Una de ellas es el factor de transcripción ABA INSENSITIVE 4 (ABI4), que reconoce la secuencia reguladora CCAC, que está muy representada en los

promotores de genes nucleares cuya expresión depende de la señalización retrógrada (Koussevitzky *et al.*, 2007). Se ha propuesto también al factor de transcripción PHD-TYPE TRANSCRIPTION FACTOR WITH TRANSMEMBRANE DOMAINS (PTM) como mediador entre las señales generadas en el cloroplasto y la expresión génica en el núcleo. La forma inmadura de PTM se localiza en la membrana externa de los cloroplastos, y su dominio aminoterminal puede ser liberado por proteolisis en respuesta a señales del cloroplasto y viajar al núcleo, en donde se acumula y activa la transcripción de *ABI4* (Sun *et al.*, 2011). Los factores de transcripción ELONGATED HYPOCOTYL 5 (HY5) y GOLDEN 2-LIKE 1 (GLK1) y GLK 2 también responden a señales procedentes de los cloroplastos, y por ello se ha propuesto que son mediadores de la señalización retrógrada (Ruckle y Larkin, 2009; Waters *et al.*, 2009).

III.4.- Las proteínas DnaJ

Los genes de *Escherichia coli* que codifican las proteínas DnaJ y DnaK forman un operón (Saito y Uchida, 1978). La proteína de choque térmico de 70 kDa (Hsp70, de Heat shock protein 70) eucariótica es homóloga de DnaK (Bardwell y Craig, 1984), y las Hsp40 de *Saccharomyces cerevisiae*, *Drosophila melanogaster*, las plantas y humana son homólogas de DnaJ (Caplan *et al.*, 1993; Kurzik-Dumke *et al.*, 1995; Venter *et al.*, 2001; Fan *et al.*, 2003; Walsh *et al.*, 2004; Rajan y D'Silva, 2009). El sistema Hsp40-Hsp70 eucariótico es equivalente al DnaJ-DnaK de *Escherichia coli*.

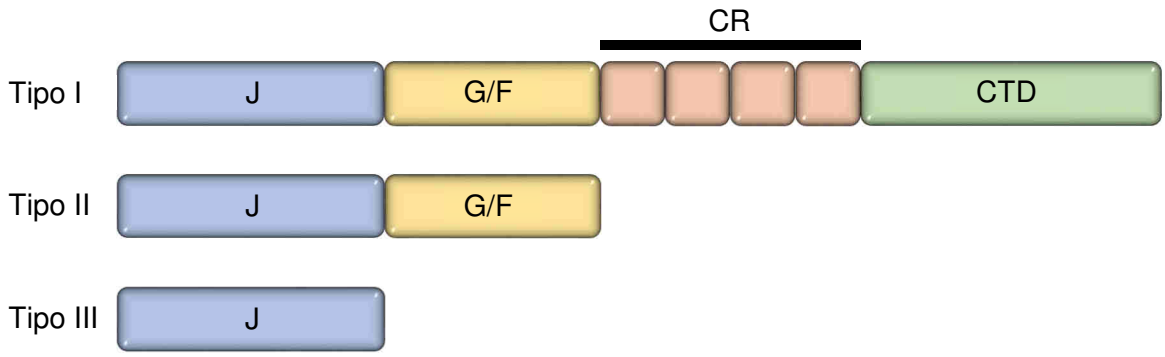


Figura 3.- Clasificación de las proteínas DnaJ según la presencia o ausencia de sus dominios característicos. Se denomina a estos dominios J, G/F (rico en glicina y fenilalanina), CR (rico en cisteína, que contiene cuatro repeticiones del motivo C-X-X-C-X-G-X-G) y CTD (carboxiterminal). Modificado a partir de Cheetham y Caplan (1998).

La proteína DnaJ de *Escherichia coli* contiene cuatro dominios característicos (Figura 3): el J en su extremo amino, formado por cuatro hélices α y que contiene un

tripéptido de histidina, prolina y ácido aspártico (motivo HPD), esencial para la función de las proteínas DnaJ, entre la segunda y tercera hélice (Qian *et al.*, 1996; Miernyk, 2001); el G/F, rico en glicina y fenilalanina; el CR, central, rico en cisteína y con cuatro repeticiones del motivo C-X-X-C-X-G-X-G (X: cualquier aminoácido) que forman dos dedos de zinc (Szabo *et al.*, 1996; Cheetham y Caplan, 1998), y el CTD, carboxiterminal y poco conservado. La proteína DnaJ de *Escherichia coli* presenta actividad disulfuro isomerasa (de Crouy-Chanel *et al.*, 1995), que se ha relacionado con los motivos C-X-X-C del dominio CR (Chivers *et al.*, 1997). Las proteínas similares a la DnaJ de *Escherichia coli* se clasifican en tres tipos, (I) las que cuentan con los cuatro dominios antes mencionados, (II) las que contienen los dominios J y G/F y (III) las que solo conservan el dominio J (Cheetham y Caplan, 1998).

Las proteínas DnaJ/Hsp40 se han conservado a lo largo de la evolución como consecuencia de su importancia en la homeostasis de las proteínas, ya que participan en procesos cruciales como su síntesis, plegamiento, translocación y degradación. Suelen actuar como cochaperonas de las proteínas DnaK/Hsp70, estabilizando su unión a ciertos polipéptidos recién sintetizados y contribuyendo así al correcto plegamiento de estos últimos (Yochem *et al.*, 1978; Frydman *et al.*, 1994; Hartl y Hayer-Hartl, 2002; Hennessy *et al.*, 2005; Qiu *et al.*, 2006). También tienen actividad por sí mismas, uniéndose a proteínas nacientes para prevenir su agregación (Langer *et al.*, 1992; Hendrick *et al.*, 1993; Cyr, 1995; Cheetham y Caplan, 1998; Meacham *et al.*, 1999).

Algunas plantas como el arroz o Arabidopsis cuentan con más de cien proteínas DnaJ (Rajan y D'Silva, 2009; Sarkar *et al.*, 2013). Muchas de ellas ejercen sus funciones en el cloroplasto y están relacionadas con la biogénesis y/o la función de este orgánulo (Chiu *et al.*, 2013), como las que se muestran en la Tabla 2, en la página 16; algunas de ellas presentan actividad disulfuro isomerasa, al igual que la proteína DnaJ de *Escherichia coli* (Shimada *et al.*, 2007; Lu *et al.*, 2011; Zhong *et al.*, 2013; Fristedt *et al.*, 2014).

III.5.- Relación entre el metabolismo del nitrógeno y la biosíntesis de aminoácidos en las plantas

III.5.1.- Implicación del ciclo GS/GOGAT en la fijación y reasimilación del nitrógeno

El nitrógeno es un nutriente básico para el crecimiento y el desarrollo de las plantas, ya que forma parte de moléculas esenciales como los ácidos nucleicos y los aminoácidos. La eficacia en el uso del nitrógeno repercute en la producción de biomasa vegetal y por consiguiente en la productividad primaria de los ecosistemas (Gaufichon *et al.*, 2016).

Tabla 2.- Proteínas DnaJ que actúan en el cloroplasto

Proteína	Especie*	Proceso en el que participa
ACCUMULATION AND REPLICATION OF CHLOROPLASTS (ARC6)[a,b]	A	División del cloroplasto
Bundle sheath defective2 (Bsd2)[c]	M	Regulación postraduccional de la subunidad grande de la RuBisCO
CELL GROWTH DEFECT FACTOR 1 (CDF1)/ CHAPERONE-LIKE PROTEIN OF POR1 (CPP1)[d]	A, Nb	Biosíntesis de clorofilas y desarrollo del cloroplasto
CHLOROPLAST J-LIKE DOMAIN 1 (CJD1)[e]	A	Regulación de la composición en ácidos grasos del cloroplasto
CHLORORESPIRATORY REDUCTION J (CRRJ)[f]	A	Flujo de electrones a través del fotosistema I
CHLORORESPIRATORY REDUCTION L (CRRL)[f]	A	Flujo de electrones a través del fotosistema I
EMBRYO SAC DEVELOPMENT ARREST 3 (EDA3)/ PSA2[g,h]	A, M	Ensamblaje y acumulación del fotosistema I
J8/ TOC12[i-k]	A, G	Estabilización del fotosistema II, regulación del metabolismo del cloroplasto en la oscuridad
J11[i]	A	Estabilización del fotosistema II
J20[i,l]	A	Estabilización del fotosistema II, biosíntesis de isoprenoides
Lycopersicon esculentum chloroplast DnaJ protein 1 (LeCDJ1)[m,n]	T	Mantenimiento del fotosistema II en condiciones de estrés abiótico
LeCDJ2[o]	T	Mantenimiento del fotosistema II en condiciones de estrés biótico y abiótico
LOW QUANTUM YIELD OF PHOTOSYSTEM II1 (LQY1)[p]	A	Mantenimiento del fotosistema II
Oryza sativa DnaJ 7/8 (OsDjA7/8)[q]	Ar	Desarrollo y diferenciación del cloroplasto
Pea chloroplast DnaJ homologue (PCJ1)[r]	G	Transporte de proteínas a través de las envueltas del cloroplasto
pTAC5[s]	A	Desarrollo y diferenciación del cloroplasto, mantenimiento de la actividad PEP en condiciones de estrés
SNOWY COTYLEDON 2 (SCO2)/SHI-YO-U MEANS COTYLEDON IN JAPANESE (CYO1)[t-y]	A, Ar, Lj	Ensamblaje y mantenimiento de los fotosistemas, transporte vesicular de proteínas fotosintéticas
SlCDJ2[z]	T	Mantenimiendo de la actividad RuBisCO en condiciones de estrés abiótico

*A, M, G, T, Ar, Nb y Lj: Arabidopsis, maíz, guisante, tomate, arroz, *Nicotiana benthamiana* y *Lotus japonicus*, respectivamente. [a]Vitha *et al.* (2003); [b]Glynn *et al.* (2008); [c]Brutnell *et al.* (1999); [d]Lee *et al.* (2013); [e]Ajjawi *et al.* (2011); [f]Yamamoto *et al.* (2011); [g]Fristedt *et al.* (2014); [h]Wang *et al.* (2016); [i]Chen *et al.* (2010); [j]Chiu *et al.* (2010); [k]Chen *et al.* (2011); [l]Pulido *et al.* (2013); [m]Kong *et al.* (2014a); [n]Kong *et al.* (2014b); [o]Wang *et al.* (2014); [p]Lu *et al.* (2011); [q]Zhu *et al.* (2015); [r]Schlicher y Soll (1997); [s]Zhong *et al.* (2013); [t]Shimada *et al.* (2007); [u]Albrecht *et al.* (2008); [v]Muranaka *et al.* (2012); [w]Tanz *et al.* (2012): [x]Tominaga *et al.* (2016); [y]Zagari *et al.* (2017), y [z]Wang *et al.* (2015).

Las plantas absorben el nitrógeno del suelo en forma de amonio y nitrato; este último es reducido a amonio por las nitrato y nitrito reductasas. El amonio es fijado gracias a la acción de la glutamina sintetasa (GS), que lo utiliza como sustrato para formar glutamina a partir de glutamato, consumiendo adenosín trifosfato (ATP). La glutamina es a su vez sustrato de la glutamato sintasa (GOGAT) para la producción de 2-oxoglutarato y dos moléculas de glutamato, completándose así el llamado ciclo GS/GOGAT. La glutamina y el glutamato sintetizados en el ciclo GS/GOGAT son los principales donadores de nitrógeno en la biosíntesis de compuestos como los aminoácidos, los ácidos nucleicos o la clorofila. El 2-oxoglutarato liberado en la síntesis de glutamato es utilizado en el ciclo de los ácidos tricarboxílicos (TCA; Miflin y Lea, 1976).

Existen dos tipos de GS en Arabidopsis: la GS1, localizada en el citoplasma y codificada por cinco genes distintos (*GLN1;1* a *GLN1;5*), y la GS2, codificada por un solo gen (*GLN2*) y localizada en los cloroplastos y las mitocondrias (McNally y Hirel, 1983; Peterman y Goodman, 1991; Taira *et al.*, 2004). La GS2 se expresa predominantemente en las células del mesófilo foliar (Thum *et al.*, 2003), mientras que las GS1 citoplásmicas lo hacen principalmente en los tejidos no fotosintéticos y son importantes para la asimilación primaria del amonio en las raíces y para la reasimilación del amonio derivado de la hidrólisis de proteínas (Thomsen *et al.*, 2014). Existen dos tipos de glutamato sintasas, que se diferencian en la molécula que utilizan como donador de electrones. La glutamato sintasa dependiente de ferredoxina (Fd-GOGAT) es codificada en Arabidopsis por dos genes, *GLU1* y *GLU2* (Coschigano *et al.*, 1998), mientras que la dependiente de NADH (NADH-GOGAT) es codificada por un único gen, *GLT*. Las Fd-GOGAT son responsables del 96% de la actividad glutamato sintasa en las hojas y del 68% en las raíces de Arabidopsis (Somerville y Ogren, 1980; Suzuki y Rothstein, 1997). La identidad entre las secuencias de aminoácidos de GLU1 y GLU2 es de un 80% (Suzuki y Knaff, 2005), pero sus patrones de expresión son contrarios: mientras que la GLU1 se expresa a altos niveles en las hojas y es responsable de la mayor parte de la actividad Fd-GOGAT en los tejidos fotosintéticos, GLU2 se expresa sobre todo en tejidos no fotosintéticos como las raíces, al igual que NADH-GOGAT (Lam *et al.*, 1995; Coschigano *et al.*, 1998; Chen *et al.*, 2016). Las glutamato sintasas se localizan principalmente en los plastos de las raíces y los cloroplastos de las hojas, aunque la presencia de la GLU1 se ha demostrado también en las mitocondrias (Jamai *et al.*, 2009; Chen *et al.*, 2016). La semejanza en los patrones de expresión de GLU1 y GLN2 en respuesta a la luz y a ciertos metabolitos sugiere que estas dos enzimas constituyen el sistema GS/GOGAT en las hojas (Coschigano *et al.*, 1998).

Otro proceso clave del metabolismo del nitrógeno en las plantas es la reasimilación del amonio liberado en la fotorrespiración. Mediante el análisis de los efectos de las mutaciones en el gen *GLU1*, se ha demostrado que la Fd-GOGAT, y en consecuencia el ciclo GS/GOGAT, juega también un papel fundamental en la reincorporación del amonio procedente de la fotorrespiración a las rutas de biosíntesis de aminoácidos. Algunos mutantes *glu1*, cuya actividad Fd-GOGAT es un 5% de la del tipo silvestre, sufren clorosis y acaban muriendo en condiciones atmosféricas, pero son capaces de crecer en presencia de altas concentraciones de CO_2, en las que se suprime la fotorrespiración en el mesófilo. El hecho de que estos mutantes crezcan en dichas condiciones indica que siguen llevando a cabo la asimilación primaria de nitrógeno, o por la actividad residual de GLU1 o por la acción de GLU2 y GLT. Esta observación también sugiere un papel esencial de GLU1 en la reasimilación del amonio liberado en la fotorrespiración por el ciclo GS/GOGAT (Somerville y Ogren, 1980; Suzuki y Rothstein, 1997; Coschigano *et al.*, 1998).

III.5.2.- Interconexión entre las rutas biosintéticas de los aminoácidos

Los aminoácidos son las unidades estructurales básicas de las proteínas, y sus rutas biosintéticas están estrechamente interconectadas. La glutamina y el glutamato que produce el ciclo GS/GOGAT participan en muchas reacciones enzimáticas como donadores de grupos amida y amino, respectivamente, para rendir otros aminoácidos como la asparagina, el aspartato, la arginina o la glicina (Figura 4; Ferrario-Méry *et al.*, 2000).

Las asparagina sintetasas (AS) catalizan la transferencia del grupo amida de la glutamina al aspartato, sintetizándose asparagina y glutamato. En Arabidopsis, las AS están codificadas por tres genes: *ASPARAGINE SYNTHETASE 1* (*ASN1*), *ASN2* y *ASN3* (Lea *et al.*, 2007; Gaufichon *et al.*, 2010). La carbamoil fosfato sintetasa (CPS), cuyas subunidades son productos de los genes *VENOSA3* (*VEN3*) y *VEN6*, utilizan glutamina y bicarbonato para rendir glutamato y carbamoil fosfato; este último es necesario para la conversión de ornitina en citrulina, un aminoácido que puede dar lugar a arginina tras la acción secuencial de las enzimas argininosuccinato sintetasa (ASS) y argininosuccinato liasa (ASL; Taira *et al.*, 2004; Mollá-Morales *et al.*, 2011). La transferencia del grupo amino del glutamato al oxalacetato, que es llevada a cabo por la aspartato aminotransferasa (AspAT), rinde 2-oxoglutarato y aspartato, que puede contribuir también a la biosíntesis de arginina (Gaufichon *et al.*, 2016). Las glutamato deshidrogenasas (GDH) catalizan reversiblemente la desaminación oxidativa del glutamato a 2-oxoglutarato (Buchanan *et al.*, 2015). En Arabidopsis, *GLUTAMATE DEHYDROGENASE 1* (*GDH1*), *GDH2* y *GDH3*

codifican las tres subunidades de las GDH (β, α y γ, respectivamente), que se combinan formando homo o heterohexámeros (Miyashita y Good, 2008; Marchi *et al.*, 2014).

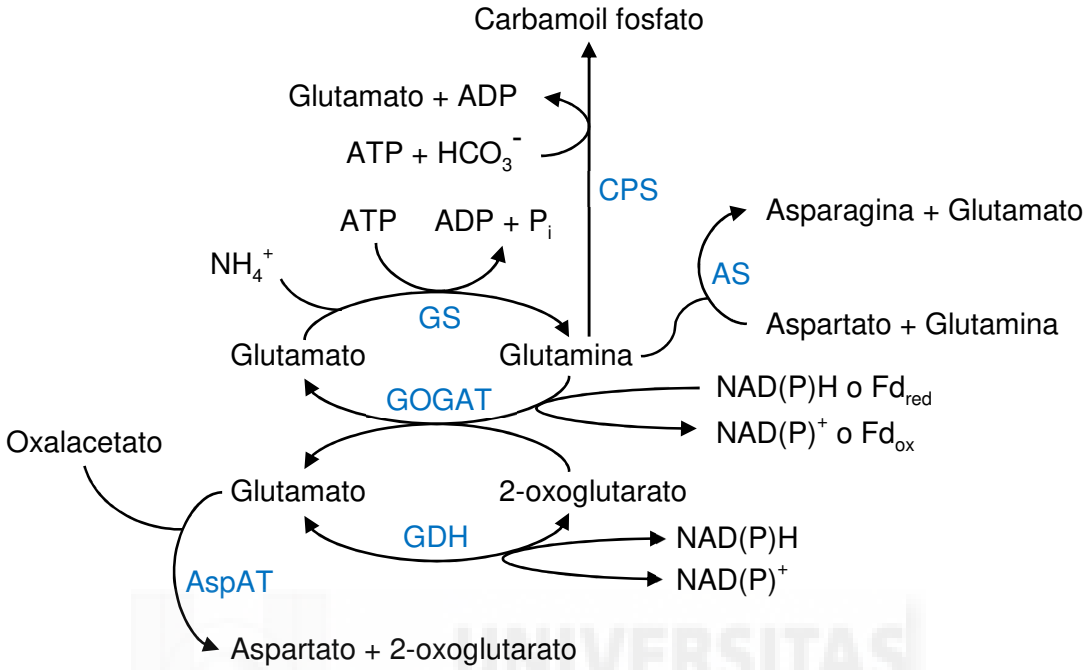


Figura 4.- Participación de la glutamina y el glutamato en el metabolismo de los aminoácidos. Se representan las principales etapas de las rutas biosintéticas de aminoácidos en las que participan la glutamina y el glutamato, y se indican en azul las enzimas que las catalizan. CPS: carbamoil fostato sintetasa. AS: asparagina sintetasa. GS: glutamina sintetasa. GOGAT: glutamina oxoglutarato aminotransferasa/glutamato sintasa. GDH: glutamato deshidrogenasa. AspAT: aspartato aminotransferasa. Modificado a partir de Buchanan *et al.* (2015).

Una consecuencia de la estrecha interconexión entre las rutas de biosíntesis de los aminoácidos es que la perturbación de una de ellas no solo altera los niveles de su producto final, sino también los de otros, perturbándose la homeostasis celular (Miesak y Coruzzi, 2002; Voll *et al.*, 2003). Además, como los aminoácidos constituyen las unidades estructurales básicas de las proteínas, la carencia de uno o varios de ellos puede provocar una reducción del contenido total de proteínas del organismo (Voll *et al.*, 2003), así como una disminución en la producción de biomasa. Las mutaciones de pérdida de función en el gen *ASN2* de Arabidopsis, por ejemplo, causan defectos en el crecimiento y una menor biomasa foliar (Gaufichon *et al.*, 2013).

III.6.- Los mutantes como herramientas para el estudio de los genes

Las estrategias para la disección de las funciones de los genes se han basado tradicionalmente en la inducción, aislamiento y caracterización de mutantes (Muller y

Altenburg, 1930). Existen dos abordajes experimentales para este fin. Por un lado, la genética clásica o directa se basa en la realización de una mutagénesis al azar, tras la cual se aíslan mutantes que manifiestan alteraciones en un determinado proceso y en los que se desconoce qué gen ha sido dañado por el mutágeno. El estudio posterior de esos mutantes y la cartografía de sus mutaciones conduce finalmente a la identificación del gen causante del fenotipo de interés (Jander *et al.*, 2002; Ostergaard y Yanofsky, 2004). La clonación posicional mediante análisis de ligamiento ha sido durante décadas una estrategia muy útil para identificar mutaciones en genes cuyas funciones son necesarias para el desarrollo foliar de Arabidopsis (Koornneef *et al.*, 1983; Bell y Ecker, 1994; Robles y Micol, 2001). En la primera década del siglo XXI, la aparición de los secuenciadores masivamente paralelos supuso una revolución en el ámbito de la genética y la genómica, reduciéndose drásticamente el gasto y el esfuerzo asociado a la secuenciación de genomas completos (Lister *et al.*, 2009; Metzker, 2010). Se denomina cartografía mediante secuenciación (mapping-by-sequencing) al método utilizado para determinar la posición de una mutación, usualmente puntual, que causa un fenotipo de interés, mediante la secuenciación masiva del ADN de individuos de una población cartográfica (James *et al.*, 2013; Candela *et al.*, 2015). Una de las muchas variantes de este abordaje es la combinación de la cartografía génica mediante análisis del ligamiento a marcadores moleculares y la secuenciación completa del genoma a estudio en un secuenciador masivamente paralelo (Mateo-Bonmatí *et al.*, 2014; Garcia *et al.*, 2016).

Se llama genética inversa a aquella en la que se modifica la secuencia de un gen, conocida *a priori*, para estudiar sus consecuencias fenotípicas. La variante más moderna de este abordaje se basa en la introducción de mutaciones mediante el sistema CRISPR (Sander y Joung, 2014). La disponibilidad de las secuencias completas de los genomas de las especies modelo ha facilitado mucho la genética inversa (Lloyd y Meinke, 2012).

Los abordajes mutacionales han sido y son una poderosa herramienta para la identificación de los genes responsables de diferentes facetas de la biología vegetal (Page y Grossniklaus, 2002). Una colección de mutantes constituye un material de partida fundamental para la comprensión de cualquier proceso biológico. La caracterización de mutantes foliares, con alteraciones en la morfología de la hoja, ha contribuido sustancialmente a la disección genética de diferentes aspectos del desarrollo de este órgano: la iniciación del primordio foliar, el establecimiento de los ejes de polaridad, la configuración del margen y la contribución que la proliferación y la expansión celular hacen a la arquitectura final del órgano (Berná *et al.*, 1999; Micol y Hake, 2003; Fleming, 2005; Horiguchi *et al.*, 2006; Micol, 2009; Pérez-Pérez *et al.*, 2009).

III.7.- El estudio del desarrollo mediante análisis clonal

Se llama mosaicos genéticos a los individuos que contienen grupos de células vecinas genotípicamente iguales entre sí pero distintas a las de su contexto. A dichos grupos de células se les denomina sectores o clones. El mosaicismo genético se suele asociar a la aparición de mutaciones, que pueden afectar a células de la línea germinal (mosaicismo gonadal) o somáticas (mosaicismo somático). En el mosaicismo gonadal, la mutación puede transmitirse a la descendencia, mientras que en el somático lo hace clonalmente y afecta solamente a las células derivadas de la que sufrió la mutación.

El análisis clonal es el estudio del tamaño, la distribución espacial y la correspondencia entre el fenotipo y el genotipo de los sectores de mosaicos genéticos inducidos. Las técnicas de análisis clonal se basan en la inducción de sectores de células mutantes relacionadas clonalmente, en un individuo que es por lo demás fenotípicamente normal. El uso de mosaicos genéticos inducidos para el estudio del desarrollo se inició en *Drosophila melanogaster* a principios del siglo XX (Sturtevant, 1932), impulsado por el descubrimiento de la recombinación mitótica entre cromosomas homólogos (Stern, 1936). La inducción de recombinación mitótica mediante irradiación con rayos X (Patterson, 1930) resultó eficaz para la generación de mosaicos somáticos en *Drosophila melanogaster*. El análisis de mosaicos genéticos en este sistema modelo ha resultado muy útil para averiguar las funciones de muchos genes durante el desarrollo (Sturtevant, 1932) y para el estudio de los linajes celulares (Garcia-Bellido y Merriam, 1969), la autonomía celular (Morgan *et al*., 1919; Sturtevant, 1920), los lugares de síntesis y actuación de moléculas difusibles (Rubin, 1989; Heitzler y Simpson, 1991) y los efectos de las mutaciones letales embrionarias (Bryant y Zornetzer, 1973), aspectos de gran importancia en la biología de cualquier organismo pluricelular.

El análisis clonal también contribuyó a la comprensión del desarrollo vegetal en el primer tercio del siglo XX (Stadler, 1930). Mucho después, la inducción de sectores que manifestaban los efectos fenotípicos de marcadores visibles resultó crucial para comprender el papel de los linajes celulares en el desarrollo normal de plantas como Arabidopsis, el maíz y el tabaco (Poethig, 1987; Furner *et al*., 1996; Hernandez *et al*., 1999), así como el grado de autonomía celular de los productos génicos de *liguleless1* (*lg1*) y *liguleless2* (*lg2*) del maíz y *FCA* de Arabidopsis (Becraft *et al*., 1990; Furner *et al*., 1996; Harper y Freeling, 1996). Además, las técnicas de análisis clonal han proporcionado información fundamental sobre las funciones postembrionarias de genes como *EMBRYO DEFECTIVE 506* (*EMB506*) y *RETINOBLASTOMA-RELATED* (*RBR*), cuyas mutaciones

causan letalidad durante el desarrollo embrionario de Arabidopsis (Latvala-Kilby y Kilby, 2006; Wachsman *et al.*, 2011).

Un aspecto crítico del análisis clonal es la disponibilidad de genes que puedan servir de marcadores fenotípicos y permitan la visualización de los sectores. Para el marcaje de las células de las plantas se han utilizado con este fin mutaciones que impiden la fotosíntesis y en consecuencia causan albinismo. Las plantas a estudio suelen ser heterocigóticas para el gen marcador y el de interés, que deben estar ligados y en acoplamiento (los dos alelos mutantes en *cis* en un cromosoma y los dos silvestres en su homólogo). Además, el marcador debe estar ubicado en el mismo brazo cromosómico que el gen a estudio y entre este último y el centrómero. Los clones se obtienen tras la inducción, usualmente por irradiación, de la deleción terminal de un segmento cromosómico que contiene los alelos silvestres del marcador y del gen a estudio; se genera así un clon de células que son albinas y deficientes para la función del gen a estudio. Si el marcador fuera distal al gen a estudio podrían obtenerse sectores albinos que conservasen el alelo funcional de este último. Los genes marcadores deben ser autónomos celulares: sus productos no deben difundir de la célula que los sintetiza a sus vecinas (Poethig, 1987; 1989). La autonomía celular del marcador fenotípico es imprescindible para determinar los límites de un sector dado.

El análisis clonal resulta de particular interés para el estudio de las mutaciones letales embrionarias, ya que posibilita la observación de sus efectos en el organismo adulto. La utilidad de la inducción de sectores mutantes en etapas postembrionarias depende de la viabilidad de las células. De hecho, resulta imposible en algunos casos encontrar sectores albinos, ya que la célula fundadora del sector muere o no puede dividirse (letalidad celular). No obstante, se ha demostrado en experimentos clásicos llevados a cabo en *Drosophila melanogaster* que los sectores homocigóticos para no pocas mutaciones letales embrionarias son viables al estar rodeados de tejido adulto silvestre (Demerec, 1936; Bryant y Zornetzer, 1973). De ser así en las plantas, el estudio de sectores albinos podría servir para determinar las funciones postembrionarias de los genes letales embrionarios.

III.8.- Análisis de las funciones postembrionarias de los genes letales tempranos

III.8.1.- Identificación de genes esenciales para el desarrollo embrionario y gametofítico en Arabidopsis

En Arabidopsis existen genes esenciales cuya ausencia de función causa letalidad, que puede ser esporofítica (embrionaria) o gametofítica. Las mutaciones letales

embrionarias interrumpen la embriogénesis, impidiendo el desarrollo de una planta a partir de una semilla. Se ha denominado *EMBRYO-DEFECTIVE* (*EMB*; Meinke y Sussex, 1979a) a los genes cuyas mutaciones de insuficiencia de función —usualmente nulas— causan letalidad embrionaria. Las mutaciones letales gametofíticas afectan a los gametofitos, es decir, a la fase haploide del ciclo de vida de la planta. Se cree que el número de genes de Arabidopsis cuya ausencia de función causa letalidad embrionaria o gametofítica es de unos 1.000 (Franzmann *et al.*, 1995; McElver *et al.*, 2001; Tzafrir *et al.*, 2003; Tzafrir *et al.*, 2004; Pagnussat *et al.*, 2005; Meinke *et al.*, 2008; Boavida *et al.*, 2009; Meinke *et al.*, 2009). Los alelos nulos y letales recesivos de estos genes deben mantenerse en heterocigosis, condición en la que no suelen causar efecto fenotípico alguno.

El procedimiento más sencillo para identificar mutaciones letales embrionarias en Arabidopsis es la mutagénesis química de semillas maduras, a las que se denomina M_0 y M_1, antes y después, respectivamente, de su inmersión en una disolución del mutágeno. La detección de mutaciones letales embrionarias se lleva a cabo inspeccionando las silicuas inmaduras de las plantas M_1 tras su autofecundación, con el fin de encontrar óvulos o semillas abortivas, sin necesidad de esperar a la generación M_2 para detectar el fenotipo mutante (Meinke y Sussex, 1979b). Las silicuas de una planta heterocigótica para una mutación letal embrionaria contienen semillas abortivas, que pueden distinguirse de las silvestres por su forma y color. Aunque ambos tipos de semillas son inicialmente iguales, en las etapas finales del desarrollo embrionario las silvestres incrementan su tamaño y adquieren un color verde, mientras que las mutantes suelen permanecer incoloras y ser más pequeñas (Meinke y Sussex, 1979b).

Las semillas homocigóticas para mutaciones letales embrionarias no rinden plántulas, y por tanto escapan a la selección de plantas M_2 viables y fértiles que suele realizarse tras una mutagénesis. No pocos genes letales embrionarios, sin embargo, desempeñan funciones en etapas del desarrollo posteriores a la formación de la semilla, tal como indican sus patrones de expresión silvestres, que están ampliamente documentados en bases de datos públicas, como eFP Browser (Winter *et al.*, 2007) y TraVA (Klepikova *et al.*, 2016). Estas funciones postembrionarias no pueden ser estudiadas mediante abordajes mutacionales convencionales.

III.8.2.- Estrategias para el estudio de las funciones postembrionarias de genes letales embrionarios y gametofíticos en Arabidopsis

El estudio de las funciones postembrionarias de los genes letales embrionarios en Arabidopsis puede realizarse mediante distintos abordajes, en caso de que no existan

alelos hipomorfos que nos permitan observar directamente un fenotipo mutante en la fase vegetativa del ciclo de vida de la planta. El objetivo común de estos abordajes es que las plantas consigan superar la fase fenocrítica de la mutación a estudio (en este caso, el desarrollo embrionario) y completen su ciclo de vida. Solo así pueden estudiarse en etapas postembrionarias, como el desarrollo vegetativo, las consecuencias de la ausencia de función de genes letales embrionarios y deducir sus funciones en órganos distintos a la semilla.

Una de las estrategias para estudiar las funciones postembrionarias de genes letales embrionarios es el rescate de embriones mutantes en condiciones de cultivo específicas. Esto es especialmente útil para los mutantes auxótrofos, cuyos embriones pueden ser rescatados cultivándolos en un medio suplementado con el producto de la reacción enzimática alterada por la mutación de la que son portadores (Patton *et al.*, 1998; Muralla *et al.*, 2007; 2008). Otro abordaje es la complementación del fenotipo mutante, empleando para ello un transgén portador de una copia del alelo silvestre del gen a estudio, controlado por un promotor activo durante la embriogénesis, como los de *ABSCISIC ACID INSENSITIVE 3* (*ABI3*; Devic *et al.*, 1996; Despres *et al.*, 2001; Garcion *et al.*, 2006; Gómez *et al.*, 2010) y *RIBOSOMAL PROTEIN S5A* (*RPS5A*; Weijers *et al.*, 2001; Johnson *et al.*, 2008). También puede emplearse un promotor inducible, que permita elegir el momento de la expresión del alelo silvestre del gen letal embrionario (Aoyama y Chua, 1997; van Dijken *et al.*, 2004). Una aproximación alternativa se basa en el silenciamiento del gen letal embrionario después de la embriogénesis, que puede lograrse transfiriendo a plantas silvestres transgenes que produzcan ARN interferentes, como los microARN artificiales (Schwab *et al.*, 2006), bajo el control de un promotor inducible o del 35S del virus del mosaico de la coliflor, que no se expresa hasta el estadio embrionario de corazón (Zhang *et al.*, 1992; Robison *et al.*, 2009). Por último, el análisis clonal también constituye un método muy útil para el estudio de las funciones postembrionarias de genes letales embrionarios mediante el análisis de sectores clonales mutantes, inducidos en plantas adultas fenotípicamente silvestres. Una ventaja del análisis clonal frente a las demás técnicas es que, debido al marcaje de los sectores mutantes, proporciona también información valiosa sobre la autonomía celular y sobre el sitio de acción de los productos génicos, y puede llevarse a cabo con alelos nulos, que facilitan la interpretación de las interacciones génicas en dobles y triples mutantes.

III.8.3.- Análisis clonal de los efectos de mutaciones letales embrionarias en tejidos adultos

La letalidad temprana puede considerarse indicativa de la importancia de la función de un gen. De ser así, deberíamos esperar en no pocos casos que los genes cuyas mutaciones causan letalidad embrionaria también desempeñen algún papel en etapas posteriores del desarrollo de la planta. Dado que sus homocigotos no son viables o no pueden obtenerse, la disección de la función de los genes letales no puede realizarse en etapas postembrionarias, o en el esporofito. A pesar de que se han aislado cientos de mutaciones letales embrionarias y gametofíticas en Arabidopsis (Meinke y Sussex, 1979b; Pagnussat *et al*., 2005), no se ha realizado ningún esfuerzo sistemático para caracterizar las funciones de los genes correspondientes en plantas adultas. Sin embargo, estudios como los del gen *FUSCA1* (*FUS1*) de Arabidopsis han puesto de manifiesto el potencial del análisis clonal para caracterizar los efectos postembrionarios de las mutaciones letales embrionarias (Miséra *et al*., 1994). En efecto, los sectores somáticos *fus1* que se indujeron mediante mutagénesis química en la epidermis de plantas *FUS1*/*fus1* adultas manifestaron tricomas supernumerarios, demostrando que la actividad del gen *FUS1* se requiere en etapas postembrionarias del desarrollo.

Diversos experimentos de análisis clonal han demostrado en el maíz el papel en el desarrollo foliar de ciertos genes cuyas mutaciones causan letalidad embrionaria, como *empty pericarp2* (*emp2*) y *defective kernel1* (*dek1*) (Becraft *et al*., 2002; Fu y Scanlon, 2004). El gen *emp2* codifica una proteína similar a HEAT SHOCK BINDING PROTEIN1 de Arabidopsis (Fu *et al*., 2002). El alelo recesivo *white seedling3* (*w3*) puede usarse como marcador, ya que está ligado a *emp2* y bloquea la biosíntesis de los carotenoides, causando albinismo por destrucción fotoquímica de la clorofila (Anderson y Robertson, 1960). La irradiación de plantas *W3+ Emp2+*/*w3 emp2* puede causar rupturas cromosómicas entre el locus *w3* y su centrómero, que producen deleciones terminales con pérdida del fragmento acéntrico, generándose sectores hemicigóticos albinos *w3 emp2*/-. El análisis de estos sectores permitió identificar en la hoja un rasgo asociado a la mutación *emp2*, lo que sugiere que el gen *emp2* es necesario para el desarrollo foliar (Fu y Scanlon, 2004). Utilizando un elemento transponible *Dissociation* (*Ds*) se indujeron en el maíz sectores mutantes *dek1* marcados con un alelo mutante del gen *viviparous5* (*vp5*), que codifica la fitoeno desaturasa, una enzima de la ruta de biosíntesis de los carotenoides, cuya insuficiencia altera la síntesis de clorofila y causa albinismo. El estudio de los sectores hemicigóticos albinos permitió demostrar el papel postembrionario de *dek1* (Becraft *et al*., 2002).

III.8.4.- Técnicas de análisis clonal

III.8.4.1.- Técnicas basadas en la irradiación y los elementos transponibles

Tradicionalmente, el mosaicismo somático se ha inducido en *Drosophila melanogaster* incrementando la tasa de recombinación mitótica mediante rayos X (Patterson, 1930; Friesen, 1934; Stern, 1936). Como consecuencia de la irradiación, en algunos casos segregan dos alelos idénticos al mismo polo del huso mitótico y se singulariza el linaje de cada una de las dos células resultantes de una mitosis, que adquieren marcadores fenotípicos distintos, generándose dos clones gemelos. Esto probablemente ocurre porque los cromosomas homólogos entran en sinapsis de manera natural durante la mitosis en *Drosophila melanogaster*.

La generación de mosaicos genéticos en las plantas se ha basado usualmente en la pérdida de segmentos cromosómicos como consecuencia de rupturas producidas por irradiación con rayos X o γ, o de la actividad del elemento transponible *Ds* (Dean *et al.*, 1992; Furner y Pumfrey, 1992; Irish y Sussex, 1992; Furner y Pumfrey, 1993; Weil y Wessler, 1993; Balcells *et al.*, 1994; Becraft *et al.*, 2002; Fu y Scanlon, 2004). Por lo tanto, mientras que en *Drosophila melanogaster* los sectores se originan sin ganancia ni pérdida de material genético, en plantas como Arabidopsis o el maíz —dos especies en las que los cromosomas no se aparean durante la mitosis— los sectores son genotípicamente distintos porque los alelos recesivos quedan en hemicigosis como consecuencia de deleciones terminales. Una ventaja de los rayos X frente a los transposones es que puede controlarse el momento en que se inducen los sectores.

Las líneas CAUT (de cell autonomy) sirven de herramienta para la realización de análisis clonal mediante métodos de irradiación clásicos en Arabidopsis (Furner *et al.*, 2008). Se trata de líneas transgénicas, homocigóticas para una mutación recesiva, *chlorata-42* (*ch-42*), que confiere a la planta una pigmentación amarillenta. Las plantas CAUT, sin embargo, son verdes, ya que son portadoras de un transgén que contiene una copia del alelo silvestre del gen *CH-42*, que codifica una quelatasa de magnesio que actúa en el cloroplasto y es necesaria para la biosíntesis de clorofila (Koncz *et al.*, 1990; Gibson *et al.*, 1996). Este marcador se ha utilizado previamente en varios estudios de autonomía celular (Furner *et al.*, 1996; Stirnberg *et al.*, 2007; Furner *et al.*, 2008). Cuando el brazo cromosómico portador del transgén *CH-42* se pierde como consecuencia de una ruptura cromosómica inducida por los rayos X, aparecen sectores marcados, de color amarillento (Figura 5, en la página 27). En una línea con el genotipo apropiado, la pérdida de un brazo cromosómico puede generar un sector hemicigótico para una mutación letal y a la vez visible por su color amarillento. La colección CAUT incluye líneas transgénicas portadoras

de inserciones del transgén *CH-42* en distintas posiciones del genoma de Arabidopsis, de manera que es posible elegir la más conveniente en función de su proximidad a un gen de interés. Las líneas CAUT permiten evitar el problema que supone la escasez de marcadores autónomos celulares adecuados, y aunque se diseñaron inicialmente para estudios de autonomía celular, también pueden usarse para obtener sectores en los que se manifieste el fenotipo de una mutación letal embrionaria. Dos ventajas adicionales de las líneas CAUT son que su uso exige la realización de pocos cruzamientos, y que el investigador controla el momento de la inducción de los sectores.

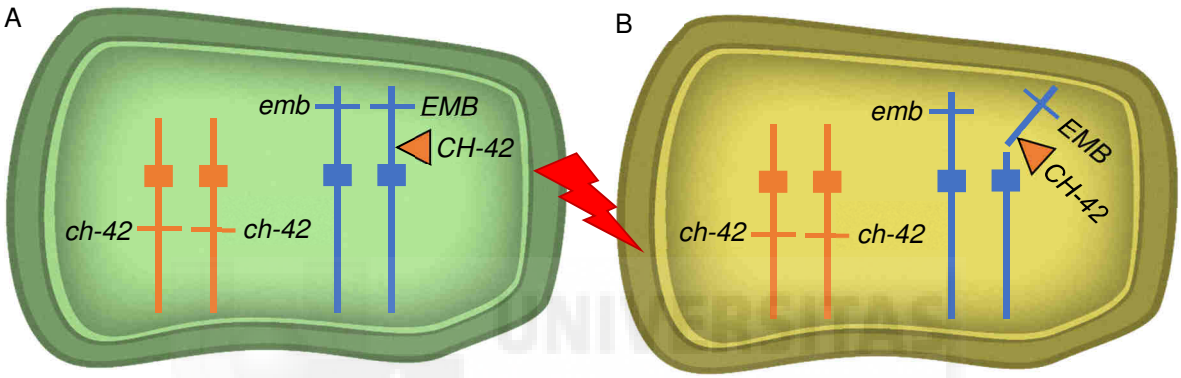


Figura 5.- Estrategia para la inducción de sectores mutantes *emb*/- empleando las líneas CAUT y rayos X. Se representan células con el genotipo apropiado (*ch-42/ch-42;emb/EMB CH-42*) para su irradiación, (A) antes y (B) después del tratamiento con rayos X, que induce una ruptura en el brazo cromosómico que contiene el alelo silvestre del gen endógeno *EMB* y el transgén *CH-42*.

III.8.4.2.- Técnicas basadas en la recombinación específica de sitio

Los métodos clásicos de inducción de mosaicos genéticos en *Drosophila melanogaster* han sido paulatinamente sustituidos por otros en los que se utilizan recombinasas específicas de sitio (Golic y Lindquist, 1989; Xu y Rubin, 1993; Siegal y Hartl, 1996; Resino *et al.*, 2002; Besse y Pret, 2003). Se ha ido extendiendo a la vez el uso de genes testigo, como los que codifican la proteína verde fluorescente de *Aequorea victoria* (GFP) y la β-glucuronidasa de *Escherichia coli* (GUS), para el marcaje de los sectores (Kilby *et al.*, 1995; Sieburth *et al.*, 1998; Kilby *et al.*, 2000; Heidstra *et al.*, 2004; Kurup *et al.*, 2005; Serralbo *et al.*, 2006; Wachsman *et al.*, 2011).

Las recombinasas específicas de sitio generan sectores mutantes al escindir la copia silvestre de un gen de interés contenida en un transgén. Las recombinasas que suelen emplearse para generar sectores mutantes son dos: FLP (flipase) del plásmido de 2 micras de la levadura *Saccharomyces cerevisiae* (McLeod *et al.*, 1986), y Cre (causes recombination) del bacteriófago P1 de *Escherichia coli* (Hoess *et al.*, 1982). El transgén

que contiene la copia silvestre del gen a estudio debe estar flanqueado por las secuencias reconocidas por la recombinasa, y su escisión acerca un promotor constitutivo a la región codificante de un gen testigo, cuya expresión marca las células mutantes. El estudio de sectores marcados ha permitido conocer los linajes celulares normales que aparecen durante el desarrollo de distintos órganos, como las raíces (Wilkins, 1992; Scheres *et al.*, 1994; Bossinger y Smyth, 1996; Kidner *et al.*, 2000; Jenik y Irish, 2001). El uso de estas estrategias basadas en transgenes es de rutina en *Drosophila melanogaster* y otros animales modelo para la inducción y estudio de mosaicos genéticos (Liu *et al.*, 2002; Besse y Pret, 2003; Dor *et al.*, 2004), pero solo recientemente se ha aplicado al análisis de las funciones postembrionarias de genes letales en Arabidopsis (Latvala-Kilby y Kilby, 2006; Wachsman *et al.*, 2011).

III.9.- Antecedentes y objetivos de esta Tesis

III.9.1.- Disección genética de procesos biológicos

El modo más simple de averiguar la función de un gen es analizar los efectos de su disfunción. La Genética se ha basado en esta idea desde principios del siglo XX, época en la que se evidenció la utilidad de las mutaciones espontáneas para la disección de procesos biológicos y comenzaron a realizarse mutagénesis en diferentes sistemas biológicos (Beadle y Tatum, 1941; Brenner, 1974; Nüsslein-Volhard y Wieschaus, 1980; Driever *et al.*, 1996). La disección genética de procesos biológicos, basada en la mutagénesis con mutágenos físicos o químicos —que dañan genes al azar—, seguida de la selección de los fenotipos de interés, ha resultado especialmente útil para desentrañar las funciones de los genes de las plantas superiores y comprender numerosos aspectos de su biología, y en particular su desarrollo (Page y Grossniklaus, 2002). Esto es especialmente patente en Arabidopsis, que ha sido objeto de numerosos análisis mutacionales en los que el aislamiento y la caracterización de mutantes ha permitido establecer las funciones de muchos genes (Baulcombe, 1999; Alonso *et al.*, 2003a; Alonso *et al.*, 2003b; Casimiro *et al.*, 2003; Koiwa *et al.*, 2006; Gille *et al.*, 2009).

La hoja es la unidad anatómica básica y el principal órgano fotosintético de una planta; su estudio y eventual manipulación pueden contribuir a incrementar la producción primaria de los cultivos que alimentan de manera directa o indirecta a toda la humanidad (Micol, 2009). Para comprender los mecanismos responsables de la morfogénesis y el crecimiento de las hojas de Arabidopsis y poder aplicar este conocimiento a otras especies vegetales, se han llevado a cabo en el laboratorio de J.L. Micol varias búsquedas de mutantes con una morfología foliar alterada. En la década de los 90 se identificaron en

estas búsquedas 153 mutantes inducidos por EMS, 28 por bombardeo con neutrones rápidos y 56 pertenecientes a una colección de dominio público, la del Arabidopsis Information Service (AIS). El análisis de complementación de estos mutantes demostró que correspondían a 94, 8 y 14 genes, respectivamente (Berná *et al.*, 1999; Robles, 1999; Serrano-Cartagena *et al.*, 1999; Pérez-Pérez *et al.*, 2009). Hasta ahora se han identificado y caracterizado 57 de estos genes, inicialmente mediante clonación posicional y más recientemente usando estrategias basadas en la secuenciación masiva (mapping-by-sequencing; Mateo-Bonmatí *et al.*, 2014).

III.9.2.- Los mutantes *angulata*

Los mutantes de la clase fenotípica Angulata (Anu) presentan hojas pálidas, apuntadas y dentadas. Estos rasgos sugieren, respectivamente, la disfunción de los cloroplastos y la alteración de la expansión lateral de la hoja. En la mutagénesis con EMS llevada a cabo en el laboratorio de J.L. Micol en los años noventa se aislaron 18 mutantes *anu*. Su análisis de complementación indicó que correspondían a 12 genes distintos (*ANU1-ANU12*; Berná *et al.*, 1999), cuyas posiciones de mapa genético se establecieron mediante análisis del ligamiento a microsatélites polimórficos (Robles y Micol, 2001).

Los mutantes *anu* presentan alteraciones severas en la estructura interna de la hoja (Pérez-Pérez *et al.*, 2011; Casanova-Sáez *et al.*, 2014). La biogénesis del cloroplasto y la morfogénesis foliar son dos procesos estrechamente relacionados, probablemente mediante señalización retrógrada del cloroplasto al núcleo (Pyke *et al.*, 2000; Rodermel, 2001; Andriankaja *et al.*, 2012; Tameshige *et al.*, 2013). Los mutantes *anu,* en consecuencia, son potencialmente útiles para la identificación de nuevas funciones del cloroplasto y para establecer la influencia de este orgánulo sobre el desarrollo foliar.

Se han clonado hasta ahora cinco genes *ANU* en el laboratorio de J.L. Micol: *ANU1*, *ANU4*, *ANU9* y *ANU12* (Mateo-Bonmatí *et al.*, 2014), y *ANU10* (Casanova-Sáez *et al.*, 2014). Todos ellos son genes nucleares cuyos productos forman parte del proteoma del cloroplasto con funciones muy diversas, como la degradación de proteínas o su transporte a través de las membranas de este orgánulo.

III.9.3.- Los mutantes *orbiculata*

Los rasgos más conspicuos de los mutantes de la clase fenotípica Orbiculata (Orb) son sus hojas vegetativas redondeadas y parcialmente despigmentadas, así como un menor tamaño de la roseta con respecto a la de su tipo silvestre Landsberg *erecta* (L*er*). Al

igual que ocurre con los mutantes *anu*, la despigmentación foliar sugiere una alteración del funcionamiento de los cloroplastos en los mutantes *orb*.

En la búsqueda de mutaciones inducidas con EMS antes mencionada (Berná *et al.*, 1999) se aislaron 6 mutantes *orb*. Su análisis de complementación indicó que correspondían a 2 genes distintos: *ORB1 y ORB2*, cuyas posiciones de mapa se determinaron mediante cartografía de baja resolución (Robles y Micol, 2001).

III.9.4.- Limitaciones del abordaje mutacional convencional

Las colecciones de mutantes tienen una utilidad indiscutible, que no está exenta de limitaciones. Una de ellas es la redundancia funcional, fenómeno que se da entre genes parálogos que son resultado de una duplicación relativamente reciente; la obtención de un alelo nulo de uno de ellos no tiene manifestación fenotípica en muchos casos. La otra dificultad a la que se enfrentan quienes emplean abordajes mutacionales es la letalidad embrionaria y/o gametofítica, ya que una parte pequeña pero importante de los genes de cualquier genoma rinden exclusivamente alelos que no son viables. Se ha estimado que el número de genes con alelos letales embrionarios en Arabidopsis es de 500-750 (Franzmann *et al.*, 1995; McElver *et al.*, 2001). Se han identificado además varios cientos de mutaciones recesivas que causan letalidad gametofítica masculina o femenina (Pagnussat *et al.*, 2005; Boavida *et al.*, 2009). En consecuencia, no son pocos los genes del genoma de Arabidopsis que codifican proteínas con funciones esenciales en las etapas tempranas de la fase diploide o la fase gametofítica. Aunque muchos de estos genes se expresan en estadios de desarrollo posteriores a la embriogénesis, sus funciones postembrionarias no pueden ser estudiadas mediante abordajes mutacionales convencionales.

Los alelos nulos de los genes *EMB* (apartado III.8.1, en la página 22) causan defectos en el embrión, que interrumpe su desarrollo en alguna de las fases de la embriogénesis. Los primeros mutantes *emb* fueron aislados y caracterizados a principios de los años 60, y constituyen una herramienta imprescindible para el estudio del desarrollo embrionario de Arabidopsis. Los primeros 72 mutantes *emb* de Arabidopsis fueron obtenidos por A.J. Müller (1963), que los clasificó en 6 clases fenotípicas, según el momento en el que se interrumpía su desarrollo embrionario, el tamaño de los embriones y su color. Otra clasificación posterior subdivide a las mutaciones *emb* según afecten a procesos celulares esenciales, a etapas concretas de la embriogénesis, o causen auxotrofías (Meinke y Sussex, 1979a).

A lo largo de las últimas décadas se han realizado en Arabidopsis varias mutagénesis con ADN-T con el fin de saturar el genoma con mutaciones letales embrionarias (Franzmann *et al.*, 1995; McElver *et al.*, 2001; Tzafrir *et al.*, 2004). Además, iniciativas como el proyecto SeedGenes pretendían identificar todos los genes con una función esencial en el desarrollo embrionario de Arabidopsis (Tzafrir *et al.*, 2003).

III.9.5.- Estudio de las funciones postembrionarias de los genes letales embrionarios

Muchos genes esenciales para el desarrollo embrionario y gametofítico permanecen activos en etapas postembrionarias del desarrollo, como la organogénesis foliar, en las que por tanto podrían desempeñar funciones importantes. Nuestra hipótesis de partida en esta Tesis fue considerar a los genes *EMB* candidatos a aportar información sobre los mecanismos que conducen a la formación de un órgano adulto: la hoja. Definimos tres posibles abordajes al estudio de las funciones postembrionarias de los genes *EMB*. El primero, y más simple, se basaba en la idea de que el catálogo de mutaciones *emb* está incompleto y que una búsqueda exhaustiva permitiría identificar alelos adicionales, alguno de los cuales sería hipomorfo, viable e informativo. El segundo abordaje, aparentemente con menos expectativas de éxito, era la búsqueda de *escapers*, individuos de genotipo letal que superan la fase fenocrítica asociada a la insuficiencia de la función que padecen y que sobreviven manifestando aberraciones morfológicas severas pero ilustrativas de los procesos en los que participa el gen a estudio. Solo se han descrito *escapers* para algunos genes letales y suelen suponer una fracción muy baja del conjunto de los individuos de genotipo letal. El tercer abordaje, técnicamente más complejo que los anteriores, era el análisis clonal: la obtención de mosaicos genéticos en los que se indujese la aparición de un clon de células homocigóticas para una mutación letal embrionaria en un tejido adulto de una planta por lo demás fenotípicamente silvestre y heterocigótica para dicha mutación recesiva (Candela *et al.*, 2011).

III.9.6.- Objetivos de esta Tesis

Los objetivos generales de esta Tesis incluían la caracterización genética y molecular de los mutantes (1) *anu7* y (2) *orb1*, y (3) la generación de herramientas de análisis clonal para el estudio de las funciones postembrionarias de genes letales embrionarios y de su posible participación en la organogénesis foliar.

En cuanto a los genes *ANU7* y *ORB1*, el primero de nuestros objetivos concretos fue la caracterización del fenotipo morfológico e histológico de los alelos hipomorfos

aislados en el laboratorio de J.L. Micol tras una mutagénesis con EMS del ecotipo silvestre L*er*. Otro de nuestros propósitos fue establecer la identidad de *ANU7* y *ORB1* mediante clonación posicional y secuenciación de genes candidatos (en el caso de *anu7-1*), o combinando la clonación posicional con las tecnologías de secuenciación masiva (en el de *orb1-1*). También teníamos interés en analizar el transcriptoma de *anu7-1* y *orb1-3*, con el fin de establecer sus diferencias con el de L*er*, y determinar qué procesos estaban alterados en estos mutantes. Además, en el caso de *anu7-1*, otro de los objetivos que nos planteamos fue estudiar sus interacciones genéticas con los genes *gun* (apartado III.3.3, en la página 12), que están relacionados con la señalización retrógrada del cloroplasto al núcleo.

El objetivo último de la caracterización de los mutantes *anu7* y *orb1* era proponer modelos que explicasen la relación entre sus alteraciones morfológicas y las funciones silvestres de sus genes mutados en el desarrollo foliar, y de la planta en su conjunto.

En lo relativo al estudio de las funciones postembrionarias de genes letales embrionarios mediante análisis clonal, nuestro primer objetivo fue la selección de un grupo de genes que pudieran resultar de interés. Elegimos genes *EMB* que solo tuvieran alelos letales embrionarios y que se expresaran en las hojas vegetativas silvestres. Nos propusimos a continuación generar una colección de líneas transgénicas con el fenotipo adecuado para la inducción de sectores mutantes mediante el uso de las líneas CAUT y la irradiación con rayos X. También pretendíamos generar una colección de líneas transgénicas con el fenotipo adecuado para la inducción de sectores mutantes mediante recombinación específica de sitio mediada por Cre, inducida por un choque térmico.

El objetivo para el cual desarrollamos estas herramientas de análisis clonal era la inducción y el posterior escrutinio y caracterización de sectores mutantes, con el fin de establecer relaciones entre los fenotipos mutantes observados en los sectores y las posibles funciones postembrionarias de los genes letales embrionarios en la organogénesis foliar.

IV.- BIBLIOGRAFÍA DE LA INTRODUCCIÓN

UNIVERSITAS
Miguel
Hernández

IV.- BIBLIOGRAFÍA DE LA INTRODUCCIÓN

Ajjawi, I., Coku, A., Froehlich, J.E., Yang, Y., Osteryoung, K.W., Benning, C., y Last, R.L. (2011). A J-like protein influences fatty acid composition of chloroplast lipids in Arabidopsis. *PLOS ONE* **6,** e25368.

Albrecht, V., Ingenfeld, A., y Apel, K. (2008). *Snowy cotyledon 2*: the identification of a zinc finger domain protein essential for chloroplast development in cotyledons but not in true leaves. *Plant Molecular Biology* **66,** 599-608.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., y Ecker, J.R. (2003a). Genome-wide insertional mutagenesis of *Arabidopsis thaliana. Science* **301,** 653-657.

Alonso, J.M., Stepanova, A.N., Solano, R., Wisman, E., Ferrari, S., Ausubel, F.M., y Ecker, J.R. (2003b). Five components of the ethylene-response pathway identified in a screen for weak ethylene-insensitive mutants in Arabidopsis. *Proceedings of the National Academy of Sciences USA* **100,** 2992-2997.

Anderson, I.C., y Robertson, D.S. (1960). Role of carotenoids in protecting chlorophyll from photodestruction. *Plant Physiology* **35,** 531-534.

Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., Mühlenbock, P., Skirycz, A., Gonzalez, N., Beemster, G.T., y Inzé, D. (2012). Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. *Developmental Cell* **22,** 64-78.

Aoyama, T., y Chua, N.H. (1997). A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant Journal* **11,** 605-612.

Arsova, B., Hoja, U., Wimmelbacher, M., Greiner, E., Üstün, S., Melzer, M., Petersen, K., Lein, W., y Börnke, F. (2010). Plastidial thioredoxin z interacts with two fructokinase-like proteins in a thiol-dependent manner: evidence for an essential role in chloroplast development in *Arabidopsis* and *Nicotiana benthamiana. Plant Cell* **22,** 1498-1515.

Baba, K., Schmidt, J., Espinosa-Ruiz, A., Villarejo, A., Shiina, T., Gardeström, P., Sane, A.P., y Bhalerao, R.P. (2004). Organellar gene transcription and early seedling development are affected in the *rpoT;2* mutant of *Arabidopsis*. *Plant Journal* **38,** 38-48.

Balcells, L., Sundberg, E., y Coupland, G. (1994). A heat-shock promoter fusion to the *Ac* transposase gene drives inducible transposition of a *Ds* element during *Arabidopsis* embryo development. *Plant Journal* **5,** 755-764.

Barajas-López, J.D., Blanco, N.E., y Strand, A. (2013). Plastid-to-nucleus communication, signals controlling the running of the plant cell. *Biochimica et Biophysica Acta* **1833,** 425-437.

Bardwell, J.C., y Craig, E.A. (1984). Major heat shock gene of *Drosophila* and the *Escherichia coli* heat-inducible *dnaK* gene are homologous. *Proceedings of the National Academy of Sciences USA* **81,** 848-852.

Baulcombe, D.C. (1999). Fast forward genetics based on virus-induced gene silencing. *Current Opinion in Plant Biology* **2,** 109-113.

Beadle, G.W., y Tatum, E.L. (1941). Genetic control of biochemical reactions in Neurospora. *Proceedings of the National Academy of Sciences USA* **27,** 499-506.

Becraft, P.W., Bongard-Pierce, D.K., Sylvester, A.W., Poethig, R.S., y Freeling, M. (1990). The *liguleless-1* gene acts tissue specifically in maize leaf development. *Developmental Biology* **141,** 220-232.

Becraft, P.W. (1999). Development of the leaf epidermis. *Current Topics in Developmental Biology* **45,** 1-40.

Becraft, P.W., Li, K., Dey, N., y Asuncion-Crabb, Y. (2002). The maize *dek1* gene functions in embryonic pattern formation and cell fate specification. *Development* **129,** 5217-5225.

Bell, C.J., y Ecker, J.R. (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19,** 137-144.

Berná, G., Robles, P., y Micol, J.L. (1999). A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana. Genetics* **152,** 729-742.

Besse, F., y Pret, A.M. (2003). Apoptosis-mediated cell death within the ovarian polar cell lineage of *Drosophila melanogaster. Development* **130,** 1017-1027.

Bilsborough, G.D., Runions, A., Barkoulas, M., Jenkins, H.W., Hasson, A., Galinha, C., Laufs, P., Hay, A., Prusinkiewicz, P., y Tsiantis, M. (2011). Model for the regulation of *Arabidopsis thaliana* leaf margin development. *Proceedings of the National Academy of Sciences USA* **108,** 3424-3429.

Boavida, L.C., Shuai, B., Yu, H.J., Pagnussat, G.C., Sundaresan, V., y McCormick, S. (2009). A collection of *Ds* insertional mutants associated with defects in male gametophyte development and function in *Arabidopsis thaliana. Genetics* **181,** 1369-1385.

Bolker, J.A. (1995). Model systems in developmental biology. *Bioessays* **17,** 451-455.

Börner, T., Aleynikova, A.Y., Zubo, Y.O., y Kusnetsov, V.V. (2015). Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochimica et Biophysica Acta* **1847,** 761-769.

Bossinger, G., y Smyth, D.R. (1996). Initiation patterns of flower and floral organ development in *Arabidopsis thaliana. Development* **122,** 1093-1102.

Bradbeer, J.W., Atkinson, Y.E., Börner, T., y Hagemann, R. (1979). Cytoplasmic synthesis of plastid polypeptides may be controlled by plastid-synthesised RNA. *Nature* **279,** 816-817.

Braybrook, S.A., y Kuhlemeier, C. (2010). How a plant builds leaves. *Plant Cell* **22,** 1006-1018.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans. Genetics* **77,** 71-94.

Brutnell, T.P., Sawers, R.J., Mant, A., y Langdale, J.A. (1999). BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the *rbcL* gene of maize. *Plant Cell* **11,** 849-864.

Bryant, P.J., y Zornetzer, M. (1973). Mosaic analysis of lethal mutations in Drosophila. *Genetics* **75,** 623-637.

Buchanan, B.B., Gruissem, W., y Jones, R.L. (2015). *Biochemistry and Molecular Biology of Plants*. John Wiley & Sons.

Byrne, M., Timmermans, M., Kidner, C., y Martienssen, R. (2001). Development of leaf shape. *Current Opinion in Plant Biology* **4,** 38-43.

Candela, H., Pérez-Pérez, J.M., y Micol, J.L. (2011). Uncovering the post-embryonic functions of gametophytic- and embryonic-lethal genes. *Trends in Plant Science* **16,** 336-345.

Candela, H., Casanova-Sáez, R., y Micol, J.L. (2015). Getting started in mapping-by-sequencing. *Journal of Integrative Plant Biology* **57,** 606-612.

Caplan, A.J., Cyr, D.M., y Douglas, M.G. (1993). Eukaryotic homologues of *Escherichia coli* dnaJ: a diverse protein family that functions with HSP70 stress proteins. *Molecular Biology of the Cell* **4,** 555-563.

Casanova-Sáez, R., Mateo-Bonmatí, E., Kangasjarvi, S., Candela, H., y Micol, J.L. (2014). *Arabidopsis* ANGULATA10 is required for thylakoid biogenesis and mesophyll development. *Journal of Experimental Botany* **65,** 2391-2404.

Casimiro, I., Beeckman, T., Graham, N., Bhalerao, R., Zhang, H., Casero, P., Sandberg, G., y Bennett, M.J. (2003). Dissecting *Arabidopsis* lateral root development. *Trends in Plant Science* **8,** 165-171.

Coen, E.S., y Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* **353,** 31-37.

Coschigano, K.T., Melo-Oliveira, R., Lim, J., y Coruzzi, G.M. (1998). Arabidopsis *gls* mutants and distinct Fd-GOGAT genes. Implications for photorespiration and primary nitrogen assimilation. *Plant Cell* **10,** 741-752.

Cyr, D.M. (1995). Cooperation of the molecular chaperone Ydj1 with specific Hsp70 homologs to suppress protein aggregation. *FEBS Letters* **359,** 129-132.

Chan, K.X., Phua, S.Y., Crisp, P., McQuinn, R., y Pogson, B.J. (2016). Learning the languages of the chloroplast: retrograde signaling and beyond. *Annual Review of Plant Biology* **67,** 25-53.

Chang, S.H., Lee, S., Um, T.Y., Kim, J.K., Do Choi, Y., y Jang, G. (2017). pTAC10, a key subunit of plastid-encoded RNA polymerase, promotes chloroplast development. *Plant Physiology* **174,** 435-449.

Cheetham, M.E., y Caplan, A.J. (1998). Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress and Chaperones* **3,** 28-36.

Chen, H., Li, C., Liu, L., Zhao, J., Cheng, X., Jiang, G., y Zhai, W. (2016). The Fd-GOGAT1 mutant gene *lc7* confers resistance to *Xanthomonas oryzae* pv. *Oryzae* in rice. *Scientific Reports* **6,** 26411.

Chen, K.M., Holmstrom, M., Raksajit, W., Suorsa, M., Piippo, M., y Aro, E.M. (2010). Small chloroplast-targeted DnaJ proteins are involved in optimization of photosynthetic reactions in *Arabidopsis thaliana. BMC Plant Biology* **10,** 43.

Chen, K.M., Piippo, M., Holmström, M., Nurmi, M., Pakula, E., Suorsa, M., y Aro, E.M. (2011). A chloroplast-targeted DnaJ protein AtJ8 is negatively regulated by light and has rapid turnover in darkness. *Journal of Plant Physiology* **168,** 1780-1783.

Chiu, C.C., Chen, L.J., y Li, H.M. (2010). Pea chloroplast DnaJ-J8 and Toc12 are encoded by the same gene and localized in the stroma. *Plant Physiology* **154,** 1172-1182.

Chiu, C.C., Chen, L.J., Su, P.H., y Li, H.M. (2013). Evolution of chloroplast J proteins. *PLOS ONE* **8,** e70384.

Chivers, P.T., Prehoda, K.E., y Raines, R.T. (1997). The CXXC motif: a rheostat in the active site. *Biochemistry* **36,** 4061-4066.

Daniell, H., Lin, C.S., Yu, M., y Chang, W.J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology* **17,** 134.

de Crouy-Chanel, A., Kohiyama, M., y Richarme, G. (1995). A novel function of *Escherichia coli* chaperone DnaJ. Protein-disulfide isomerase. *Journal of Biological Chemistry* **270,** 22669-22672.

Dean, C., Sjodin, C., Page, T., Jones, J., y Lister, C. (1992). Behaviour of the maize transposable element *Ac* in *Arabidopsis thaliana. Plant Journal* **2,** 69-81.

Dekker, J.P., y Boekema, E.J. (2005). Supramolecular organization of thylakoid membrane proteins in green plants. *Biochimica et Biophysica Acta* **1706,** 12-39.

Demerec, M. (1936). Frequency of "cell-lethals" among lethals obtained at random in the X-chromosome of *Drosophila melanogaster. Proceedings of the National Academy of Sciences USA* **22,** 350-354.

Despres, B., Delseny, M., y Devic, M. (2001). Partial complementation of *embryo defective* mutations: a general strategy to elucidate gene function. *Plant Journal* **27,** 149-159.

Devic, M., Albert, S., y Delseny, M. (1996). Induction and expression of seed-specific promoters in *Arabidopsis* embryo-defective mutants. *Plant Journal* **9,** 205-215.

Dor, Y., Brown, J., Martinez, O.I., y Melton, D.A. (2004). Adult pancreatic β-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429,** 41-46.

Driever, W., Solnica-Krezel, L., Schier, A.F., Neuhauss, S.C., Malicki, J., Stemple, D.L., Stainier, D.Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., y Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123,** 37-46.

Dyall, S.D., Brown, M.T., y Johnson, P.J. (2004). Ancient invasions: from endosymbionts to organelles. *Science* **304,** 253-257.

Eshed, Y., Izhaki, A., Baum, S.F., Floyd, S.K., y Bowman, J.L. (2004). Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. *Development* **131,** 2997-3006.

Fan, C.Y., Lee, S., y Cyr, D.M. (2003). Mechanisms for regulation of Hsp70 function by Hsp40. *Cell Stress and Chaperones* **8,** 309-316.

Ferrario-Méry, S., Suzuki, A., Kunz, C., Valadier, M.H., Roux, Y., Hirel, B., y Foyer, C.H. (2000). Modulation of amino acid metabolism in transformed tobacco plants deficient in Fd-GOGAT. *Plant and Soil* **221,** 67-79.

Fleming, A.J. (2005). Formation of primordia and phyllotaxy. *Current Opinion in Plant Biology* **8,** 53-58.

Franzmann, L.M., Yoon, E.S., y Meinke, D.W. (1995). Saturating the genetic map of *Arabidopsis thaliana* with embryonic mutations. *Plant Journal* **7,** 341-350.

Friesen, H. (1934). Causes of suppression of crossing-over in males of *Drosophila melanogaster. Nature* **134**.

Fristedt, R., Williams-Carrier, R., Merchant, S.S., y Barkan, A. (2014). A thylakoid membrane protein harboring a DnaJ-type zinc finger domain is required for photosystem I accumulation in plants. *Journal of Biological Chemistry* **289,** 30657-30667.

Frydman, J., Nimmesgern, E., Ohtsuka, K., y Hartl, F.U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370,** 111-117.

Fu, S., Meeley, R., y Scanlon, M.J. (2002). *empty pericarp2* encodes a negative regulator of the heat shock response and is required for maize embryogenesis. *Plant Cell* **14,** 3119-3132.

Fu, S., y Scanlon, M.J. (2004). Clonal mosaic analysis of EMPTY PERICARP2 reveals nonredundant functions of the duplicated HEAT SHOCK FACTOR BINDING PROTEINs during maize shoot development. *Genetics* **167,** 1381-1394.

Furner, I., Ellis, L., Bakht, S., Mirza, B., y Sheikh, M. (2008). CAUT lines: a novel resource for studies of cell autonomy in Arabidopsis. *Plant Journal* **53,** 645-660.

Furner, I.J., y Pumfrey, J.A. (1992). Cell fate in the shoot apical meristem of *Arabidopsis thaliana. Development* **115,** 755-764.

Furner, I.J., y Pumfrey, J.E. (1993). Cell fate in the inflorescence meristem and floral buttress of *Arabidopsis thaliana. Plant Journal* **4,** 917-931.

Furner, I.J., Ainscough, J.F., Pumfrey, J.A., y Petty, L.M. (1996). Clonal analysis of the late flowering *fca* mutant of *Arabidopsis thaliana*: cell fate and cell autonomy. *Development* **122,** 1041-1050.

Gao, Z.P., Yu, Q.B., Zhao, T.T., Ma, Q., Chen, G.X., y Yang, Z.N. (2011). A functional component of the transcriptionally active chromosome complex, Arabidopsis pTAC14, interacts with pTAC12/HEMERA and regulates plastid gene expression. *Plant Physiology* **157,** 1733-1745.

Gao, Z.P., Chen, G.X., y Yang, Z.N. (2012). Regulatory role of Arabidopsis pTAC14 in chloroplast development and plastid gene expression. *Plant Signaling and Behavior* **7,** 1354-1356.

Garcia-Bellido, A., y Merriam, J.R. (1969). Cell lineage of the imaginal discs in *Drosophila gynandromorphs. Journal of Experimental Zoology* **170,** 61-75.

Garcia, M., Myouga, F., Takechi, K., Sato, H., Nabeshima, K., Nagata, N., Takio, S., Shinozaki, K., y Takano, H. (2008). An *Arabidopsis* homolog of the bacterial peptidoglycan synthesis enzyme MurE has an essential role in chloroplast development. *Plant Journal* **53,** 924-934.

Garcia, V., Bres, C., Just, D., Fernandez, L., Tai, F.W., Mauxion, J.P., Le Paslier, M.C., Bérard, A., Brunel, D., Aoki, K., Alseekh, S., Fernie, A.R., Fraser, P.D., y Rothan, C. (2016). Rapid identification of causal mutations in tomato EMS populations via mapping-by-sequencing. *Nature Protocols* **11,** 2401-2418.

Garcion, C., Guilleminot, J., Kroj, T., Parcy, F., Giraudat, J., y Devic, M. (2006). AKRP and EMB506 are two ankyrin repeat proteins essential for plastid differentiation and plant development in Arabidopsis. *Plant Journal* **48,** 895-906.

Gaufichon, L., Reisdorf-Cren, M., Rothstein, S.J., Chardon, F., y Suzuki, A. (2010). Biological functions of asparagine synthetase in plants. *Plant Science* **179,** 141-153.

Gaufichon, L., Masclaux-Daubresse, C., Tcherkez, G., Reisdorf-Cren, M., Sakakibara, Y., Hase, T., Clement, G., Avice, J.C., Grandjean, O., Marmagne, A., Boutet-Mercey, S., Azzopardi, M., Soulay, F., y Suzuki, A. (2013). *Arabidopsis thaliana ASN2* encoding asparagine synthetase is involved in the control of nitrogen assimilation and export during vegetative growth. *Plant, Cell and Environment* **36,** 328-342.

Gaufichon, L., Rothstein, S.J., y Suzuki, A. (2016). Asparagine metabolic pathways in Arabidopsis. *Plant and Cell Physiology* **57,** 675-689.

Gibson, L.C., Marrison, J.L., Leech, R.M., Jensen, P.E., Bassham, D.C., Gibson, M., y Hunter, C.N. (1996). A putative Mg chelatase subunit from *Arabidopsis thaliana* cv C24. *Plant Physiology* **111,** 61-71.

Gilkerson, J., Perez-Ruiz, J.M., Chory, J., y Callis, J. (2012). The plastid-localized pfkB-type carbohydrate kinases FRUCTOKINASE-LIKE 1 and 2 are essential for growth and development of *Arabidopsis thaliana. BMC Plant Biology* **12,** 102.

Gille, S., Hansel, U., Ziemann, M., y Pauly, M. (2009). Identification of plant cell wall mutants by means of a forward chemical genetic approach using hydrolases. *Proceedings of the National Academy of Sciences USA* **106,** 14699-14704.

Glynn, J.M., Froehlich, J.E., y Osteryoung, K.W. (2008). *Arabidopsis* ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane space. *Plant Cell* **20,** 2460-2470.

Golic, K.G., y Lindquist, S. (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. *Cell* **59,** 499-509.

Gómez, L.D., Gilday, A., Feil, R., Lunn, J.E., y Graham, I.A. (2010). *AtTPS1*-mediated trehalose 6-phosphate synthesis is essential for embryogenic and vegetative growth and responsiveness to ABA in germinating seeds and stomatal guard cells. *Plant Journal* **64,** 1-13.

Ha, C.M., Jun, J.H., Nam, H.G., y Fletcher, J.C. (2007). *BLADE-ON-PETIOLE 1* and *2* control Arabidopsis lateral organ fate through regulation of LOB domain and adaxial-abaxial polarity genes. *Plant Cell* **19,** 1809-1825.

Hajdukiewicz, P.T., Allison, L.A., y Maliga, P. (1997). The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO Journal* **16,** 4041-4048.

Harper, L., y Freeling, M. (1996). Interactions of *liguleless1* and *liguleless2* function during ligule induction in maize. *Genetics* **144,** 1871-1882.

Hartl, F.U., y Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295,** 1852-1858.

Heidstra, R., Welch, D., y Scheres, B. (2004). Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis* SCARECROW action in asymmetric cell division. *Genes and Development* **18,** 1964-1969.

Heitzler, P., y Simpson, P. (1991). The choice of cell fate in the epidermis of Drosophila. *Cell* **64,** 1083-1092.

Hendrick, J.P., Langer, T., Davis, T.A., Hartl, F.U., y Wiedmann, M. (1993). Control of folding and membrane translocation by binding of the chaperone DnaJ to nascent polypeptides. *Proceedings of the National Academy of Sciences USA* **90,** 10216-10220.

Hennessy, F., Nicoll, W.S., Zimmermann, R., Cheetham, M.E., y Blatch, G.L. (2005). Not all J domains are created equal: Implications for the specificity of Hsp40–Hsp70 interactions. *Protein Science* **14,** 1697-1709.

Hernandez, M.L., Passas, H.J., y Smith, L.G. (1999). Clonal analysis of epidermal patterning during maize leaf development. *Developmental Biology* **216,** 646-658.

Hess, W.R., y Börner, T. (1999). Organellar RNA polymerases of higher plants. *International Review of Cytology* **190,** 1-59.

Hoess, R.H., Ziese, M., y Sternberg, N. (1982). P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proceedings of the National Academy of Sciences USA* **79,** 3398-3402.

Horiguchi, G., Ferjani, A., Fujikura, U., y Tsukaya, H. (2006). Coordination of cell proliferation and cell expansion in the control of leaf size in *Arabidopsis thaliana. Journal of Plant Research* **119,** 37-42.

Hricová, A., Quesada, V., y Micol, J.L. (2006). The *SCABRA3* nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in Arabidopsis. *Plant Physiology* **141,** 942-956.

Huang, M., Friso, G., Nishimura, K., Qu, X., Olinares, P.D., Majeran, W., Sun, Q., y van Wijk, K.J. (2013). Construction of plastid reference proteomes for maize and *Arabidopsis* and evaluation of their orthologous relationships; the concept of orthoproteomics. *Journal of Proteome Research* **12,** 491-504.

Irish, V.F., y Sussex, I.M. (1992). A fate map of the *Arabidopsis* embryonic shoot apical meristem. *Development* **115,** 745-753.

Jamai, A., Salomé, P.A., Schilling, S.H., Weber, A.P., y McClung, C.R. (2009). Arabidopsis photorespiratory serine hydroxymethyltransferase activity requires the mitochondrial accumulation of ferredoxin-dependent glutamate synthase. *Plant Cell* **21,** 595-606.

James, G.V., Patel, V., Nordström, K.J., Klasen, J.R., Salomé, P.A., Weigel, D., y Schneeberger, K. (2013). User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biology* **14,** R61.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M., y Last, R.L. (2002). Arabidopsis map-based cloning in the post-genome era. *Plant Physiology* **129,** 440-450.

Jenik, P.D., e Irish, V.F. (2001). The *Arabidopsis* floral homeotic gene *APETALA3* differentially regulates intercellular signaling required for petal and stamen development. *Development* **128,** 13-23.

Jeon, Y., Jung, H.J., Kang, H., Park, Y.I., Lee, S.H., y Pai, H.S. (2012). S1 domain-containing STF modulates plastid transcription and chloroplast biogenesis in *Nicotiana benthamiana. New Phytologist* **193,** 349-363.

Johnson, H.B. (1975). Plant pubescence: An ecological perspective. *Botanical Review* **41,** 233-258.

Johnson, K.L., Faulkner, C., Jeffree, C.E., e Ingram, G.C. (2008). The phytocalpain *defective kernel 1* is a novel *Arabidopsis* growth regulator whose activity is regulated by proteolytic processing. *Plant Cell* **20,** 2619-2630.

Jung, H.S., y Chory, J. (2010). Signaling between chloroplasts and the nucleus: can a systems biology approach bring clarity to a complex and highly regulated pathway? *Plant Physiology* **152,** 453-459.

Kidner, C., Sundaresan, V., Roberts, K., y Dolan, L. (2000). Clonal analysis of the *Arabidopsis* root confirms that position, not lineage, determines cell fate. *Planta* **211,** 191-199.

Kilby, N.J., Davies, G.J., y Snaith, M.R. (1995). FLP recombinase in transgenic plants: constitutive activity in stably transformed tobacco and generation of marked cell clones in *Arabidopsis*. *Plant Journal* **8,** 637-652.

Kilby, N.J., Fyvie, M.J., Sessions, R.A., Davies, G.J., y Murray, J.A. (2000). Controlled induction of GUS marked clonal sectors in *Arabidopsis*. *Journal of Experimental Botany* **51,** 853-863.

Kindgren, P., Kremnev, D., Blanco, N.E., Barajas-López, J.D., Fernández, A.P., Tellgren-Roth, C., Kleine, T., Small, I., y Strand, A. (2012). The plastid *redox insensitive 2* mutant of Arabidopsis is impaired in PEP activity and high light-dependent plastid redox signalling to the nucleus. *Plant Journal* **70,** 279-291.

Kindgren, P., y Strand, A. (2015). Chloroplast transcription, untangling the Gordian Knot. *New Phytologist* **206,** 889-891.

Kleine, T., Maier, U.G., y Leister, D. (2009). DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology* **60,** 115-138.

Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., y Penin, A.A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant Journal* **88,** 1058-1070.

Koiwa, H., Bressan, R.A., y Hasegawa, P.M. (2006). Identification of plant stress-responsive determinants in Arabidopsis by large-scale forward genetic screens. *Journal of Experimental Botany* **57,** 1119-1128.

Koncz, C., Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Reiss, B., Redei, G.P., y Schell, J. (1990). Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana. EMBO Journal* **9,** 1337-1346.

Kong, F., Deng, Y., Wang, G., Wang, J., Liang, X., y Meng, Q. (2014a). LeCDJ1, a chloroplast DnaJ protein, facilitates heat tolerance in transgenic tomatoes. *Journal of Integrative Plant Biology* **56,** 63-74.

Kong, F., Deng, Y., Zhou, B., Wang, G., Wang, Y., y Meng, Q. (2014b). A chloroplast-targeted DnaJ protein contributes to maintenance of photosystem II under chilling stress. *Journal of Experimental Botany* **65,** 143-158.

Koornneef, M., Van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J., y Feenstra, W.J. (1983). Linkage map of *Arabidopsis thaliana. Journal of Heredity* **74** 265-272.

Koornneef, M., y Meinke, D. (2010). The development of *Arabidopsis* as a model plant. *Plant Journal* **61,** 909-921.

Koussevitzky, S., Nott, A., Mockler, T.C., Hong, F., Sachetto-Martins, G., Surpin, M., Lim, J., Mittler, R., y Chory, J. (2007). Signals from chloroplasts converge to regulate nuclear gene expression. *Science* **316,** 715-719.

Kurup, S., Runions, J., Köhler, U., Laplaze, L., Hodge, S., y Haseloff, J. (2005). Marking cell lineages in living tissues. *Plant Journal* **42,** 444-453.

Kurzik-Dumke, U., Gundacker, D., Renthrop, M., y Gateff, E. (1995). Tumor suppression in *Drosophila* is causally related to the function of the *lethal(2) tumorous imaginal discs* gene, a *dnaJ* homolog. *Developmental Genetics* **16,** 64-76.

Lam, H.M., Coschigano, K., Schultz, C., Melo-Oliveira, R., Tjaden, G., Oliveira, I., Ngai, N., Hsieh, M.H., y Coruzzi, G. (1995). Use of Arabidopsis mutants and genes to study amide amino acid biosynthesis. *Plant Cell* **7,** 887-898.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K., y Hartl, F.U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356,** 683-689.

Larkin, R.M., Alonso, J.M., Ecker, J.R., y Chory, J. (2003). GUN4, a regulator of chlorophyll synthesis and intracellular signaling. *Science* **299,** 902-906.

Latvala-Kilby, S.M., y Kilby, N.J. (2006). Uncovering the post-embryonic role of embryo essential genes in *Arabidopsis* using the controlled induction of visibly marked genetic mosaics: *EMB506*, an illustration. *Plant Molecular Biology* **61,** 179-194.

Lea, P.J., Sodek, L., Parry, M.A., Shewry, P.R., y Halford, N.G. (2007). Asparagine in plants. *Annals of Applied Biology* **150,** 1-26.

Lee, J.Y., Lee, H.S., Song, J.Y., Jung, Y.J., Reinbothe, S., Park, Y.I., Lee, S.Y., y Pai, H.S. (2013). CELL GROWTH DEFECT FACTOR1/CHAPERONE-LIKE PROTEIN OF POR1 plays a role in stabilization of light-dependent protochlorophyllide oxidoreductase in *Nicotiana benthamiana* and *Arabidopsis*. *Plant Cell* **25,** 3944-3960.

Leister, D. (2003). Chloroplast research in the genomic age. *Trends in Genetics* **19,** 47-56.

Leister, D. (2012). Retrograde signaling in plants: from simple to complex scenarios. *Frontiers in Plant Science* **3,** 135.

Liere, K., Weihe, A., y Börner, T. (2011). The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *Journal of Plant Physiology* **168,** 1345-1360.

Linnaeus, C. (1751). *Philosophia botanica*. G. Kiesewetter.

Lister, R., Gregory, B.D., y Ecker, J.R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current Opinion in Plant Biology* **12,** 107-118.

Liu, P., Jenkins, N.A., y Copeland, N.G. (2002). Efficient Cre-*loxP*-induced mitotic recombination in mouse embryonic stem cells. *Nature Genetics* **30,** 66-72.

Lopez-Juez, E., y Pyke, K.A. (2004). Plastids unleashed: their development and their integration in plant development. *International Journal of Developmental Biology* **49,** 557-577.

Lu, Y., Hall, D.A., y Last, R.L. (2011). A small zinc finger thylakoid protein plays a role in maintenance of photosystem II in *Arabidopsis thaliana. Plant Cell* **23,** 1861-1875.

Lloyd, J., y Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant Physiology* **158,** 1115-1129.

Marchi, L., Polverini, E., Degola, F., Baruffini, E., y Restivo, F.M. (2014). Glutamate dehydrogenase isoenzyme 3 (GDH3) of *Arabidopsis thaliana* is less thermostable than GDH1 and GDH2 isoenzymes. *Plant Physiology and Biochemistry* **83,** 225-231.

Mateo-Bonmatí, E., Casanova-Sáez, R., Candela, H., y Micol, J.L. (2014). Rapid identification of *angulata* leaf mutations using next-generation sequencing. *Planta* **240,** 1113-1122.

McElver, J., Tzafrir, I., Aux, G., Rogers, R., Ashby, C., Smith, K., Thomas, C., Schetter, A., Zhou, Q., Cushman, M.A., Tossberg, J., Nickle, T., Levin, J.Z., Law, M., Meinke, D., y Patton, D. (2001). Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana. Genetics* **159,** 1751-1763.

McLeod, M., Craft, S., y Broach, J.R. (1986). Identification of the crossover site during FLP-mediated recombination in the *Saccharomyces cerevisiae* plasmid 2 microns circle. *Molecular and Cellular Biology* **6,** 3357-3367.

McNally, S., y Hirel, B. (1983). Glutamine synthetase isoforms in higher plants. *Plant Physiology* **21,** 761-774.

Meacham, G.C., Lu, Z., King, S., Sorscher, E., Tousson, A., y Cyr, D.M. (1999). The Hdj-2/Hsc70 chaperone pair facilitates early steps in CFTR biogenesis. *EMBO Journal* **18,** 1492-1505.

Meinke, D., Muralla, R., Sweeney, C., y Dickerman, A. (2008). Identifying essential genes in *Arabidopsis thaliana. Trends in Plant Science* **13,** 483-491.

Meinke, D., Sweeney, C., y 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.

Meinke, D.W., y Sussex, I.M. (1979a). Isolation and characterization of six embryo-lethal mutants of *Arabidopsis thaliana. Developmental Biology* **72,** 62-72.

Meinke, D.W., y Sussex, I.M. (1979b). Embryo-lethal mutants of *Arabidopsis thaliana*. A model system for genetic analysis of plant embryo development. *Developmental Biology* **72,** 50-61.

Melonek, J., Matros, A., Trosch, M., Mock, H.P., y Krupinska, K. (2012). The core of chloroplast nucleoids contains architectural SWIB domain proteins. *Plant Cell* **24,** 3060-3073.

Metzker, M.L. (2010). Sequencing technologies – the next generation. *Nature Reviews Genetics* **11,** 31-46.

Meyerowitz, E.M. (1987). *Arabidopsis thaliana. Annual Review of Genetics* **21,** 93-111.

Meyerowitz, E.M. (2001). Prehistory and history of Arabidopsis research. *Plant Physiology* **125,** 15-19.

Micol, J.L., y Hake, S. (2003). The development of plant leaves. *Plant Physiology* **131,** 389-394.

Micol, J.L. (2009). Leaf development: time to turn over a new leaf? *Current Opinion in Plant Biology* **12,** 9-16.

Miernyk, J.A. (2001). The J-domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones. *Cell Stress and Chaperones* **6,** 209-218.

Miesak, B.H., y Coruzzi, G.M. (2002). Molecular and physiological analysis of Arabidopsis mutants defective in cytosolic or chloroplastic aspartate aminotransferase. *Plant Physiology* **129,** 650-660.

Miflin, B.J., y Lea, P.J. (1976). The pathway of nitrogen assimilation in plants. *Phytochemistry* **15,** 873-885.

Miséra, S., Müller, A.J., Weiland-Heidecker, U., y Jürgens, G. (1994). The *FUSCA* genes of *Arabidopsis*: negative regulators of light responses. *Molecular and General Genetics* **244,** 242-252.

Miyashita, Y., y Good, A.G. (2008). Glutamate deamination by glutamate dehydrogenase plays a central role in amino acid catabolism in plants. *Plant Signaling and Behavior* **3,** 842-843.

Mochizuki, N., Brusslan, J.A., Larkin, R., Nagatani, A., y Chory, J. (2001). Arabidopsis *genomes uncoupled 5* (*GUN5*) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proceedings of the National Academy of Sciences USA* **98,** 2053-2058.

Mollá-Morales, A., Sarmiento-Mañús, R., Robles, P., Quesada, V., Pérez-Pérez, J.M., González-Bayón, R., Hannah, M.A., Willmitzer, L., Ponce, M.R., y Micol, J.L. (2011). Analysis of *ven3* and *ven6* reticulate mutants reveals the importance of arginine biosynthesis in Arabidopsis leaf development. *Plant Journal* **65,** 335-345.

Morgan, T.H., Bridges, C.B., y Sturtevant, A.H. (1919). *The origin of gynandromorphs*. Carnegie Institution of Washington.

Müller, A.J. (1963). Embryonentest zum Nachweis rezessiver Letalfaktoren bei *Arabidopsis thaliana. Biologisches Zentralblatt* **82,** 133-163.

Muller, H.J., y Altenburg, E. (1930). The frequency of translocations produced by X-rays in *Drosophila. Genetics* **15,** 283-311.

Muralla, R., Sweeney, C., Stepansky, A., Leustek, T., y Meinke, D. (2007). Genetic dissection of histidine biosynthesis in Arabidopsis. *Plant Physiology* **144,** 890-903.

Muralla, R., Chen, E., Sweeney, C., Gray, J.A., Dickerman, A., Nikolau, B.J., y Meinke, D. (2008). A bifunctional locus (*BIO3-BIO1*) required for biotin biosynthesis in Arabidopsis. *Plant Physiology* **146,** 60-73.

Muranaka, A., Watanabe, S., Sakamoto, A., y Shimada, H. (2012). *Arabidopsis* cotyledon chloroplast biogenesis factor CYO1 uses glutathione as an electron donor and interacts with PSI (A1 and A2) and PSII (CP43 and CP47) subunits. *Journal of Plant Physiology* **169,** 1212-1215.

Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M., y Shinozaki, K. (2008). A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell* **20,** 3148-3162.

Nüsslein-Volhard, C., y Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila. Nature* **287,** 795-801.

Ostergaard, L., y Yanofsky, M.F. (2004). Establishing gene function by mutagenesis in *Arabidopsis thaliana. Plant Journal* **39,** 682-696.

Page, D.R., y Grossniklaus, U. (2002). The art and design of genetic screens: *Arabidopsis thaliana. Nature Reviews: Genetics* **3,** 124-136.

Pagnussat, G.C., Yu, H.J., Ngo, Q.A., Rajani, S., Mayalagu, S., Johnson, C.S., Capron, A., Xie, L.F., Ye, D., y Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis. Development* **132,** 603-614.

Patterson, J.T. (1930). Somatic segregation produced by X-rays in *Drosophila melanogaster. Proceedings of the National Academy of Sciences USA* **16,** 109-111.

Patton, D.A., Schetter, A.L., Franzmann, L.H., Nelson, K., Ward, E.R., y Meinke, D.W. (1998). An embryo-defective mutant of Arabidopsis disrupted in the final step of biotin synthesis. *Plant Physiology* **116,** 935-946.

Pérez-Pérez, J.M., Candela, H., Robles, P., Quesada, V., Ponce, M.R., y Micol, J.L. (2009). Lessons from a search for leaf mutants in *Arabidopsis thaliana. International Journal of Developmental Biology* **53,** 1623-1634.

Pérez-Pérez, J.M., Rubio-Díaz, S., Dhondt, S., Hernández-Romero, D., Sánchez-Soriano, J., Beemster, G.T., Ponce, M.R., y Micol, J.L. (2011). Whole organ, venation and epidermal cell morphological variations are correlated in the leaves of *Arabidopsis* mutants. *Plant, Cell and Environment* **34,** 2200-2211.

Peterman, T.K., y Goodman, H.M. (1991). The glutamine synthetase gene family of *Arabidopsis thaliana*: light-regulation and differential expression in leaves, roots and seeds. *Molecular and General Genetics* **230,** 145-154.

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.J., y Oelmüller, R. (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* **18,** 176-197.

Pfalz, J., y Pfannschmidt, T. (2013). Essential nucleoid proteins in early chloroplast development. *Trends in Plant Science* **18,** 186-194.

Pfannschmidt, T., Blanvillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., Grubler, B., Hommel, E., y Lerbs-Mache, S. (2015). Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. *Journal of Experimental Botany* **66,** 6957-6973.

Poethig, R.S. (1987). Clonal analysis of cell lineage patterns in plant development. *American Journal of Botany* **74,** 581-594.

Poethig, S. (1989). Genetic mosaics and cell lineage analysis in plants. *Trends in Genetics* **5,** 273-277.

Pulido, P., Toledo-Ortiz, G., Phillips, M.A., Wright, L.P., y Rodríguez-Concepción, M. (2013). *Arabidopsis* J-protein J20 delivers the first enzyme of the plastidial isoprenoid pathway to protein quality control. *Plant Cell* **25,** 4183-4194.

Pyke, K., Zubko, M.K., y Day, A. (2000). Marking cell layers with spectinomycin provides a new tool for monitoring cell fate during leaf development. *Journal of Experimental Botany* **51,** 1713-1720.

Qian, Y.Q., Patel, D., Hartl, F.U., y McColl, D.J. (1996). Nuclear magnetic resonance solution structure of the human Hsp40 (HDJ-1) J-domain. *Journal of Molecular Biology* **260,** 224-235.

Qiao, J., Li, J., Chu, W., y Luo, M. (2013). PRDA1, a novel chloroplast nucleoid protein, is required for early chloroplast development and is involved in the regulation of plastid gene expression in Arabidopsis. *Plant and Cell Physiology* **54,** 2071-2084.

Qiu, X.B., Shao, Y.M., Miao, S., y Wang, L. (2006). The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and Molecular Life Sciences* **63,** 2560-2570.

Rajan, V.B., y D'Silva, P. (2009). *Arabidopsis thaliana* J-class heat shock proteins: cellular stress sensors. *Functional and Integrative Genomics* **9,** 433-446.

Resino, J., Salama-Cohen, P., y García-Bellido, A. (2002). Determining the role of patterned cell proliferation in the shape and size of the *Drosophila* wing. *Proceedings of the National Academy of Sciences USA* **99,** 7502-7507.

Reumann, S., y Keegstra, K. (1999). The endosymbiotic origin of the protein import machinery of chloroplastic envelope membranes. *Trends in Plant Science* **4,** 302-307.

Robison, M.M., Ling, X., Smid, M.P., Zarei, A., y Wolyn, D.J. (2009). Antisense expression of mitochondrial ATP synthase subunits OSCP (ATP5) and gamma (ATP3) alters leaf morphology, metabolism and gene expression in Arabidopsis. *Plant and Cell Physiology* **50,** 1840-1850.

Robles, P. (1999). *Análisis genético de mutantes de Arabidopsis thaliana con alteraciones en la morfología de la hoja*. Tesis doctoral. Universidad Miguel Hernández de Elche.

Robles, P., y Micol, J.L. (2001). Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. *Molecular Genetics and Genomics* **266,** 12-19.

Rochaix, J.-D. (1997). Chloroplast reverse genetics: new insights into the function of plastid genes. *Trends in Plant Science* **2,** 419-425.

Rodermel, S. (2001). Pathways of plastid-to-nucleus signaling. *Trends in Plant Science* **6,** 471-478.

Rubin, G.M. (1989). Development of the Drosophila retina: inductive events studied at single cell resolution. *Cell* **57,** 519-520.

Ruckle, M.E., y Larkin, R.M. (2009). Plastid signals that affect photomorphogenesis in *Arabidopsis thaliana* are dependent on GENOMES UNCOUPLED 1 and cryptochrome 1. *New Phytologist* **182,** 367-379.

Saito, H., y Uchida, H. (1978). Organization and expression of the *dnaJ* and *dnaK* genes of *Escherichia coli* K12. *Molecular and General Genetics* **164,** 1-8.

Sander, J.D., y Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology* **32,** 347-355.

Santis-Maciossek, D., Kofer, W., Bock, A., Schoch, S., Maier, R.M., Wanner, G., Rüdiger, W., Koop, H.U., y Herrmann, R.G. (1999). Targeted disruption of the plastid RNA polymerase genes *rpoA*, *B* and *C1*: molecular biology, biochemistry and ultrastructure. *Plant Journal* **18,** 477-489.

Sarkar, N.K., Thapar, U., Kundnani, P., Panwar, P., y Grover, A. (2013). Functional relevance of J-protein family of rice (*Oryza sativa*). *Cell Stress and Chaperones* **18,** 321-331.

Sato, N., Terasawa, K., Miyajima, K., y Kabeya, Y. (2003). Organization, developmental dynamics, and evolution of plastid nucleoids. *International Review of Cytology* **232,** 217-262.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., y Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana. DNA Research* **6,** 283-290.

Scheres, B., Wolkenfelt, H., Willemsen, V., Terlouw, M., Lawson, E., Dean, C., y Weisbeek, P. (1994). Embryonic origin of the *Arabidopsis* primary root and root meristem initials. *Development* **120,** 2475-2487.

Scheres, B., y Wolkenfelt, H. (1998). The *Arabidopsis* root as a model to study plant development. *Plant Physiology and Biochemistry* **36,** 21-32.

Schlicher, T., y Soll, J. (1997). Chloroplastic isoforms of DnaJ and GrpE in pea. *Plant Molecular Biology* **33,** 181-185.

Schröter, Y., Steiner, S., Matthai, K., y Pfannschmidt, T. (2010). Analysis of oligomeric protein complexes in the chloroplast sub-proteome of nucleic acid-binding proteins from mustard reveals potential redox regulators of plastid gene expression. *Proteomics* **10,** 2191-2204.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N., y Weigel, D. (2006). Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell* **18,** 1121-1133.

Serralbo, O., Pérez-Pérez, J.M., Heidstra, R., y Scheres, B. (2006). Non-cell-autonomous rescue of anaphase-promoting complex function revealed by mosaic analysis of *HOBBIT*, an *Arabidopsis CDC27* homolog. *Proceedings of the National Academy of Sciences USA* **103,** 13250-13255.

Serrano-Cartagena, J., Robles, P., Ponce, M.R., y Micol, J.L. (1999). Genetic analysis of leaf form mutants from the *Arabidopsis* Information Service collection. *Molecular and General Genetics* **261,** 725-739.

Shimada, H., Mochizuki, M., Ogura, K., Froehlich, J.E., Osteryoung, K.W., Shirano, Y., Shibata, D., Masuda, S., Mori, K., y Takamiya, K. (2007). *Arabidopsis* cotyledon-specific chloroplast biogenesis factor CYO1 is a protein disulfide isomerase. *Plant Cell* **19,** 3157-3169.

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.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., y Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO Journal* **5,** 2043-2049.

Sieburth, L.E., Drews, G.N., y Meyerowitz, E.M. (1998). Non-autonomy of *AGAMOUS* function in flower development: use of a Cre/*loxP* method for mosaic analysis in *Arabidopsis*. *Development* **125,** 4303-4312.

Siegal, M.L., y Hartl, D.L. (1996). Transgene coplacement and high efficiency site-specific recombination with the Cre/*loxP* system in Drosophila. *Genetics* **144,** 715-726.

Singh, R., Singh, S., Parihar, P., Singh, V.P., y Prasad, S.M. (2015). Retrograde signaling between plastid and nucleus: A review. *Journal of Plant Physiology* **181,** 55-66.

Somerville, C., y Koornneef, M. (2002). A fortunate choice: the history of *Arabidopsis* as a model plant. *Nature reviews: Genetics* **3,** 883-889.

Somerville, C.R., y Ogren, W.L. (1980). Inhibition of photosynthesis in *Arabidopsis* mutants lacking leaf glutamate synthase activity. *Nature* **286,** 257-259.

Stadler, L.J. (1930). Some genetic effects of X-rays in plants. *Journal of Heredity* **21,** 3-20.

Steiner, S., Schröter, Y., Pfalz, J., y Pfannschmidt, T. (2011). Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. *Plant Physiology* **157,** 1043-1055.

Stern, C. (1936). Somatic crossing over and segregation in *Drosophila melanogaster. Genetics* **21,** 625-730.

Stirnberg, P., Furner, I.J., y Ottoline Leyser, H.M. (2007). MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. *Plant Journal* **50,** 80-94.

Strand, A., Asami, T., Alonso, J., Ecker, J.R., y Chory, J. (2003). Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrin IX. *Nature* **421,** 79-83.

Sturtevant, A.H. (1920). The *vermilion* gene and gynandromorphism. *Experimental Biology and Medicine* **17,** 70-71.

Sturtevant, A.H. (1932). The use of mosaics in the study of the developmental effects of genes, en *VI International Congress of Genetics*.

Sun, X., Feng, P., Xu, X., Guo, H., Ma, J., Chi, W., Lin, R., Lu, C., y Zhang, L. (2011). A chloroplast envelope-bound PHD transcription factor mediates chloroplast signals to the nucleus. *Nature Communications* **2,** 477.

Susek, R.E., Ausubel, F.M., y Chory, J. (1993). Signal transduction mutants of Arabidopsis uncouple nuclear *CAB* and *RBCS* gene expression from chloroplast development. *Cell* **74,** 787-799.

Suzuki, A., y Rothstein, S. (1997). Structure and regulation of ferredoxin-dependent glutamase synthase from *Arabidopsis thaliana*. Cloning of cDNA expression in different tissues of wild-type and *gltS* mutant strains, and light induction. *European Journal of Biochemistry* **243,** 708-718.

Suzuki, A., y Knaff, D.B. (2005). Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism. *Photosynthesis Research* **83,** 191-217.

Szabo, A., Korszun, R., Hartl, F.U., y Flanagan, J. (1996). A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO Journal* **15,** 408-417.

Tadini, L., Pesaresi, P., Kleine, T., Rossi, F., Guljamow, A., Sommer, F., Mühlhaus, T., Schroda, M., Masiero, S., Pribil, M., Rothbart, M., Hedtke, B., Grimm, B., y Leister, D. (2016). GUN1 controls accumulation of the plastid ribosomal protein S1 at the protein level and interacts with proteins involved in plastid protein homeostasis. *Plant Physiology* **170,** 1817-1830.

Taira, M., Valtersson, U., Burkhardt, B., y Ludwig, R.A. (2004). *Arabidopsis thaliana GLN2*-encoded glutamine synthetase is dual targeted to leaf mitochondria and chloroplasts. *Plant Cell* **16,** 2048-2058.

Tameshige, T., Fujita, H., Watanabe, K., Toyokura, K., Kondo, M., Tatematsu, K., Matsumoto, N., Tsugeki, R., Kawaguchi, M., Nishimura, M., y Okada, K. (2013). Pattern dynamics in adaxial-abaxial specific gene expression are modulated by a plastid retrograde signal during *Arabidopsis thaliana* leaf development. *PLOS Genetics* **9,** e1003655.

Tanz, S.K., Kilian, J., Johnsson, C., Apel, K., Small, I., Harter, K., Wanke, D., Pogson, B., y Albrecht, V. (2012). The SCO2 protein disulphide isomerase is required for thylakoid biogenesis

and interacts with LHCB1 chlorophyll a/b binding proteins which affects chlorophyll biosynthesis in Arabidopsis seedlings. *Plant Journal* **69,** 743-754.

Terry, M.J., y Smith, A.G. (2013). A model for tetrapyrrole synthesis as the primary mechanism for plastid-to-nucleus signaling during chloroplast biogenesis. *Frontiers in Plant Science* **4,** 14.

Thomsen, H.C., Eriksson, D., Møller, I.S., y Schjoerring, J.K. (2014). Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency? *Trends in Plant Science* **19,** 656-663.

Thum, K.E., Shasha, D.E., Lejay, L.V., y Coruzzi, G.M. (2003). Light- and carbon-signaling pathways. Modeling circuits of interactions. *Plant Physiology* **132,** 440-452.

Tominaga, J., Mizutani, H., Horikawa, D., Nakahara, Y., Takami, T., Sakamoto, W., Sakamoto, A., y Shimada, H. (2016). Rice CYO1, an ortholog of *Arabidopsis thaliana* cotyledon chloroplast biogenesis factor AtCYO1, is expressed in leaves and involved in photosynthetic performance. *Journal of Plant Physiology* **207,** 78-83.

Turner, S., y Sieburth, L.E. (2002). Vascular patterning, en *The Arabidopsis Book*. American Society of Plant Biologists.

Tzafrir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., Patton, D., y Meinke, D. (2003). The Arabidopsis SeedGenes Project. *Nucleic Acids Research* **31,** 90-93.

Tzafrir, I., Pena-Muralla, R., Dickerman, A., Berg, M., Rogers, R., Hutchens, S., Sweeney, T.C., McElver, J., Aux, G., Patton, D., y Meinke, D. (2004). Identification of genes required for embryo development in Arabidopsis. *Plant Physiology* **135,** 1206-1220.

van Dijken, A.J., Schluepmann, H., y Smeekens, S.C. (2004). Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. *Plant Physiology* **135,** 969-977.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz,

R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., y Zhu, X. (2001). The sequence of the human genome. *Science* **291,** 1304-1351.

Vinti, G., Hills, A., Campbell, S., Bowyer, J.R., Mochizuki, N., Chory, J., y Lopez-Juez, E. (2000). Interactions between *hy1* and *gun* mutants of *Arabidopsis*, and their implications for plastid/nuclear signalling. *Plant Journal* **24,** 883-894.

Vitha, S., Froehlich, J.E., Koksharova, O., Pyke, K.A., van Erp, H., y Osteryoung, K.W. (2003). ARC6 is a J-domain plastid division protein and an evolutionary descendant of the cyanobacterial cell division protein Ftn2. *Plant Cell* **15,** 1918-1933.

Voll, L., Hausler, R.E., Hecker, R., Weber, A., Weissenbock, G., Fiene, G., Waffenschmidt, S., y Flugge, U.I. (2003). The phenotype of the Arabidopsis *cue1* mutant is not simply caused by a general restriction of the shikimate pathway. *Plant Journal* **36,** 301-317.

Von Goethe, J.W. (1816). *Italian Journey: 1786-1788*. Penguin Classics.

Wachsman, G., Heidstra, R., y Scheres, B. (2011). Distinct cell-autonomous functions of *RETINOBLASTOMA-RELATED* in Arabidopsis stem cells revealed by the Brother of Brainbow clonal analysis system. *Plant Cell* **23,** 2581-2591.

Walsh, P., Bursac, D., Law, Y.C., Cyr, D., y Lithgow, T. (2004). The J-protein family: modulating protein assembly, disassembly and translocation. *EMBO Reports* **5,** 567-571.

Wang, G., Cai, G., Kong, F., Deng, Y., Ma, N., y Meng, Q. (2014). Overexpression of tomato chloroplast-targeted DnaJ protein enhances tolerance to drought stress and resistance to *Pseudomonas solanacearum* in transgenic tobacco. *Plant Physiology and Biochemistry* **82,** 95-104.

Wang, G., Kong, F., Zhang, S., Meng, X., Wang, Y., y Meng, Q. (2015). A tomato chloroplast-targeted DnaJ protein protects Rubisco activity under heat stress. *Journal of Experimental Botany* **66,** 3027-3040.

Wang, Y.W., Chen, S.M., Wang, W.J., Huang, X.Q., Zhou, C.F., Zhuang, Z., y Lu, S. (2016). The DnaJ-like zinc finger domain protein PSA2 affects light acclimation and chloroplast development in *Arabidopsis thaliana. Frontiers in Plant Science* **7,** 360.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., y Langdale, J.A. (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. *Plant Cell* **21,** 1109-1128.

Weijers, D., Franke-van Dijk, M., Vencken, R.J., Quint, A., Hooykaas, P., y Offringa, R. (2001). An *Arabidopsis* Minute-like phenotype caused by a semi-dominant mutation in a *RIBOSOMAL PROTEIN S5* gene. *Development* **128,** 4289-4299.

Weil, C.F., y Wessler, S.R. (1993). Molecular evidence that chromosome breakage by *Ds* elements is caused by aberrant transposition. *Plant Cell* **5,** 515-522.

Wildwater, M., Campilho, A., Pérez-Pérez, J.M., Heidstra, R., Blilou, I., Korthout, H., Chatterjee, J., Mariconti, L., Gruissem, W., y Scheres, B. (2005). The *RETINOBLASTOMA-RELATED* gene regulates stem cell maintenance in *Arabidopsis* roots. *Cell* **123,** 1337-1349.

Wilkins, A.S. (1992). *Genetic analysis of animal development*. 2ª Edición. Wiley-Liss.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., y Provart, N.J. (2007). An "electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLOS ONE* **2,** e718.

Xu, T., y Rubin, G.M. (1993). Analysis of genetic mosaics in developing and adult Drosophila tissues. *Development* **117,** 1223-1237.

Yagi, Y., Ishizaki, Y., Nakahira, Y., Tozawa, Y., y Shiina, T. (2012). Eukaryotic-type plastid nucleoid protein pTAC3 is essential for transcription by the bacterial-type plastid RNA polymerase. *Proceedings of the National Academy of Sciences USA* **109,** 7541-7546.

Yagi, Y., y Shiina, T. (2014). Recent advances in the study of chloroplast gene expression and its evolution. *Frontiers in Plant Science* **5,** 61.

Yamamoto, H., Peng, L., Fukao, Y., y Shikanai, T. (2011). An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. *Plant Cell* **23,** 1480-1493.

Yin, C., Richter, U., Börner, T., y Weihe, A. (2010). Evolution of plant phage-type RNA polymerases: the genome of the basal angiosperm *Nuphar advena* encodes two mitochondrial and one plastid phage-type RNA polymerases. *BMC Evolutionary Biology* **10,** 379.

Yochem, J., Uchida, H., Sunshine, M., Saito, H., Georgopoulos, C.P., y Feiss, M. (1978). Genetic analysis of two genes, *dnaJ* and *dnaK*, necessary for *Escherichia coli* and bacteriophage lambda DNA replication. *Molecular and General Genetics* **164,** 9-14.

Yu, Q.B., Lu, Y., Ma, Q., Zhao, T.T., Huang, C., Zhao, H.F., Zhang, X.L., Lv, R.H., y Yang, Z.N. (2013). TAC7, an essential component of the plastid transcriptionally active chromosome complex, interacts with FLN1, TAC10, TAC12 and TAC14 to regulate chloroplast gene expression in *Arabidopsis thaliana. Physiologia Plantarum* **148,** 408-421.

Yu, Q.B., Huang, C., y Yang, Z.N. (2014). Nuclear-encoded factors associated with the chloroplast transcription machinery of higher plants. *Frontiers in Plant Science* **5,** 316.

Zagari, N., Sandoval-Ibañez, O., Sandal, N., Su, J., Rodríguez-Concepción, M., Stougaard, J., Pribil, M., Leister, D., y Pulido, P. (2017). *SNOWY COTYLEDON 2* promotes chloroplast development and has a role in leaf variegation in both *Lotus japonicus* and *Arabidopsis thaliana. Molecular Plant* **10,** 721-734.

Zhang, H., Scheirer, D.C., Fowle, W.H., y Goodman, H.M. (1992). Expression of antisense or sense RNA of an ankyrin repeat-containing gene blocks chloroplast differentiation in Arabidopsis. *Plant Cell* **4,** 1575-1588.

Zhao, L., y Sack, F.D. (1999). Ultrastructure of stomatal development in *Arabidopsis* (Brassicaceae) leaves. *American Journal of Botany* **86,** 929-939.

Zhong, L., Zhou, W., Wang, H., Ding, S., Lu, Q., Wen, X., Peng, L., Zhang, L., y Lu, C. (2013). Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in *Arabidopsis* under heat stress. *Plant Cell* **25,** 2925-2943.

Zhu, X., Liang, S., Yin, J., Yuan, C., Wang, J., Li, W., He, M., Wang, J., Chen, W., Ma, B., Wang, Y., Qin, P., Li, S., y Chen, X. (2015). The DnaJ OsDjA7/8 is essential for chloroplast development in rice (*Oryza sativa*). *Gene* **574,** 11-19.

UNIVERSITAS
Miguel
Hernández

V.- PUBLICACIONES

UNIVERSITAS
Miguel
Hernández

the plant journal

The Plant Journal (2016)

The *ANGULATA7* gene encodes a DnaJ-like zinc finger-domain protein involved in chloroplast function and leaf development in Arabidopsis

Tamara Muñoz-Nortes, José Manuel Pérez-Pérez, María Rosa Ponce, Héctor Candela and José Luis Micol*
Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, Elche 03202, Spain

Received 27 April 2016; revised 16 December 2016; accepted 19 December 2016.
*For correspondence (e-mail jlmicol@umh.es).

SUMMARY

The characterization of mutants with altered leaf shape and pigmentation has previously allowed the identification of nuclear genes that encode plastid-localized proteins that perform essential functions in leaf growth and development. A large-scale screen previously allowed us to isolate ethyl methanesulfonate-induced mutants with small rosettes and pale green leaves with prominent marginal teeth, which were assigned to a phenotypic class that we dubbed Angulata. The molecular characterization of the 12 genes assigned to this phenotypic class should help us to advance our understanding of the still poorly understood relationship between chloroplast biogenesis and leaf morphogenesis. In this article, we report the phenotypic and molecular characterization of the *angulata7-1* (*anu7-1*) mutant of *Arabidopsis thaliana*, which we found to be a hypomorphic allele of the *EMB2737* gene, which was previously known only for its embryonic-lethal mutations. *ANU7* encodes a plant-specific protein that contains a domain similar to the central cysteine-rich domain of DnaJ proteins. The observed genetic interaction of *anu7-1* with a loss-of-function allele of *GENOMES UNCOUPLED1* suggests that the *anu7-1* mutation triggers a retrograde signal that leads to changes in the expression of many genes that normally function in the chloroplasts. Many such genes are expressed at higher levels in *anu7-1* rosettes, with a significant overrepresentation of those required for the expression of plastid genome genes. Like in other mutants with altered expression of plastid-encoded genes, we found that *anu7-1* exhibits defects in the arrangement of thylakoidal membranes, which appear locally unappressed.

Keywords: ***Arabidopsis thaliana*, leaf development, chloroplast, DnaJ proteins, thylakoidal membranes.**

INTRODUCTION

The characterization of leaf shape mutants has allowed the identification of genes that control key events in the development and growth of leaves, from the recruitment of initial cells at the flank of the shoot apical meristem to the patterning of its constituent tissues along the abaxial-adaxial, proximal-distal and medial-lateral axes (Pérez-Pérez *et al.*, 2009; González and Inzé, 2015; Sluis and Hake, 2015). While some of these genes have been found to encode transcriptional or post-transcriptional regulators, a significant proportion of the mutations that affect the size and shape of plant leaves damage proteins that perform a variety of housekeeping functions, from enzymes that catalyze steps of biosynthetic pathways to subunits of the cytosolic ribosome (Pérez-Pérez *et al.*, 2009). In line with this idea, the systematic cloning of genes identified in our large-scale screen for ethyl methanesulfonate (EMS)-induced mutants with abnormal leaf size and/or shape has evidenced that not few of such genes code for enzymes that catalyze steps in the biosynthesis pathways of arginine (Mollá-Morales *et al.*, 2011), cellulose (Rubio-Díaz *et al.*, 2012) or tetrapyrroles (Quesada *et al.*, 2013). The two most abundant phenotypic classes identified in this screen comprise mutants carrying alleles of nuclear genes that encode subunits of the cytosolic ribosome (Van Minnebruggen *et al.*, 2010; Horiguchi *et al.*, 2011; Casanova-Sáez *et al.*, 2014a) and chloroplast-localized proteins, respectively (Casanova-Sáez *et al.*, 2014b; Mateo-Bonmatí *et al.*, 2014, 2015). Remarkably, many of the mutations in genes that appear to be housekeeping do not cause a mere growth retardation, but instead they cause specific phenotypes, as other authors have previously reported (Tsukaya *et al.*, 2013).

We recently reported the cloning of five genes whose mutant alleles cause phenotypes that we assigned to one

such classes, the Angulata (Anu) phenotypic class, which is defined by mutant alleles of 12 different genes, named *ANU1* to *ANU12* (Berná *et al.*, 1999; Pérez-Pérez *et al.*, 2009). Mutations in all these genes cause a pleiotropic phenotype that includes both defects in leaf size and shape and pale green pigmentation. The genes cloned so far include *ANU1*, *ANU4*, *ANU9*, *ANU10* and *ANU12* (Casanova-Sáez *et al.*, 2014b; Mateo-Bonmatí *et al.*, 2014), which in most cases encode proteins that contain an N-terminal chloroplast transit peptide. We found our *anu1*, *anu4*, *anu9* and *anu12* mutants to be loss-of-function, viable alleles of the *SECA2*, *TRANSLOCON AT THE OUTER MEMBRANE OF CHLOROPLASTS 33* (*TOC33*), *NON-INTRINSIC ABC PROTEIN 14* (*NAP14*) and *CLP PROTEASE PROTEOLYTIC SUBUNIT 1* (*CLPR1*) genes (Jarvis *et al.*, 1998; Koussevitzky *et al.*, 2007; Shimoni-Shor *et al.*, 2010; Skalitzky *et al.*, 2011). The *ANU10* gene, by contrast, was found to encode a conserved, previously undescribed protein that is widespread throughout the plant kingdom and is required for thylakoid stacking in chloroplasts (Casanova-Sáez *et al.*, 2014b), demonstrating that mutational approaches remain a powerful tool to identify new molecular components even for processes that have been intensively surveyed over the last decades, such as photosynthetic function.

Here we report the cloning of an additional gene in this phenotypic class, *ANU7*, which encodes a plant-specific protein containing a domain with conserved cysteine and glycine residues that is similar to an incomplete central cysteine-rich (CR) domain, which accounts for the disulfide isomerase activity of DnaJ proteins. DnaJ proteins normally function as chaperones, either alone or in combination with heat shock protein 70 (Hsp70), and have been proposed to participate in the folding, unfolding, assembly and degradation of proteins, maintaining protein homeostasis under normal or stress conditions (Hennessy *et al.*, 2005; Craig *et al.*, 2006; Liu *et al.*, 2007). Although the complete loss of *ANU7* function causes embryonic lethality, our EMS-induced allele is hypomorphic and viable, and has allowed us to demonstrate that *ANU7* is required for the accumulation of photosynthetic pigments and the correct organization of the thylakoid membrane system.

RESULTS

Phenotype of the *anu7-1* mutant

In order to further our understanding of the basis of the Angulata phenotype, we have characterized the *angulata7-1* (*anu7-1*) mutant, which was isolated in a large-scale screen for EMS-induced mutants with abnormal leaf size, shape or pigmentation (Berná *et al.*, 1999). The *anu7-1* allele is recessive and causes a phenotype similar to that caused by mutant alleles of other *ANU* genes, including pale green leaves with prominent marginal teeth and reduced leaf growth (Figure 1a–h). In line with this observation, the projected area of the basal rosettes of *anu7-1* plants was significantly reduced ($P < 0.001$) relative to the wild-type parental line, the Landsberg *erecta* (L*er*) accession (Figure 1p). This reduction became apparent soon after germination, persisted throughout the lifespan of the plants, and was accompanied by reductions in fresh and dry weights that were statistically significant 14 days after stratification (das; Figures 1q and S1). The *anu7-1* mutant seemed to recover ~4 weeks after stratification, when no significant difference in fresh and dry weight with L*er* was detected (Figures 1o and S1). A severe reduction in stem length was observed, despite no differences in flowering time were detected between mutant and wild-type plants (Figures S1 and S2).

The pale green phenotype of *anu7-1* leaves was found to be linked to a significant reduction in the levels of chlorophylls *a* and *b*, as well as in the levels of carotenoids (Figure 2). The presence of hydrogen peroxide, a reactive oxygen species (ROS), was tested by incubating the leaves with 3,3′-diaminobenzidine (DAB), but a clear difference between mutant and wild-type plants was not observed (Figure S3).

Histological characterization of the *anu7* mutant

We examined the developmental defects of the *anu7* mutant using light and electron microscopy. To this end, we initially cleared rosette leaves from the first and third nodes of *anu7-1* mutant plants using chloral hydrate and examined their internal tissues using differential interference contrast (DIC) microscopy, which uncovered a vascular network with apparently normal topology. An irregular, wider distribution of palisade mesophyll cell sizes, accompanied by increased airspaces, was observed in paradermal and transverse sections of *anu7-1* mutant leaves (Figure 3). This observation links the *anu7-1* mutant to other mutants in the Angulata phenotypic class, such as the *anu10* mutants, which also display an altered distribution of cell sizes in the palisade mesophyll (Casanova-Sáez *et al.*, 2014b).

The ultrastructure of the chloroplasts in mesophyll cells of plants grown under continuous light was studied using transmission electron microscopy (TEM; Figure 4). In the mutant, the chloroplasts contained large starch granules (similar to those of the wild-type) and abundant thylakoid membranes, which seemed to fill a larger fraction of their internal space. In general, both the mutant and the wild-type chloroplasts contained stacks of tightly appressed thylakoidal membranes (i.e. grana). However, these membranes occasionally appeared unappressed in the *anu7-1* chloroplasts, resulting in a locally increased luminal space between the thylakoidal membranes (white arrowheads in Figure 4f).

Positional cloning of the *ANU7* gene

To gain insight into the molecular processes that are disrupted in the *anu7-1* mutant, we followed a map-based

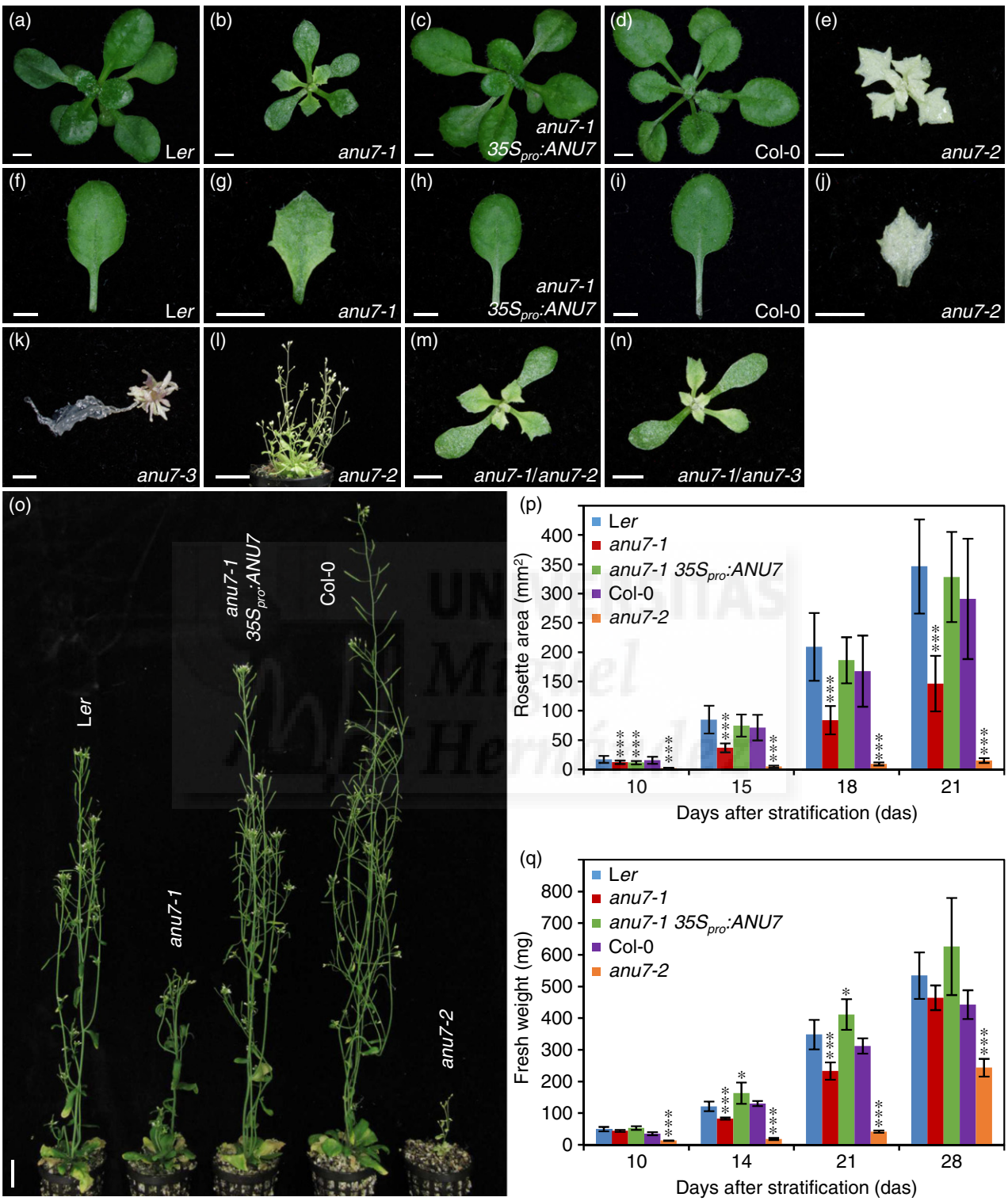


Figure 1. Phenotypes of homozygous and heterozygous plants for *anu7* mutant alleles and phenotypic rescue of *anu7-1.*
(a–e, k, m, n) Rosettes, (f–j) third-node leaves, and (l, o) adult plants. Pictures were taken (a–c, f–h, m, n) 16, (d, e, i, j) 21, (k) 34, (o) 42, and (l) 60 days after stratification (das). Scale bars indicate (a–k, m, n) 2 mm, and (l, o) 2 cm.
(p, q) Quantification of phenotypic traits of *anu7* mutant alleles: (p) Rosette area and (q) fresh weight of L*er*, *anu7-1*, *anu7-1 35S$_{pro}$:ANU7*, Col-0 and *anu7-2* plants. Error bars indicate standard deviations. Asterisks indicate values significantly different from the corresponding wild-type in (p) Student's *t*-test (***$P < 0.001$, $n = 20$–30), and (q) Mann–Whitney *U*-test (*$P < 0.05$, ***$P < 0.001$, $n = 8$).

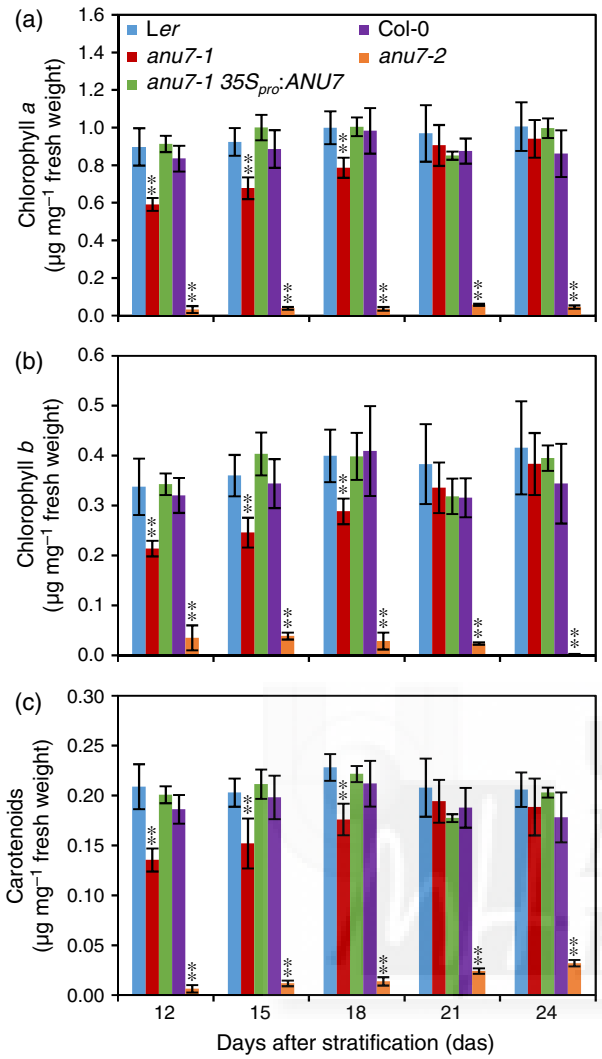


Figure 2. Pigment concentration in *anu7* mutants.
(a–c) (a) Chlorophyll *a*, (b) chlorophyll *b*, and (c) carotenoids content in L*er*, *anu7-1*, *anu7-1 35S$_{pro}$:ANU7*, Col-0, and *anu7-2* above-ground tissues. Plants were collected 12, 15, 18, 21 and 24 days after stratification (das). Error bars indicate standard deviations. Asterisks indicate values significantly different from the corresponding wild-type in a Mann–Whitney *U*-test (**$P < 0.01$, $n = 6$).

approach to identify the *ANU7* gene (Figure 5). We first mapped the gene to chromosome 5 using microsatellite and single nucleotide polymorphism (SNP) markers (Figure 5a; Ponce *et al.*, 2006; Robles and Micol, 2001). The gene was subsequently mapped to a shorter interval delimited by the At5g53700 and At5g53902 genes, which encompassed 26 genes (Figure 5b). Sequencing of candidate genes within this interval allowed us to identify a single G-to-A transition mutation in the coding sequence (CDS) of the At5g53860 gene (Figure 5c). This gene, also known as *EMBRYO DEFECTIVE 2737* (*EMB2737*), is present in a single copy in the genome of Arabidopsis. Two additional, embryo-lethal alleles have been described, *emb2737-1* (*anu7-2*) and *emb2737-2* (*anu7-3*), which arrest the embryogenesis at the cotyledon stage (Tzafrir *et al.*, 2003).

To demonstrate that *ANU7* is the same gene as *EMB2737*, we performed complementation crosses between *anu7-1* and plants heterozygous for each of the embryo-lethal alleles, *emb2737-1* and *emb2737-2*. The F_1 progenies of these crosses comprised phenotypically wild-type plants and plants similar to *anu7-1* homozygotes in a 1:1 ratio, as expected from crosses involving plants that were heterozygous for the lethal alleles (Figure 1m,n). The observed non-complementation indicates that the *anu7-1*, *emb2737-1* and *emb2737-2* all damage the same gene. The viability of the *anu7-1*/*emb2737-1* and *anu7-1*/*emb2737-2* heterozygotes suggests that *anu7-1* is a partial loss-of-function (hypomorphic) allele of the *ANU7* gene, while *emb2737-1* and *emb2737-2* might be null alleles. An additional confirmation for the correct identification of *ANU7* as the At5g53860 gene came from the complementation of the mutant phenotype of *anu7-1* with a *35S$_{pro}$:ANU7* transgene, which expresses a wild-type copy of the *ANU7* gene under the control of the 35S promoter. We isolated 12 independent transformants in the *anu7-1* background. Six of these transformants exhibited a completely wild-type phenotype (Figure 1c, h, o), and six more displayed an intermediate (partially complemented) phenotype. The phenotypic rescue was observed for all the studied traits, including rosette area (Figure 1p), fresh and dry weight (Figures 1q and S1), pigment content (Figure 2), and histology (Figures 3 and 4g–i). In addition, we also isolated four independent transformants carrying insertions of a *35S$_{pro}$:ANU7:GFP* translational fusion, which also complemented the mutant phenotype.

Data available from the Arabidopsis eFP Browser (Winter *et al.*, 2007; http://www.bar.utoronto.ca/) indicate that the *ANU7* gene is broadly expressed in both the root system and above-ground tissues. To characterize the *ANU7* expression pattern with higher resolution, we isolated transgenic lines expressing an *ANU7$_{pro}$:GUS* transgene. In young (7 das) L*er* seedlings, we detected GUS signal in the primary root, hypocotyl, cotyledons, shoot apical meristem and leaf primordia (Figure 6a). In cotyledons and leaves, the expression was most conspicuous at the venation pattern and hydathodes (Figure 6b, c), but GUS signal was also detected in mesophyll cells and the guard cells of stomata (Figure 6d, e). In roots, the expression was most intense at the root apex and the vascular cylinder (Figure 6a, f). Residual expression was detected in cauline leaves (Figure 6g). In immature flowers, the signal was highest in the pistil and petals (Figure 6h). In mature flowers, expression was detected at the receptacle, stigmata, anther filaments, and the venation of petals (Figure 6g–i). In siliques, we also observed GUS signal in the funiculi and the micropylar region of fertilized ovules and developing seeds (Figure 6j).

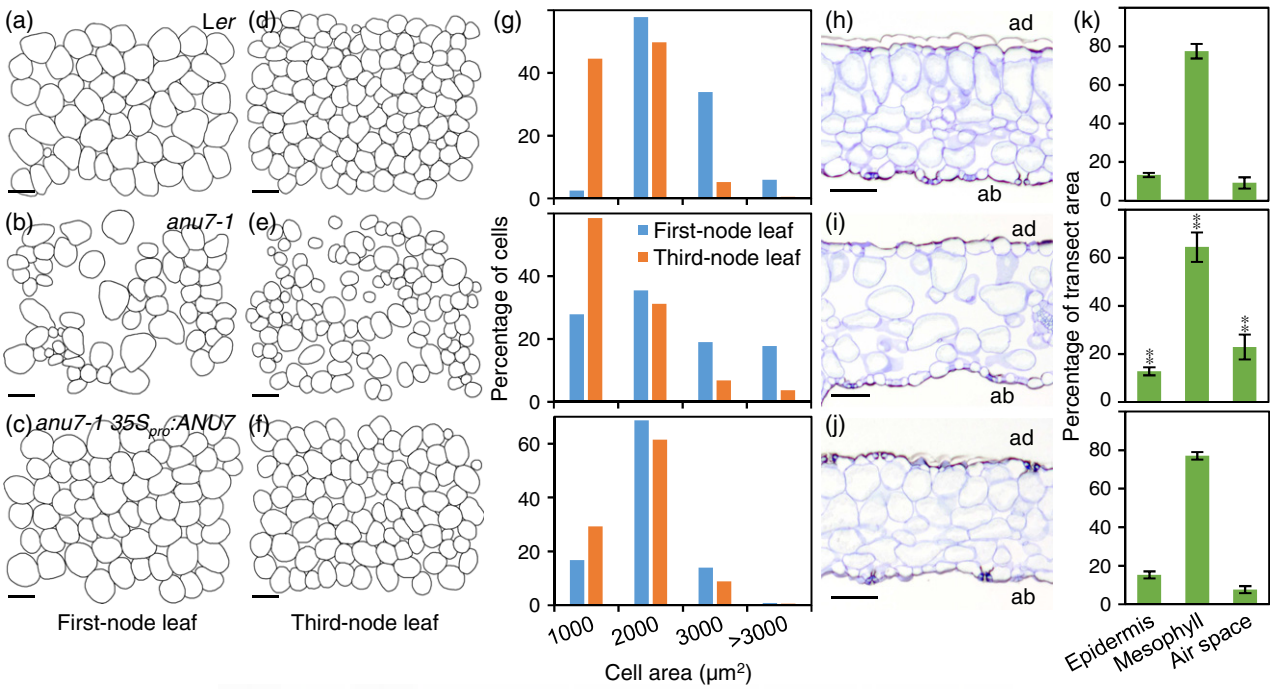


Figure 3. Morphometry of palisade mesophyll cells from L*er*, *anu7-1*, and *anu7-1 35S$_{pro}$:ANU7* plants.
(a–f) Representative diagrams of the subepidermal layer of palisade mesophyll cells from (a–c) first-node leaves and (d–f) third-node leaves.
(g) Distribution of palisade mesophyll cell sizes ($n \geq 600$) in first-node and third-node leaves.
(h–j) Third-node leaf transverse sections. ad: adaxial surface, ab: abaxial surface. Plants were collected 21 days after stratification (das). Scale bars indicate 50 μm. (k) Percentage of leaf transect area occupied by epidermis, mesophyll (including palisade and spongy mesophyll cells, and bundle sheath cells) or air space in third-node leaves. Asterisks indicate values significantly different from L*er* in a Mann–Whitney *U*-test (**$P < 0.01$, $n = 6$).

The *ANU7* gene encodes a plastid-localized protein of unknown function

The most recent annotation of the Arabidopsis genome (TAIR10) includes several splice forms for the *At5g53860* gene. The longest form (At5g53860.4) encodes a protein that consists of 459 amino acids and has a predicted molecular mass of 52.9 kDa. This isoform spans over a genomic region of 3.5 kb and comprises 13 exons, with the translation initiation codon (ATG) being located in exon 1 and the stop codon located in exon 13. The *anu7-1* mutation resides in exon 4, five nucleotides upstream of the donor site of intron 5, and causes a nonsynonymous Gly-to-Glu substitution at codon 173 of the protein (Figure 5). All the sequenced RT-PCR products spanning the site of the *anu7-1* mutation were correctly spliced. We also confirmed the presence of T-DNA insertions in the *emb2737-1* and *emb2737-2* mutants, which are located at the third and the next-to-last intron, respectively.

Other shorter transcripts have been described, which are generated by exon skipping (e.g. of exon 5 in At5g53860.1, of exon 8 in all but At5g53860.4) and/or by using alternative splice donor and/or acceptor sites (e.g. as in isoforms At5g53860.3 and At5g53860.5). In all cases, these alternative transcripts are predicted to encode shorter proteins with in-frame insertions and/or deletions. To evaluate the significance of these splice forms, we retrieved 87 expressed sequence tags (ESTs) that were available for this gene from the Unigene database (Unigene At.22001), and assembled them using CAP3 (Huang and Madan, 1999). Our assembly indicates that most of the available sequences (79 out of 87) can be assembled into a contig (Figure S4) that perfectly matches the sequence of the At5g53860.2 isoform (accession number NM_124765.4), which lacks exon 8 (relative to At5 g53860.4) and encodes a protein of 422 amino acids and a predicted molecular mass of 48.3 kDa, suggesting that this is the predominant isoform of the protein.

We carried out BLASTP and PSI-BLAST searches in order to identify the closest homologues from other species. In line with the result of our CAP3 assembly, we identified more than 50 highly similar sequences that best match the structure of the At5g53860.2 isoform. We found similar sequences from the moss *Physcomitrella patens* (XP_001783194), the lycophyte *Selaginella moellendorffii* (XP_002985760), and numerous flowering plants, including the basal angiosperm *Amborella trichopoda* (XP_006859011.1), many dicots (including the basal dicot *Nelumbo nucifera*, XP_010260241.1) and some monocot lineages (Figure 5d). Interestingly, our PSI-BLAST searches failed to identify hits to proteins from gymnosperms or the

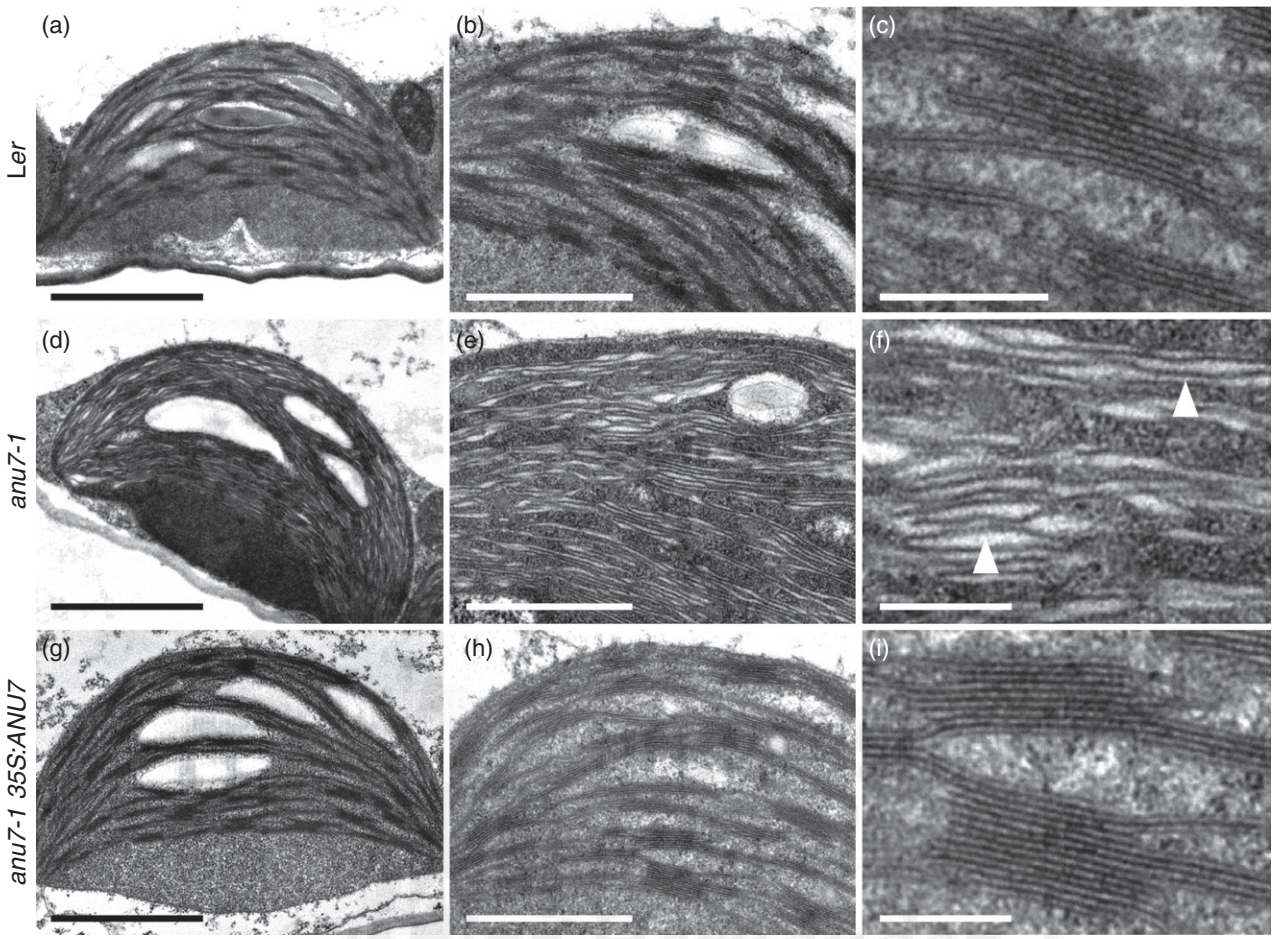


Figure 4. Ultrastructure of *anu7-1* chloroplasts.
(a–i) Transmission electron micrographs of palisade mesophyll cell chloroplasts from (a–c) L*er*, (d–f) *anu7-1*, and (g–i) *anu7-1 35S$_{pro}$:ANU7*. Pictures were taken from third-node leaves collected 21 days after stratification (das). Scale bars indicate (a, d, g) 2 μm, (b, e, h) 1 μm, and (c, f, i) 0.25 μm. Arrowheads in (f) indicate unappressed thylakoidal membranes in the *anu7-1* mutant.

grasses, as well as hits to proteins outside the plant kingdom, suggesting that ANU7 is the founding member of a plant-specific protein family that has been lost at least twice during the evolution of land plants (Figure S5 and Table S1).

All these proteins lack a known function and do not contain previously described domains, although a motif with partial similarity to the central cysteine-rich (CR) domain of DnaJ proteins is detected in some of them (PF00684, IPR001305). We performed a multiple sequence alignment using the amino acid sequences from ANU7 and other closely related proteins, which were selected from among the output of the BLASTP searches (Figure S5a). Different from DnaJ CR-type proteins, which contain four repeats of the C-x(2)-C-x-G-x-G motif, our alignment only uncovered two highly conserved C-x(2)-C-x-G-x-G repeats separated by 50 or 51 amino acid residues, which were immediately adjacent to two additional, incomplete C-x(2)-C-x-G or C-x(2)-C repeats. These repeats were present in all ANU7-like proteins. Previous authors have proposed that the C-x(2)-C-x-G-x-G-x(42,53)-C-x(2)-C-x-G-x-G motif can fold into a single zinc binding center (Hsu *et al.*, 2011), different from typical DnaJ CR-type proteins, which contain two zinc fingers. Interestingly, the third Gly residue in this motif is the one damaged by the *anu7-1* mutation, highlighting its functional importance (Figure 5d).

Despite the fact that the ANU7 protein has been detected in the chloroplast (plastid) proteome (Zybailov *et al.*, 2008; Huang *et al.*, 2013), several bioinformatics tools, including ChloroP (Emanuelsson *et al.*, 1999) and PCLR (Schein *et al.*, 2001), failed to identify an N-terminal chloroplast transit peptide (cTP) in the amino acid sequences of ANU7 and its homologues, suggesting that these proteins are non-canonical chloroplast proteins (Armbruster *et al.*, 2009). To confirm this subcellular localization pattern, we isolated transgenic plants expressing C-terminal fusions of the ANU7 protein to GFP, YFP and CFP, as well as an N-terminal fusion of ANU7 to GFP. Despite our *35S$_{pro}$:ANU7:*

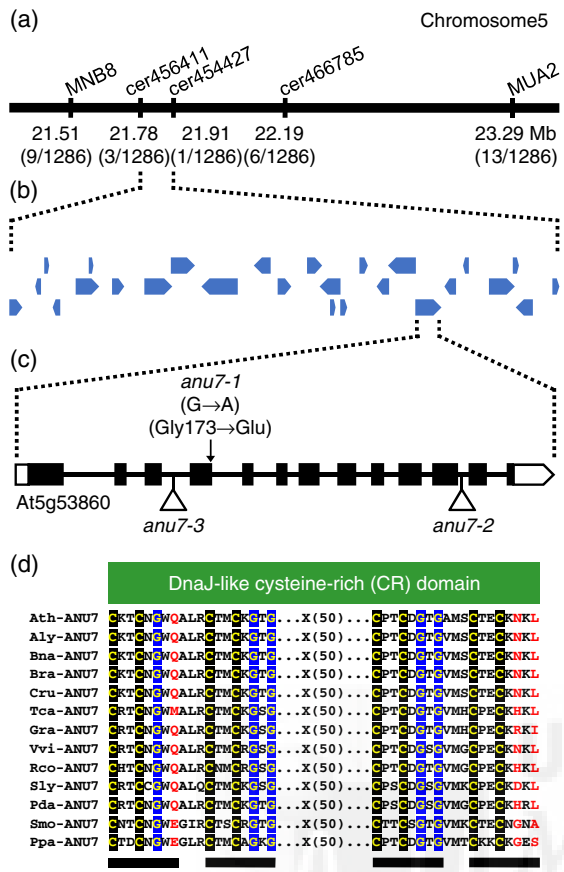


Figure 5. Positional cloning of the *ANU7* gene and conservation of the DnaJ-like cysteine-rich (CR) domain in the ANU7 protein and its orthologs from different species.

(a) A mapping population of 643 F_2 plants derived from an *anu7-1* × Col-0 cross allowed us to define a candidate region of 187.42 kb on chromosome 5. Names and physical map positions of the molecular markers used for linkage analysis are shown. The number of recombinant chromosomes found and the total of chromosomes analyzed are indicated in parentheses.

(b, c) (b) Genes present in the candidate interval of the *anu7-1* mutation (c) Structure of the *ANU7* gene, with indication of the nature and position of the *anu7* mutations. Boxes and lines between boxes indicate exons and introns, respectively. White boxes represent the 5′- and 3′-UTRs. Triangles indicate T-DNA insertions.

(d) Detail of an alignment of the ANU7 protein and some of its orthologs, showing the DnaJ-like CR domain. The alignment was generated with T-Coffee (see Experimental Procedures). The four zinc binding motifs of the DnaJ-like CR domain are underlined in black. Conserved cysteines and glycines are shaded black and blue, respectively. Plant species used for the multiple sequence alignment: Ath, *Arabidopsis thaliana* (NP_568801.2); Aly, *Arabidopsis lyrata* (XP_002864286.1); Bna, *Brassica napus* (CDY20217.1); Bra, *Brassica rapa* (XP_009119926.1); Cru, *Capsella rubella* (XP_006280533.1); Tca, *Theobroma cacao* (XP_007021511.1); Gra, *Gossypium raimondii* (XP_012473151.1); Vvi, *Vitis vinifera* (XP_002285056.1); Rco, *Ricinus communis* (XP_002528255.1); Sly, *Solanum lycopersicum* (XP_004251862.1); Pda, *Phoenix dactylifera* (XP_008794135.1); Smo, *Selaginella moellendorffii* (XP_002985760.1); Ppa, *Physcomitrella patens* (XP_00783194.1). The sequences for Smo and Ppa were completed using genomic sequences available from Phytozome (Goodstein *et al.*, 2012).

GFP constructs were able to complement the phenotype of *anu7-1* mutants, we were not able to detect fluorescence in the plastids using any of these constructs.

Plants homozygous for the embryo-lethal *emb2737-1* and *emb2737-2* alleles can grow postembryonically in the presence of sucrose

Families segregating the *emb2737-1* and *emb2737-2* mutations were sown in Murashige and Skoog medium containing 1% (m/v) sucrose. In these conditions, the *emb2737-1*/*emb2737-1* and *emb2737-2*/*emb2737-2* homozygotes were able to germinate, although at a low rate, and gave rise to albino plants with a growth phenotype that was more severe than those of the *anu7-1*/*anu7-1* homozygotes and the *anu7-1*/*emb2737-1* and *anu7-1*/*emb2737-2* heterozygotes (Figure 1m, n). In the presence of sucrose, the phenotype of the *emb2737-2* (*anu7-3*) mutant was more severe than that of the *emb2737-1* (*anu7-2*) mutant, showing that the *anu7-2* mutation does not completely abolish the function of the protein. Similar to the *anu7-1* allele, the phenotype of the *anu7-2* allele partly recovered in older plants, although not as much as the phenotype of the *anu7-1* allele. To determine if photosynthesis was impaired in these mutants, we carried out measurements of the F_v/F_m ratio in mutant and wild-type plants. Despite the fact that *anu7-1* had slightly lower chlorophyll amounts, *anu7-1* and L*er* showed similar F_m and F_v/F_m values at 16 das. We performed similar measurements using *anu7-2* plants grown in the presence of sucrose until 38 das (when the plants displayed a small rosette with albino leaves). Under these conditions, the *anu7-2* mutant had very low F_m and F_v/F_m values relative to Col-0 (Figure S6). Despite the low stature of their inflorescences, adult *anu7-2* plants were bushier than the wild-type (Figure 1l, o). The severe phenotype of *anu7-2* plants grown in the presence of sucrose closely resembled the phenotype of the *scabra3-2* mutant, which carries a strong allele of the gene encoding the nuclear-encoded plastid-targeted RNA polymerase (Hricová *et al.*, 2006). Based on this observation, we hypothesized that a general defect in the transcription of plastid-encoded genes might be at the origin of the phenotype of the *anu7* mutants.

Microarray analysis of the *anu7-1* mutant

To identify genes that are differentially expressed in the leaves of the *anu7-1* mutant and the L*er* wild-type, we performed a microarray analysis with three biological replicates per genotype using Affymetrix ATH1 genome arrays. In order to increase the power of the statistical analysis and to eliminate non-relevant probe sets, we used the interquartile range (IQR) as a filtering criterion. We focused on those probe sets whose $\log_2$ (intensity) values had an IQR greater than 0.5. Among the 22 810 probe sets present in the arrays, 1652 passed this criterion. We selected differentially expressed genes in L*er* samples compared to *anu7-1* samples using *t* and *F* statistics as described by Smyth (2004). After FDR correction (Benjamini and

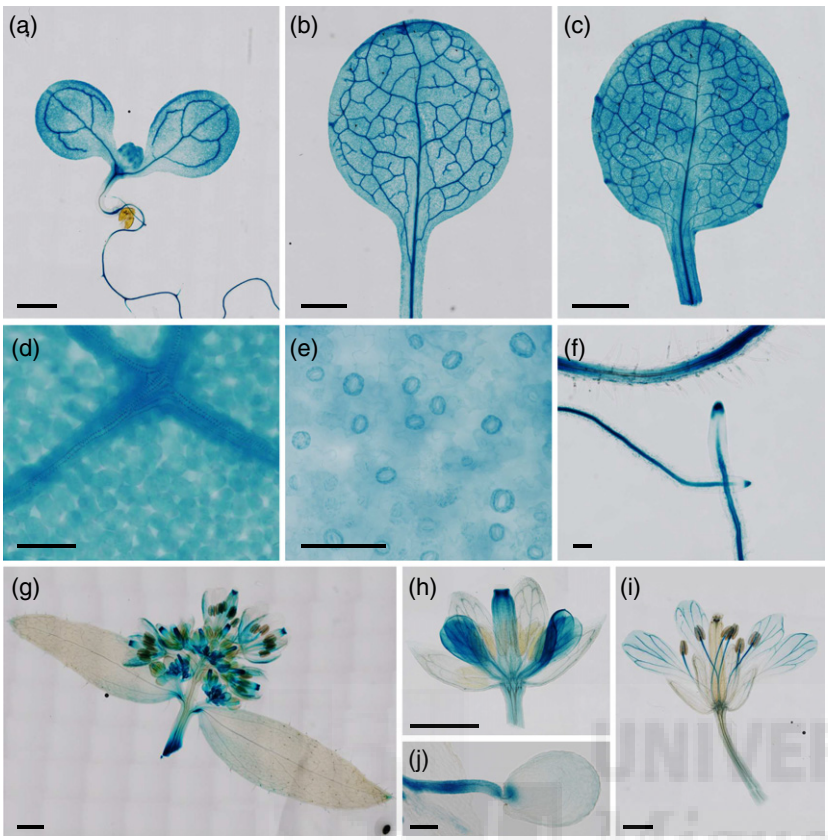


Figure 6. Visualization of *ANU7*$_{pro}$*:GUS* activity in a wild-type L*er* background.
(a–j) (a) Seedling, (b) first-node leaf, (c) third-node leaf, (d) first-node leaf veins and palisade mesophyll, (e) stomata, (f) roots, (g) inflorescence and cauline leaves, (h) immature flower, (i) mature flower, (j) funiculus and developing seed. Pictures were taken (a) 7, (b–f) 14, (g–i) 32 or (j) 37 days after stratification (das). Scale bars indicate (a–c, g–i) 1 mm, and (d–f, j) 100 μm.

Hochberg, 1995), 826 probe sets were detected as being signficantly differentially expressed, with $\log_2$ (ratio) values ranging from 0.486 to 4.88 ($1.4 \leq$ ratio ≤ 29.4). Of these, 280 probe sets corresponded to genes that were expressed at lower levels in *anu7-1*, and 546 corresponded to genes that were expressed at higher levels in *anu7-1* (Table S2). To better understand these results, we searched for enriched gene ontology (GO) terms in these sets using the agriGO analysis tool (Du *et al.*, 2010). Forty-seven GO terms were found to be significantly enriched in the subset of probes that were expressed at higher levels in the *anu7-1* mutant, including four in the 'Cellular Component' category, five in the 'Molecular Function' category, and 38 in the 'Biological Process' category (Table S3). In line with the observed phenotype, 98 probes shared the significantly overrepresented terms 'plastid' (GO:0009536) and 'chloroplast' (GO:0009507). Eight probes shared an additional term, 'nucleoid' (GO:0009295), and corresponded to seven genes that encode components of plastid transcriptionally active chromosome (pTAC) complexes, such as *FE SUPEROXIDE DISMUTASE 2* (*FSD2*; At5 g51100), *FSD3* (At5 g23310), *FRUCTOKINASE-LIKE 1* (*FLN1*; At3 g54090), *FLN2* (At1 g69200), *PLASTID TRANSCRIPTIONALLY ACTIVE 2* (*PTAC2*; At1 g74850, also known as *PIGMENT DEFECTIVE 343* or *PDE343*), *PTAC10* (At3 g48500, also known as *PDE312*) and *PTAC14* (At4 g20130). Mutations in these genes cause severe defects in chloroplast development and thylakoid organization (Pfalz *et al.*, 2006; Myouga *et al.*, 2008; Steiner *et al.*, 2011; Gilkerson *et al.*, 2012), and their upregulation might reflect a compensatory response that is triggered in response to the plastid defects seen in *anu7-1*. The set of genes that are expressed at higher levels also included many genes encoded in the plastid genome, with a majority of them corresponding to genes transcribed by the nuclear-encoded RNA polymerase (NEP). Among these genes were three subunits of the plastid-encoded RNA polymerase (PEP; *rpoB*, *rpoC1* and *rpoC2*), many genes encoding ribosomal proteins (*rpl14*, *rpl20*, *rpl22*, *rpl23*, *rpl36*, *rps15*, and *rps16*), as well as other genes thought to be transcribed exclusively by NEP (*accD*, *ycf2*) or both by NEP and PEP (*ndhB*, *ycf1*). In order to validate the results of our microarray analysis using qRT-PCR, we selected seven genes encoding proteins associated to pTAC complexes: *FSD2*, *FSD3*, *FLN1*, *FLN2*, *PTAC2*, *PTAC10* and *PTAC14*. As in the microarray, these seven genes were significantly expressed at higher levels in the *anu7-1* mutant relative to the wild-type L*er* (Figure 7).

Eighteen additional GO terms were found to be significantly enriched in the subset of probes that were expressed at lower levels in the *anu7-1* mutant, including 10 in the 'Biological Process' category and eight in the 'Molecular Function' category (Table S4). The most significantly

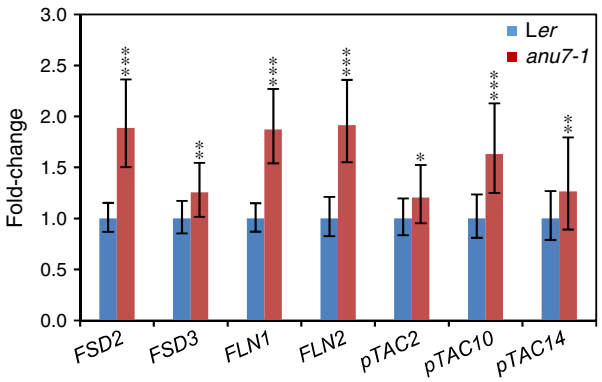


Figure 7. Expression of genes encoding pTAC components in the *anu7-1* mutant.
qRT-PCR analysis of the nuclear genes *FSD2, FSD3, FLN1, FLN2, pTAC2, pTAC10* and *pTAC14* in L*er* and *anu7-1* rosettes collected 16 days after stratification (das). Bars indicate relative expression levels, as determined using the comparative C_T method and normalized with the *OTC* housekeeping gene. Error bars indicate $2^{-(\Delta\Delta C_T)} \pm$ SD. Asterisks indicate significantly different ΔC_T values using Mann–Whitney *U*-tests (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, $n = 9$).

enriched term in this subset was 'response to hormone stimulus' (GO:0009725), which was shared by 36 probe sets. Of these, 20 probe sets shared the 'response to auxin stimulus' (GO:0009733) term and eight the 'response to cytokinin stimulus' (GO:0009735) term. In link with the defects observed in the leaf mesophyll, numerous auxin-responsive genes were expressed at lower levels in the *anu7-1* mutant, including numerous genes encoding members of the Aux/IAA, ARF and SAUR protein families (Table S3). In addition to these, four genes of the ARR cytokinin response family were also expressed at lower levels in *anu7-1.*

The *anu7-1* mutation genetically interacts with *gun1-1*

To investigate if the plastid defects observed in the *anu7-1* mutant trigger a retrograde (plastid-to-nucleus) signal, we crossed *anu7-1* to several non-allelic *genomes uncoupled* (*gun*) mutants, all of which exhibit defective retrograde signaling (Cottage *et al.*, 2010). In our growth conditions, *gun1-1* plants are clearly distinct from *anu7-1* plants, as the former have a wild-type morphology with green, normally-shaped leaves. We isolated *anu7-1 gun1-1* double mutant plants in the selfed progeny of *anu7-1*/*anu7-1*;*GUN1*/*gun1-1* sesquimutant plants. The genotype of the double mutants was confirmed by genotyping the plants by PCR and sequencing using primers from Table S5. The *anu7-1 gun1-1* double mutant exhibited a synergistic phenotype characterized by the presence of sectors with distinct pigmentation in the leaf lamina, from albino to completely green (Figure 8). The shape, orientation and distribution of these sectors were reminiscent of those seen in X-ray-induced chimeras (Steeves and Sussex, 1989), strongly suggesting that distinct lineages of mesophyll cells are stably propagated through the mitotic divisions that take place during leaf development (Figure 8e–h). Whether these sectors consist of mesophyll cells arising through a mechanism involving damage to the organellar genomes (like in the *chloroplast mutator* mutant of Arabidopsis; Martínez-Zapater, 1993) or through a different mechanism, remains an open question. In addition, the prominent marginal teeth characteristic of the *anu7-1* single mutant were clearly, albeit not completely, suppressed in the *anu7-1 gun1-1* double mutant (Figure 8). The increased intercellular spaces and the irregular distribution of palisade mesophyll cell sizes observed in *anu7-1* were also suppressed in *anu7-1 gun1-1* plants (Figure S7). Other measured parameters, such as rosette area and pigment content, remained significantly different from those of the wild-type Col-0 (Figures S8 and S9).

We also obtained *anu7-1 gun2-1*, *anu7-1 gun3-1* and *anu7-1 gun4-1* double mutants, which were very similar to one another and displayed an additive phenotype (Figure S10). The lack of a genetic interaction between *anu7-1* and these mutations can be attributed to the participation of *GUN2*, *GUN3* and *GUN4* in tetrapyrrole biosynthesis, while *GUN1* seems to participate in an independent pathway (Vinti *et al.*, 2000; Strand *et al.*, 2003).

DISCUSSION

Our characterization of the *anu7-1* mutant has allowed us to identify the *ANU7* gene, which was found to be the same gene as *EMB2737*, initially defined by its embryonic-lethal alleles (Tzafrir *et al.*, 2003). As an essential gene, our expectation was to find homologues in all major clades of land plants. However, we were not able to find homologues in two clades that have been subjected to extensive genome sequencing efforts: the grasses and the gymnosperms. Except for these two exceptions, we found genes encoding highly similar proteins in all groups of land plants, including the bryophyte *Physcomitrella patens* and the lycophyte fern *Selaginella moellendorffii.* This result demonstrates that the common ancestor of land plants already had an *ANU7*-like gene in its genome. The absence of similar sequences in the genomes of grasses and gymnosperms indicates that this gene has possibly been lost independently at least twice during the evolution of land plants. Different from many other plant-specific genes that function in the plastids, we were not able to identify a distant homologue in Cyanobacteria, and all our attempts to identify a chloroplast transit peptide using computational tools failed. However, ANU7 has been identified as a plastid-localized protein by previous authors who have detected it in the plastids using different proteomic approaches (Zybailov *et al.*, 2008; Huang *et al.*, 2013). We unsuccessfully tried to confirm the chloroplast localization pattern of ANU7 by expressing N- and C-terminal fusions of ANU7 to three different fluorescent proteins.

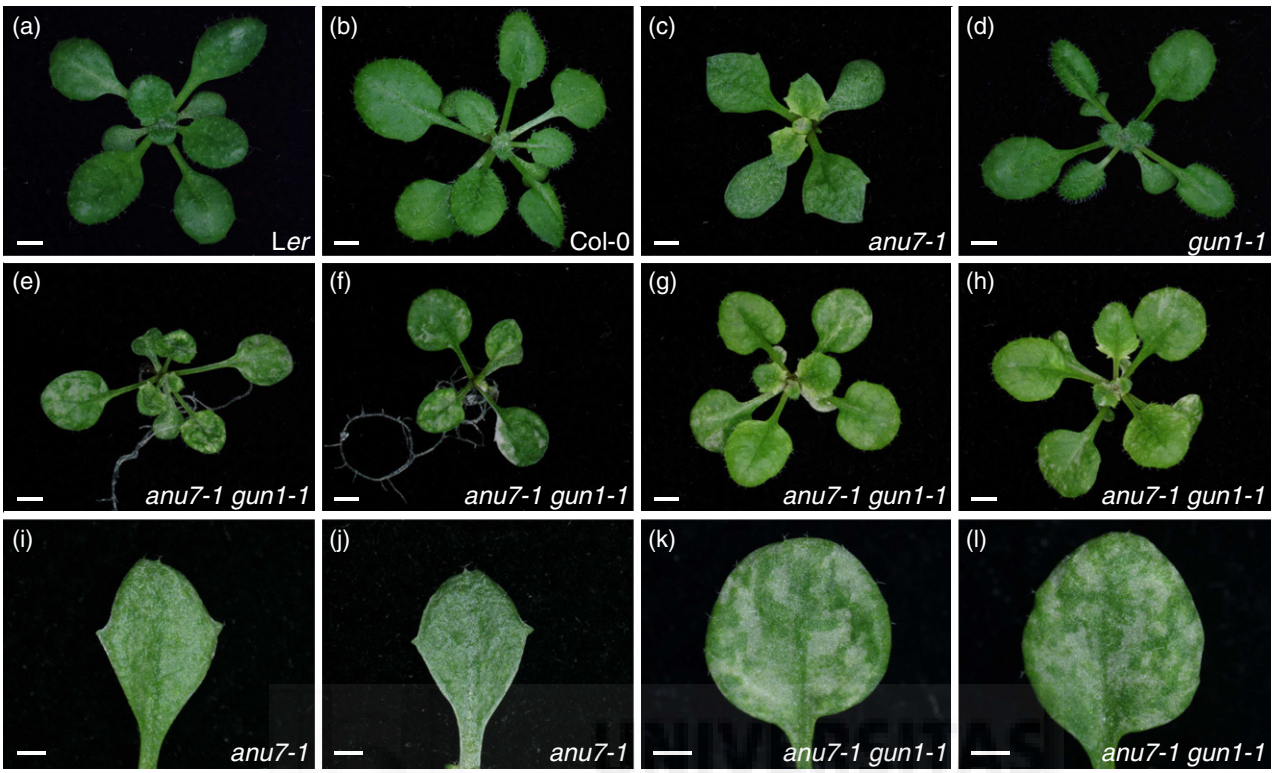


Figure 8. Leaf phenotype of the *anu7-1 gun1-1* double mutant.
(a–l) (a–h) Rosettes, and (i–l) first-node leaves of (a) L*er*, (b) Col-0, (c, i, j) *anu7-1*, and (e–h, k, l) *anu7-1 gun1-1*. Pictures were taken (a–f, i–l) 16 days after stratification (das), and (g, h) 18 das. Scale bars indicate (a–h) 2 mm, and (i–l) 1 mm.

The role of ANU7 in the plastids, however, is granted by the defects in plastid ultrastructure and the reduced F_v/F_m ratio observed in *anu7* mutants. Indeed, the observation that the embryonic-lethal mutants can grow in the presence of an external carbon source (e.g. sucrose added to the culture medium) further supports the idea that the *ANU7* gene plays a previously undescribed role in plastid function and carbon fixation, as also indicated by the locally unappressed thylakoidal membranes observed in the *anu7-1* chloroplasts.

Considering the similarity of the central CR domains of ANU7 and other DnaJ proteins, ANU7 proteins might have evolved from an ancestral DnaJ protein at an early stage in the evolution of land plants. Indeed, numerous DnaJ proteins have been shown to function in the chloroplasts, being involved in processes as diverse as the assembly and maintenance of photosynthetic complexes (Shimada *et al.*, 2007; Albrecht *et al.*, 2008; Chen *et al.*, 2010; Lu *et al.*, 2011; Muranaka *et al.*, 2012; Tanz *et al.*, 2012; Fristedt *et al.*, 2014; Kong *et al.*, 2014), the assembly and regulation of Rubisco activity (Brutnell *et al.*, 1999; Chen *et al.*, 2010; Wang *et al.*, 2015), chloroplast development and differentiation (Vitha *et al.*, 2003; Zhong *et al.*, 2013; Zhu *et al.*, 2015; Wang *et al.*, 2016), thylakoid membrane formation (Liu *et al.*, 2005) and resistance against biotic stresses (Wang *et al.*, 2014). Canonical DnaJ proteins contain four characteristic domains: a J-domain, a glycine/phenylalanine (G/F) rich domain, the above-mentioned CR domain and a carboxy-terminal domain. However, numerous DnaJ-like proteins lacking one or more of these domains have also been described. Based on the domains present, these proteins have been classified as type I (containing all the domains), type II (containing only the J-domain and the G/F domain), and type III (containing only the J-domain; Cheetham and Caplan, 1998). In addition, some authors have found proteins that do not fit well into this classification. Our results suggest that ANU7 belongs to a distinct class, which we refer to as type IV, that comprises DnaJ-like proteins, such as the chloroplast-localized LOW QUANTUM YIELD OF PHOTOSYSTEM II1 (LQY1; Lu *et al.*, 2011), SNOWY COTYLEDON 2 (SCO2; Shimada *et al.*, 2007; Albrecht *et al.*, 2008; Muranaka *et al.*, 2012; Tanz *et al.*, 2012), PLASTID TRANSCRIPTIONALLY ACTIVE 5 (PTAC5; Zhong *et al.*, 2013) and PSA2 (Fristedt *et al.*, 2014; Wang *et al.*, 2016) proteins, which only contain the CR domain and lack other domains characteristic of DnaJ proteins. LQY1, SCO2, PTAC5 and PSA2 have been shown to exhibit protein disulfide isomerase (PDI) activity *in vitro*. Because this domain is only partially conserved in ANU7 and all its putative orthologs, it is presently unclear whether ANU7 can catalyze the reduction of disulfide bonds or if it can help other proteins to fold.

Supporting a role for ANU7 in the plastids, a significant number of nuclear genes whose products normally function in the plastids were found overexpressed in the *anu7-1* mutant, including genes that code for seven subunits of pTAC complexes, which regulate gene expression in the plastid genome. In close agreement with this, we also detected higher transcript levels for numerous plastid-encoded genes, particularly those transcribed by the NEP, such as the *rpo* genes encoding subunits of the PEP. The elevated levels of nucleus-encoded pTAC components in the *anu7-1* mutant are likely to represent a compensatory response triggered by the observed defects in chloroplast organization and function. The higher expression levels seen for plastid genes that are transcribed by NEP can also be seen as part of this compensatory response. Interestingly, the accumulation of transcripts for pTAC and PEP subunits was not paralleled by a concomitant increase in the expression of PEP-transcribed genes, suggesting that ANU7 might function in the correct folding or assembly of the plastid transcriptional machinery. Mutations that damage two different bacterial-type sigma subunits, SIGMA2 (SIG2) and SIG6, which are required for the activity of PEP, lead to a similar overexpression of NEP-transcribed plastid genes (Nagashima *et al.*, 2004; Ishizaki *et al.*, 2005). To illustrate this, 17 out of 30 plastid-encoded genes reported to be expressed at higher levels in a *sig2-1* mutant were also overexpressed in the *anu7-1* mutant (Nagashima *et al.*, 2004).

The work of Zhong *et al.* (2013) on PTAC5 allows us to speculate on the possible function of ANU7 in the plastid. These authors hypothesized that the PDI activity of PTAC5 might be involved in disulfide bond formation in CR proteins that form part or interact with PEP complexes, such as PTAC2 and the plastid-encoded subunits of PEP, which were overexpressed in the *anu7-1* mutant. The synergistic interaction between the *anu7-1* mutation and a loss-of-function mutation of *GUN1*, which encodes a protein most similar to the PTAC2 component of pTAC complexes (Koussevitzky *et al.*, 2007; Pesaresi *et al.*, 2007; Pogson *et al.*, 2008), points to a functional relationship between ANU7 and GUN1, both of which have been isolated in the nucleoid fraction of plastids (Cottage *et al.*, 2008; Huang *et al.*, 2013). Because GUN1 contains 13 Cys residues and co-localizes with PTAC2, we hypothesize that ANU7 and PTAC5 might play similar roles in protein folding, perhaps by acting on a shared set of targets, including GUN1, PTAC2 and/or other cysteine-containing proteins. According to this idea, the redundant roles played by nucleoid-localized type-IV DnaJ proteins might be at the basis of the observed synergistic interaction between *anu7-1* and *gun-1.* Indeed, GUN1 has been recently found to coimmunoprecipitate with PTAC2, PTAC5 and several Hsp70 proteins (Tadini *et al.*, 2016). Although ANU7 was not detected in this experiment, it might still interact with GUN1 indirectly, because DnaJ proteins are known partners of Hsp70 proteins.

Two intriguing aspects of the *anu7-1* allele are the suppression of the margin serration and its ability to produce a variegated phenotype in a double mutant combination involving the hypomorphic *gun1-1* allele. Previous authors have established a link between variegation and plastid translation, as mutations in two genes encoding different translation initiation factors, SUPPRESSOR OF VARIEGATION 9 (SVR9) and FU-GAERI 1 (FUG1), can suppress the effects of *variegated1* (*var1*) and *var2* alleles (Miura *et al.*, 2007; Zheng *et al.*, 2016). The suppression of *var2* variegation is mediated by a reduction in plastid translation, indicating that less FtsH2 protein is needed when translation decreases. GUN1, which is known to coimmunoprecipitate with FUG1 (Tadini *et al.*, 2016), is involved in attenuating the expression of chloroplast-targeted proteins when plastids are dysfunctional (Kakizaki *et al.*, 2009), presumably as a protective mechanism to help to cope with the dysfunction. As a result, loss of GUN1 will lead to elevated expression of some genes, and enhancement of some plastid activities, which may actually be harmful. Loss of the GUN1-mediated attenuation would allow a more wild-type-like leaf margin morphology of *anu7-1 gun1-1*, and might also explain its variegated phenotype.

The compromised translation of plastid proteins in *svr9* mutants not only suppresses the variegation, but also causes increased intercellular air space and a serrated leaf margin that are very similar to those of *anu7-1* leaves. The margin serration of *svr9* is thought to occur as a consequence of changes in auxin homeostasis (Zheng *et al.*, 2016). The changes in *anu7-1* might also result from defective auxin biosynthesis or transport, as pointed out by Zheng *et al.* (2016). We found numerous auxin-responsive genes expressed at lower levels in the *anu7-1* mutant, including genes encoding members of the Aux/IAA, ARF and SAUR protein families. In addition to these, four members of the ARR family of cytokinin-responsive genes were also expressed at lower levels in *anu7-1.* Previous work in our laboratory (Pérez-Pérez *et al.*, 2010) has linked some of these genes with the proliferation of mesophyll cells in the leaves of *Arabidopsis thaliana*, suggesting that the abnormal growth of *anu7-1* leaves might, at least in part, be a consequence of an altered auxin response. The work of Andriankaja *et al.* (2012) has shown that the exit of leaf mesophyll cells from a proliferative state and their progression towards a differentiated state depend on signals produced by the chloroplast. The lower expression levels of auxin- and cytokinin-related genes in the *anu7-1* mutant suggest that the chloroplast might trigger such a developmental transition by means of hormonal signals.

EXPERIMENTAL PROCEDURES

Plant material, growth conditions and crosses

Seeds of the *Arabidopsis thaliana* (L.) Heynh. wild-type accessions Col-0 and L*er*, as well as T-DNA heterozygous lines *EMB2737*/*emb2737-1* (N16187) and *EMB2737*/*emb2737-2* (N24089), were obtained from the Nottingham Arabidopsis Stock Centre (NASC). The *anu7-1* mutant was isolated in a L*er* background after EMS-induced mutagenesis as previously described in Berná *et al.* (1999). Seeds of *gun1-1*, *gun2-1*, *gun3-1* and *gun4-1* mutants used to test genetic interactions were kindly provided by Dr N. Mochizuki. The *gun1-1* to *gun4-1* mutants were isolated in a screen for EMS-induced mutants resistant to the herbicide Norflurazon carried out on the pOCA107-2 transgenic line in the Col-0 background (Susek *et al.*, 1993; Mochizuki *et al.*, 2001; Larkin *et al.*, 2003).

Seed sterilization and sowing, plant culture, crosses and allelism tests were performed as previously described (Ponce *et al.*, 1998; Berná *et al.*, 1999). The genotypes of double mutants and homozygous T-DNA insertions were confirmed by PCR amplification and/or sequencing.

Plant morphometry, histology and microscopy

Pictures from Arabidopsis whole plants were taken using a Canon PowerShot SX200 IS digital camera. Individual rosettes and leaves were imaged using a Nikon SMZ1500 stereomicroscope equipped with a Nikon DXM1200F digital camera. Micrographs of internal leaf tissues were obtained from cleared leaves using DIC optics on a Leica DMRB microscope equipped with a Nikon DXM1200 digital camera (Candela *et al.*, 1999; Pérez-Pérez *et al.*, 2011). Diagrams of palisade mesophyll cells were generated by drawing the outlines of the cells using a Wacom DTF-720 Pen Display and the Adobe Photoshop CS3 software, as previously described (Pérez-Pérez *et al.*, 2011; Rubio-Díaz *et al.*, 2012; Ferrández-Ayela *et al.*, 2013). Cell area measurements were performed from these diagrams, using the NIS Elements AR 3.1 image analysis package.

Transverse sections of leaves were obtained by fixing and dehydrating plant material as described by Serrano-Cartagena *et al.* (2000). Samples were embedded in Technovit7100 resin and 5-μm thick sections were obtained using a Microm HM350S microtome. Sections were stained with 0.1% toluidine blue and imaged using a Leica DMRB microscope equipped with a Nikon DXM1200 digital camera under bright-field illumination (Hricová *et al.*, 2006).

For TEM, leaf tissue was fixed and prepared as previously described (Hricová *et al.*, 2006; Casanova-Sáez *et al.*, 2014b). Samples were visualized at 80 kV using a JEOL 1011 transmission electron microscope equipped with a Gatan 792 BioScan digital camera.

DAB and GUS staining

To determine the accumulation of hydrogen peroxide, third-node leaves from L*er* and *anu7-1* plants were incubated in 1 mg mL^{-1} 3,3′-diaminobenzidine (DAB; Sigma-Aldrich, Darmstadt, Germany) and cleared as described in Casanova-Sáez *et al.* (2014b). GUS staining of plant tissues was performed as reported in Robles *et al.* (2010).

Dry weight, pigment and chlorophyll fluorescence determination

Dry weight quantification and pigment concentration of L*er*, *anu7-1*, Col-0, *anu7-2* and $35S_{pro}$*:ANU7* transgenic plants at different ages were performed as previously described in Casanova-Sáez *et al.* (2014b). Chlorophyll fluorescence was analysed using pulse amplitude modulation; the maximum efficiency of PSII (F_v/F_m ratio) was measured using a DUAL-PAM/F system (Walz). Before the measurements, L*er* and mutant plants were dark-adapted for 30 min.

Positional cloning of the *anu7-1* mutation

Low-resolution mapping of the *anu7-1* mutation was carried out as described in Ponce *et al.* (2006). For fine mapping, 16 different molecular markers were used (Table S1). To find the *anu7-1* mutation, a 3349 bp fragment encompassing the At5g53860 transcription unit was amplified by PCR from L*er* and *anu7-1* genomic DNA, and sequenced on an ABI PRISM 3130×l Genetic Analyser (Applied Biosystems, Waltham, MA, USA). Primers used for cloning and sequencing the At5g53860 gene can be found in Table S1.

Constructs and plant transformation

To prepare Gateway entry clones for $35S_{pro}$*:ANU7*, $35S_{pro}$*:ANU7:GFP*, $35S_{pro}$*:ANU7:YFP*, $35S_{pro}$*:ANU7:CFP* and $35S_{pro}$*:GFP:ANU7* constructs, the full-length CDS of At5g53860.2 was amplified from Col-0 cDNA using ANU7-cds-F and ANU7-cds-R primers containing attB1 and attB2 sites (Table S1). The ANU7-cds-F1 and ANU7-cds-F3 primers contain a translation start site (ATG) in different frames. The ANU7-cds-R1 primer contains a stop codon. The ANU7-cds-R2 primer lacks a stop codon and was designed to facilitate the making of C-terminal fusions. All inserts were amplified using Phusion polymerase (Thermo Scientific, Waltham, MA, USA) and the amplification products were cloned into the pGEM-T Easy221 vector (kindly provided by Prof. B. Scheres) using BP clonase II (Life Technologies, Waltham, MA, USA). Overexpression constructs ($35S_{pro}$*:ANU7*) were prepared by recombining an entry clone (derived from the ANU7-cds-F1/ANU7-cds-R1 PCR product) and the pMDC32 destination vector (Curtis and Grossniklaus, 2003) using LR clonase II (Life technologies). Constructs to express a C-terminal fusion of ANU7 to GFP ($35S_{pro}$*:ANU7:GFP*), YFP ($35S_{pro}$*:ANU7:YFP*), and CFP ($35S_{pro}$*:ANU7:CFP*) were generated by recombining an entry clone (derived from the ANU7-cds-F1/ANU7-cds-R2 PCR product) and pMDC83 (Curtis and Grossniklaus, 2003), pEarleyGate101 and pEarleyGate102 (Earley *et al.*, 2006), respectively. Constructs to express an N-terminal fusion of ANU7 to GFP ($35S_{pro}$*:GFP:ANU7*) were generated by recombining an entry clone (derived from the ANU7-cds-F3/ANU7-cds-R1 PCR product) and pMDC45 (Curtis and Grossniklaus, 2003). To generate the $ANU7_{pro}$:GUS construct, a 533 bp fragment encompassing the intergenic region between At5 g53850 and At5 g53860, including the 3′-untranslated region (3′-UTR) of At5 g53850 and 5′-UTR of At5 g53860, was amplified from Col-0 genomic DNA using the ANU7-pro-F and ANU7-pro-R primers (which contain attB1 and attB2 sites; Table S1) and cloned into pGEM-T Easy221 using BP clonase II. The insert of the resulting entry clone was then transferred to pMDC163 (Curtis and Grossniklaus, 2003) using LR clonase II. L*er* and *anu7-1* plants were transformed with appropriate constructs using the floral-dip method (Clough and Bent, 1998) using *Agrobacterium tumefaciens* strain C58C1. T1 transgenic plants were selected on Petri dishes supplemented with 15 μg mL^{-1} hygromycin B (Invitrogen, Waltham, MA, USA).

Bioinformatics

We used the command-line version of CAP3 with default parameters (http://pbil.univ-lyon1.fr/cap3.php) (Huang and Madan, 1999) to assemble 87 expressed sequence tag (EST) sequences (downloaded from GenBank on 28 January 2016) in an attempt to identify the biologically relevant transcript isoform of *ANU7*. The identification of putative ANU7 homologues from other plant species was carried out with BLASTP and PSI-BLAST (Altschul *et al.*, 1997). Amino acid sequences were aligned using the consistency-

based method implemented in T-Coffee (Notredame *et al.*, 2000). Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6 (Tamura *et al.*, 2013; http://www.megasoftware.net). The lowest BIC (Bayesian Information Criterion) score corresponded to a Jones-Taylor-Thornton (JTT) substitution pattern, modeling the non-uniformity of evolutionary rates by using a discrete Gamma distribution (+G) with five rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). The analysis involved 61 amino acid sequences. All positions containing gaps and missing data were eliminated, leaving a total of 396 positions in the final dataset.

Statistical analyses

Statistical analyses were carried out in order to compare the *anu7-1*, *anu7-2* and *anu7-1 35S$_{pro}$:ANU7* phenotypic traits with their corresponding wild-type plants. Depending on the number of samples, we used the Mann–Whitney *U*-test ($n \leq 10$) or Student's *t*-test ($n > 10$; Robles *et al.*, 2010).

Microarray

Total RNA was isolated using the Qiagen RNeasy Mini Kit (Qiagen, Hilden, Germany) from three biological replicates of L*er* and *anu7-1* rosettes collected 21 das (12 μg of total RNA per replicate, which were isolated from 100 mg of tissue from at least nine different plants). Microarray experiments were performed at DNAVision (Charleroi, Belgium). Total RNA (3 μg) was labeled with the One-cycle Target Labeling kit (Affymetrix, Santa Clara, CA, USA) according to the manufacturer's protocol. The resulting biotinylated cRNA was fragmented and hybridized to Arabidopsis ATH1 Genome Array GeneChips, which were processed using an Affymetrix Model 450 Fluidics Station and scanned on an Affymetrix Model 3000 scanner. Raw data were normalized using the Robust Microarray Analysis algorithm (Irizarry *et al.*, 2003a,b). Differentially expressed genes were identified using the IQR as a filtering criterion. The IQR is a measure of the data variability, and it is calculated as the difference between the Q3 and Q1 quartiles. We considered as invariant genes those with an IQR below 0.5. GO terms significantly enriched among the differentially expressed genes were identified using the Singular Enrichment Analysis tool implemented in the agriGO server, selecting 'Affymetrix ATH1 Genome Array (GPL198)' as the background reference. Fisher's statistical tests with Yekutieli (FDR under dependency) multi-test correction were performed with a 0.05 significance level.

RNA isolation, cDNA synthesis and RT-qPCR

Total RNA was extracted from ~100 mg of L*er* and *anu7-1* rosettes collected 16 das using TRI Reagent (Sigma-Aldrich). Genomic DNA was eliminated with the TURBO DNA-free Kit (Invitrogen). Reverse transcription was carried out using random hexamers and the Maxima Reverse Transcriptase system (Fermentas, Waltham, MA, USA). Triplicate PCR reactions of four different biological replicates were prepared as described in Mateo-Bonmatí *et al.* (2015). The *ORNITHINE TRANSCARBAMILASE* (*OTC*) gene (Quesada *et al.*, 1999) was used as the internal control. Relative quantification of gene expression data was performed using the comparative C_T method (Schmittgen and Livak, 2008) on a Step One Plus Real-Time PCR System (Applied Biosystems). The primers used for qPCR analyses are listed on Table S1.

ACKNOWLEDGEMENTS

We thank J.M. Serrano, F.M. Lozano, A. Torregrosa and J.M. Sánchez-Larrosa for their excellent technical assistance, Prof. Ben Scheres for the pGEM-T Easy221 plasmid, and Dr. Nobuyoshi Mochizuki for mutant *gun1-1*, *gun2-1*, *gun3-1* and *gun4-1* seeds. Research in the laboratory of J.L.M. was supported by grants from the Ministerio de Economía y Competitividad of Spain (BIO2014-53063-P) and the Generalitat Valenciana (PROMETEOII/2014/006). H.C. was a recipient of a Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073). T.M.-N. held a predoctoral fellowship from the Generalitat Valenciana (ACIF/2013/273). The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Weight and height of *anu7* mutants.

Figure S2. Rosette leaf number at bolting in L*er*, *anu7-1* and *anu7-1 35S$_{pro}$:ANU7* plants.

Figure S3. Diaminobenzidine staining of *anu7-1* and wild-type leaves.

Figure S4. CAP3 assembly of *ANU7* expressed sequence tags.

Figure S5. Multiple sequence alignment and phylogenic tree of ANU7 and its orthologs.

Figure S6. Photosynthetic parameters in L*er*, *anu7-1*, Col-0 and *anu7-2* plants.

Figure S7. Morphometry of palisade mesophyll cells from L*er*, *anu7-1*, Col-0, *gun1-1* and *anu7-1 gun1-1* plants.

Figure S8. Rosette area of *anu7-1*, *gun1-1* and *anu7-1 gun1-1* plants.

Figure S9. Pigment content in L*er*, *anu7-1*, Col-0, *gun1-1* and *anu7-1 gun1-1.*

Figure S10. Vegetative phenotype of the *anu7-1 gun2-1*, *anu7-1 gun3-1* and *anu7-1 gun4-1* double mutants.

Table S1. Identity and similarity of ANU7 and its orthologs.

Table S2. List of genes differentially expressed between mutant and wild-type plants.

Table S3. Gene Ontology terms significantly enriched in genes that are expressed at higher levels.

Table S4. Gene Ontology terms significantly enriched in genes that are expressed at lower levels.

Table S5. Primer sets used in this work.

REFERENCES

Albrecht, V., Ingenfeld, A. and Apel, K. (2008) *Snowy cotyledon 2:* the identification of a zinc finger domain protein essential for chloroplast development in cotyledons but not in true leaves. *Plant Mol. Biol.* **66**, 599–608.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Andriankaja, M., Dhondt, S., De Bodt, S. *et al.* (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana:* a not-so-gradual process. *Dev. Cell*, **22**, 64–78.

Armbruster, U., Hertle, A., Makarenko, E. *et al.* (2009) Chloroplast proteins without cleavable transit peptides: rare exceptions or a major constituent of the chloroplast proteome? *Mol. Plant*, **2**, 1325–1335.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, **57**, 289–300.

Berná, G., Robles, P. and Micol, J.L. (1999) A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana. Genetics*, **152**, 729–742.

Brutnell, T.P., Sawers, R.J., Mant, A. and Langdale, J.A. (1999) BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the *rbcL* gene of maize. *Plant Cell*, **11**, 849–864.

Candela, H., Martínez-Laborda, A. and Micol, J.L. (1999) Venation pattern formation in *Arabidopsis thaliana* vegetative leaves. *Dev. Biol.* **205**, 205–216.

Casanova-Sáez, R., Candela, H. and Micol, J.L. (2014a) Combined haploinsufficiency and purifying selection drive retention of RPL36a paralogs in Arabidopsis. *Sci. Rep.* **4**, 4122.

Casanova-Sáez, R., Mateo-Bonmatí, E., Kangasjärvi, S., Candela, H. and Micol, J.L. (2014b) Arabidopsis ANGULATA10 is required for thylakoid biogenesis and mesophyll development. *J. Exp. Bot.* **65**, 2391–2404.

Cheetham, M.E. and Caplan, A.J. (1998) Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress Chaperones.* **3**, 28–36.

Chen, K.M., Holmstrom, M., Raksajit, W., Suorsa, M., Piippo, M. and Aro, E.M. (2010) Small chloroplast-targeted DnaJ proteins are involved in optimization of photosynthetic reactions in *Arabidopsis thaliana*. *BMC Plant Biol.* **10**, 43.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Cottage, A.J., Mott, E.K., Wang, J.-H. *et al.* (2008) *GUN1* (*GENOMES UNCOUPLED1*) encodes a pentatricopeptide repeat (PPR) protein involved in plastid protein synthesis-responsive retrograde signaling to the nucleus. In *Photosynthesis. Energy From the Sun: 14th International Congress on Photosynthesis* (Allen, J.F., Gantt, E., Golbeck, J.H. and Osmond, B. eds). Dordrecht: Springer Netherlands, pp. 1201–1205.

Cottage, A., Mott, E.K., Kempster, J.A. and Gray, J.C. (2010) The *Arabidopsis* plastid-signalling mutant *gun1* (*genomes uncoupled1*) shows altered sensitivity to sucrose and abscisic acid and alterations in early seedling development. *J. Exp. Bot.* **61**, 3773–3786.

Craig, E.A., Huang, P., Aron, R. and Andrew, A. (2006) The diverse roles of J-proteins, the obligate Hsp70 co-chaperone. *Rev. Physiol. Biochem. Pharmacol.* **156**, 1–21.

Curtis, M.D. and Grossniklaus, U. (2003) A Gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629.

Emanuelsson, O., Nielsen, H. and von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8**, 978–984.

Ferrández-Ayela, A., Alonso-Peral, M.M., Sánchez-García, A.B., Micol-Ponce, R., Pérez-Pérez, J.M., Micol, J.L. and Ponce, M.R. (2013) Arabidopsis *TRANSCURVATA1* encodes NUP58, a component of the nucleopore central channel. *PLoS ONE*, **8**, e67661.

Fristedt, R., Williams-Carrier, R., Merchant, S.S. and Barkan, A. (2014) A thylakoid membrane protein harboring a DnaJ-type zinc finger domain is required for photosystem I accumulation in plants. *J. Biol. Chem.* **289**, 30657–30667.

Gilkerson, J., Perez-Ruiz, J.M., Chory, J. and Callis, J. (2012) The plastid-localized pfkB-type carbohydrate kinases FRUCTOKINASE-LIKE 1 and 2 are essential for growth and development of *Arabidopsis thaliana*. *BMC Plant Biol.* **12**, 102.

González, N. and Inzé, D. (2015) Molecular systems governing leaf growth: from genes to networks. *J. Exp. Bot.* **66**, 1045–1054.

Goodstein, D.M., Shu, S., Howson, R. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186.

Hennessy, F., Nicoll, W.S., Zimmermann, R., Cheetham, M.E. and Blatch, G.L. (2005) Not all J domains are created equal: implications for the specificity of Hsp40-Hsp70 interactions. *Protein Sci.* **14**, 1697–1709.

Horiguchi, G., Mollá-Morales, A., Pérez-Pérez, J.M., Kojima, K., Robles, P., Ponce, M.R., Micol, J.L. and Tsukaya, H. (2011) Differential contributions of ribosomal protein genes to *Arabidopsis thaliana* leaf development. *Plant J.* **65**, 724–736.

Hricová, A., Quesada, V. and Micol, J.L. (2006) The *SCABRA3* nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in Arabidopsis. *Plant Physiol.* **141**, 942–956.

Hsu, C.M., Chen, C.Y. and Liu, B.J. (2011) WildSpan: mining structured motifs from protein sequences. *Algorithms Mol. Biol.* **6**, 6.

Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877.

Huang, M., Friso, G., Nishimura, K., Qu, X., Olinares, P.D., Majeran, W., Sun, Q. and van Wijk, K.J. (2013) Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. *J. Proteome Res.* **12**, 491–504.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. and Speed, T.P. (2003a) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003b) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

Ishizaki, Y., Tsunoyama, Y., Hatano, K., Ando, K., Kato, K., Shinmyo, A., Kobori, M., Takeba, G., Nakahira, Y. and 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.

Jarvis, P., Chen, L.J., Li, H., Peto, C.A., Fankhauser, C. and Chory, J. (1998) An *Arabidopsis* mutant defective in the plastid general protein import apparatus. *Science*, **282**, 100–103.

Kakizaki, T., Matsumura, H., Nakayama, K., Che, F.S., Terauchi, R. and Inaba, T. (2009) Coordination of plastid protein import and nuclear gene expression by plastid-to-nucleus retrograde signaling. *Plant Physiol.* **151**, 1339–1353.

Kong, F., Deng, Y., Zhou, B., Wang, G., Wang, Y. and Meng, Q. (2014) A chloroplast-targeted DnaJ protein contributes to maintenance of photosystem II under chilling stress. *J. Exp. Bot.* **65**, 143–158.

Koussevitzky, S., Stanne, T.M., Peto, C.A., Giap, T., Sjogren, L.L., Zhao, Y., Clarke, A.K. and Chory, J. (2007) An *Arabidopsis thaliana* virescent mutant reveals a role for ClpR1 in plastid development. *Plant Mol. Biol.* **63**, 85–96.

Larkin, R.M., Alonso, J.M., Ecker, J.R. and Chory, J. (2003) GUN4, a regulator of chlorophyll synthesis and intracellular signaling. *Science*, **299**, 902–906.

Liu, C., Willmund, F., Whitelegge, J.P., Hawat, S., Knapp, B., Lodha, M. and 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. and Schroda, M. (2007) The chloroplast HSP70B-CDJ2-CGE1 chaperones catalyse assembly and disassembly of VIPP1 oligomers in *Chlamydomonas*. *Plant J.* **50**, 265–277.

Lu, Y., Hall, D.A. and Last, R.L. (2011) A small zinc finger thylakoid protein plays a role in maintenance of photosystem II in *Arabidopsis thaliana*. *Plant Cell*, **23**, 1861–1875.

Martínez-Zapater, J.M. (1993) Genetic analysis of variegated mutants in *Arabidopsis*. *J. Hered.* **84**, 138–140.

Mateo-Bonmatí, E., Casanova-Sáez, R., Candela, H. and Micol, J.L. (2014) Rapid identification of *angulata* leaf mutations using next-generation sequencing. *Planta*, **240**, 1113–1122.

Mateo-Bonmatí, E., Casanova-Sáez, R., Quesada, V., Hricová, A., Candela, H. and Micol, J.L. (2015) Plastid control of abaxial-adaxial patterning. *Sci. Rep.* **5**, 15975.

Miura, E., Kato, Y., Matsushima, R., Albrecht, V., Laalami, S. and Sakamoto, W. (2007) The balance between protein synthesis and degradation in chloroplasts determines leaf variegation in Arabidopsis *yellow variegated* mutants. *Plant Cell*, **19**, 1313–1328.

Mochizuki, N., Brusslan, J.A., Larkin, R., Nagatani, A. and Chory, J. (2001) *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.* **98**, 2053–2058.

Mollá-Morales, A., Sarmiento-Mañús, R., Robles, P., Quesada, V., Pérez-Pérez, J.M., González-Bayón, R., Hannah, M.A., Willmitzer, L., Ponce, M.R. and Micol, J.L. (2011) Analysis of *ven3* and *ven6* reticulate mutants reveals the importance of arginine biosynthesis in Arabidopsis leaf development. *Plant J.* **65**, 335–345.

Muranaka, A., Watanabe, S., Sakamoto, A. and Shimada, H. (2012) *Arabidopsis* cotyledon chloroplast biogenesis factor CYO1 uses glutathione as an electron donor and interacts with PSI (A1 and A2) and PSII (CP43 and CP47) subunits. *J. Plant Physiol.* **169**, 1212–1215.

Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M. and Shinozaki, K. (2008) A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell*, **20**, 3148–3162.

Nagashima, A., Hanaoka, M., Motohashi, R., Seki, M., Shinozaki, K., Kanamaru, K., Takahashi, H. and Tanaka, K. (2004) DNA microarray analysis of plastid gene expression in an *Arabidopsis* mutant deficient in a plastid transcription factor sigma, SIG2. *Biosci. Biotechnol. Biochem.* **68**, 694–704.

Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.

Pérez-Pérez, J.M., Candela, H., Robles, P., López-Torrejón, G., del Pozo, J.C. and Micol, J.L. (2010) A role for *AUXIN RESISTANT3* in the coordination of leaf growth. *Plant Cell Physiol.* **51**, 1661–1673.

Pérez-Pérez, J.M., Candela, H., Robles, P., Quesada, V., Ponce, M.R. and Micol, J.L. (2009) Lessons from a search for leaf mutants in *Arabidopsis thaliana*. *Int. J. Dev. Biol.* **53**, 1623–1634.

Pérez-Pérez, J.M., Rubio-Díaz, S., Dhondt, S., Hernández-Romero, D., Sánchez-Soriano, J., Beemster, G.T., Ponce, M.R. and Micol, J.L. (2011) Whole organ, venation and epidermal cell morphological variations are correlated in the leaves of *Arabidopsis* mutants. *Plant, Cell Environ.* **34**, 2200–2211.

Pesaresi, P., Schneider, A., Kleine, T. and Leister, D. (2007) Interorganellar communication. *Curr. Opin. Plant Biol.* **10**, 600–606.

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.J. and Oelmuller, R. (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell*, **18**, 176–197.

Pogson, B.J., Woo, N.S., Forster, B. and Small, I.D. (2008) Plastid signalling to the nucleus and beyond. *Trends Plant Sci.* **13**, 602–609.

Ponce, M.R., Quesada, V. and Micol, J.L. (1998) Rapid discrimination of sequences flanking and within T-DNA insertions in the *Arabidopsis* genome. *Plant J.* **14**, 497–501.

Ponce, M.R., Robles, P., Lozano, F.M., Brotons, M.A. and Micol, J.L. (2006) Low-resolution mapping of untagged mutations. *Methods Mol. Biol.* **323**, 105–113.

Quesada, V., Ponce, M.R. and Micol, J.L. (1999) *OTC* and *AUL1*, two convergent and overlapping genes in the nuclear genome of *Arabidopsis thaliana*. *FEBS Lett.* **461**, 101–106.

Quesada, V., Sarmiento-Mañús, R., González-Bayón, R., Hricová, A., Ponce, M.R. and Micol, J.L. (2013) *PORPHOBILINOGEN DEAMINASE* deficiency alters vegetative and reproductive development and causes lesions in Arabidopsis. *PLoS ONE*, **8**, e53378.

Robles, P. and Micol, J.L. (2001) Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. *Mol. Genet. Genomics*, **266**, 12–19.

Robles, P., Fleury, D., Candela, H. *et al.* (2010) The *RON1*/*FRY1*/*SAL1* gene is required for leaf morphogenesis and venation patterning in Arabidopsis. *Plant Physiol.* **152**, 1357–1372.

Rubio-Díaz, S., Pérez-Pérez, J.M., González-Bayón, R. *et al.* (2012) Cell expansion-mediated organ growth is affected by mutations in three *EXIGUA* genes. *PLoS ONE*, **7**, e36500.

Schein, A.I., Kissinger, J.C. and Ungar, L.H. (2001) Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic Acids Res.* **29**, E82.

Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* **3**, 1101–1108.

Serrano-Cartagena, J., Candela, H., Robles, P., Ponce, M.R., Pérez-Pérez, J.M., Piqueras, P. and Micol, J.L. (2000) Genetic analysis of *incurvata* mutants reveals three independent genetic operations at work in *Arabidopsis* leaf morphogenesis. *Genetics*, **156**, 1363–1377.

Shimada, H., Mochizuki, M., Ogura, K., Froehlich, J.E., Osteryoung, K.W., Shirano, Y., Shibata, D., Masuda, S., Mori, K. and Takamiya, K. (2007) *Arabidopsis* cotyledon-specific chloroplast biogenesis factor CYO1 is a protein disulfide isomerase. *Plant Cell*, **19**, 3157–3169.

Shimoni-Shor, E., Hassidim, M., Yuval-Naeh, N. and Keren, N. (2010) Disruption of Nap14, a plastid-localized non-intrinsic ABC protein in *Arabidopsis thaliana* results in the over-accumulation of transition metals and in aberrant chloroplast structures. *Plant, Cell Environ.* **33**, 1029–1038.

Skalitzky, C.A., Martin, J.R., Harwood, J.H., Beirne, J.J., Adamczyk, B.J., Heck, G.R., Cline, K. and Fernández, D.E. (2011) Plastids contain a second sec translocase system with essential functions. *Plant Physiol.* **155**, 354–369.

Sluis, A. and Hake, S. (2015) Organogenesis in plants: initiation and elaboration of leaves. *Trends Genet.* **31**, 300–306.

Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 1–25. Article3.

Steeves, T.A. and Sussex, I.M. (1989). In *Patterns in Plant Development*. Cambridge: University Press, pp. 76–80.

Steiner, S., Schroter, Y., Pfalz, J. and Pfannschmidt, T. (2011) Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. *Plant Physiol.* **157**, 1043–1055.

Strand, A., Asami, T., Alonso, J., Ecker, J.R. and Chory, J. (2003) Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrin IX. *Nature*, **421**, 79–83.

Susek, R.E., Ausubel, F.M. and Chory, J. (1993) Signal transduction mutants of Arabidopsis uncouple nuclear *CAB* and *RBCS* gene expression from chloroplast development. *Cell*, **74**, 787–799.

Tadini, L., Pesaresi, P., Kleine, T. *et al.* (2016) GUN1 controls accumulation of the plastid ribosomal protein S1 at the protein level and interacts with proteins involved in plastid protein homeostasis. *Plant Physiol.* **170**, 1817–1830.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.

Tanz, S.K., Kilian, J., Johnsson, C., Apel, K., Small, I., Harter, K., Wanke, D., Pogson, B. and Albrecht, V. (2012) The SCO2 protein disulphide isomerase is required for thylakoid biogenesis and interacts with LHCB1 chlorophyll *a*/*b* binding proteins which affects chlorophyll biosynthesis in Arabidopsis seedlings. *Plant J.* **69**, 743–754.

Tsukaya, H., Byrne, M.E., Horiguchi, G., Sugiyama, M., Van Lijsebettens, M. and Lenhard, M. (2013) How do 'housekeeping' genes control organogenesis? Unexpected new findings on the role of housekeeping genes in cell and organ differentiation. *J. Plant. Res.* **126**, 3–15.

Tzafrir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., Patton, D. and Meinke, D. (2003) The Arabidopsis SeedGenes project. *Nucleic Acids Res.* **31**, 90–93.

Van Minnebruggen, A., Neyt, P., De Groeve, S., Coussens, G., Ponce, M.R., Micol, J.L. and Van Lijsebettens, M. (2010) The *ang3* mutation identified the ribosomal protein gene *RPL5B* with a role in cell expansion during organ growth. *Physiol. Plant.* **138**, 91–101.

Vinti, G., Hills, A., Campbell, S., Bowyer, J.R., Mochizuki, N., Chory, J. and Lopez-Juez, E. (2000) Interactions between *hy1* and *gun* mutants of *Arabidopsis*, and their implications for plastid/nuclear signalling. *Plant J.* **24**, 883–894.

Vitha, S., Froehlich, J.E., Koksharova, O., Pyke, K.A., van Erp, H. and Osteryoung, K.W. (2003) ARC6 is a J-domain plastid division protein and an evolutionary descendant of the cyanobacterial cell division protein Ftn2. *Plant Cell*, **15**, 1918–1933.

Wang, G., Cai, G., Kong, F., Deng, Y., Ma, N. and Meng, Q. (2014) Overexpression of tomato chloroplast-targeted DnaJ protein enhances tolerance to drought stress and resistance to *Pseudomonas solanacearum* in transgenic tobacco. *Plant Physiol. Biochem.* **82**, 95–104.

Wang, G., Kong, F., Zhang, S., Meng, X., Wang, Y. and Meng, Q. (2015) A tomato chloroplast-targeted DnaJ protein protects Rubisco activity under heat stress. *J. Exp. Bot.* **66**, 3027–3040.

Wang, Y.W., Chen, S.M., Wang, W.J., Huang, X.Q., Zhou, C.F., Zhuang, Z. and Lu, S. (2016) The DnaJ-like zinc finger domain protein PSA2 affects light acclimation and chloroplast development in *Arabidopsis thaliana*. *Front. Plant. Sci.* **7**, 360.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, **2**, e718.

Zheng, M., Liu, X., Liang, S., Fu, S., Qi, Y., Zhao, J., Shao, J., An, L. and Yu, F. (2016) Chloroplast translation initiation factors regulate leaf variegation and development. *Plant Physiol.* **172**, 1117–1130.

Zhong, L., Zhou, W., Wang, H., Ding, S., Lu, Q., Wen, X., Peng, L., Zhang, L. and Lu, C. (2013) Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in *Arabidopsis* under heat stress. *Plant Cell*, **25**, 2925–2943.

Zhu, X., Liang, S., Yin, J. *et al.* (2015) The DnaJ OsDjA7/8 is essential for chloroplast development in rice (*Oryza sativa*). *Gene*, **574**, 11–19.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q. and van Wijk, K.J. (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE*, **3**, e1994.

UNIVERSITAS
Miguel
Hernández

The *ANGULATA7* gene encodes a DnaJ-like zinc-finger-domain protein involved in chloroplast function and leaf development in Arabidopsis

Tamara Muñoz-Nortes, José Manuel Pérez-Pérez, María Rosa Ponce, Héctor Candela, and José Luis Micol

Supporting Information

(Figures S1-S10 and Tables S1, S3-S5)

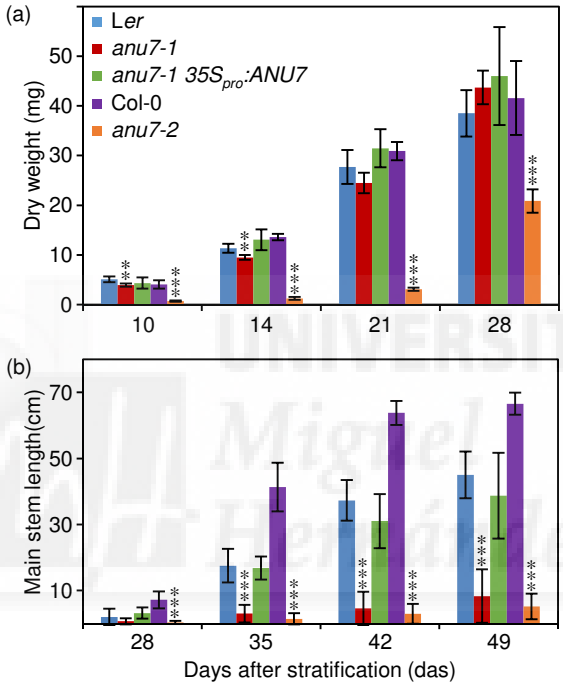


Figure S1. Weight and height of *anu7* mutants. (a) Dry weight and (b) main stem length of L*er*, *anu7-1*, *anu7-1 35S$_{pro}$:ANU7*, Col-0 and *anu7-2* plants. Plants were collected 28, 35, 42 and 49 das. Error bars indicate standard deviations. Asterisks indicate values significantly different from the corresponding wild type in (a) a Mann-Whitney U-test (**p<0.01, ***p<0.001, n=8), and (b) a Student's t-test (***p<0.001, n=12).

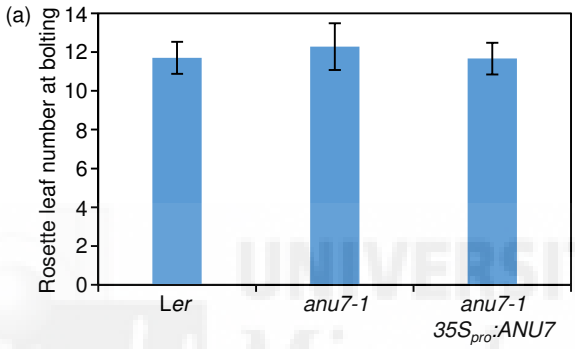


Figure S2. Rosette leaf number at bolting in L*er*, *anu7-1* and *anu7-1 35S$_{pro}$:ANU7* plants. No values significantly different from the corresponding wild type were found in a Student's *t*-test (n=14).

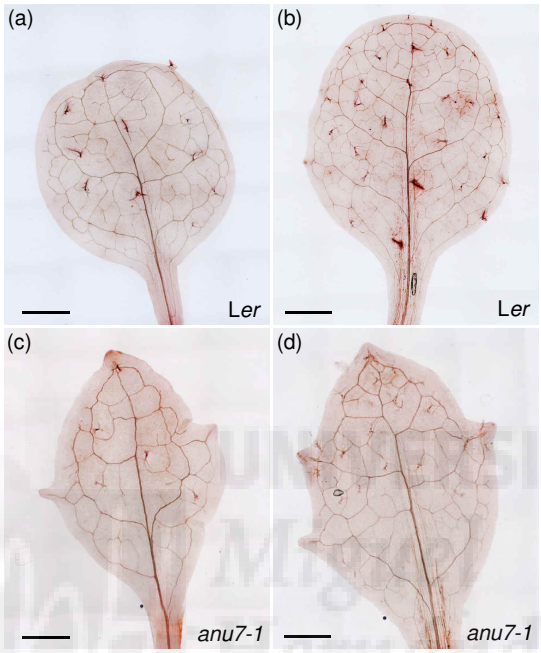


Figure S3. Diaminobenzidine staining of *anu7-1* and wild-type leaves. (a, c) First- and (b, d) third-node leaves from (a, b) L*er* and (c, d) *anu7-1* plants collected 21 das. Scale bars indicate 1 mm.

```
                         .    :    .    :    .    :    .    :    .    :    .    :
BP820279.1+          AGAATAGTGTAAAAACCACTCTTGAAACCTCCTCTTCAAAAATACCAGAGAGCGAGAAAA
DR281354.1+                          CTCTTGAAACCTCCTCTTCAAAAATACCAGAGAGCGAGAAAA
BP857405.1+                          CTCTTGAAACCTCCTCTTCAAAAATACCAGAGAGCGAGAAAA
DR281353.1+                                      CTCTTCAAAAATACCAGAGAGCGAGAAAA
DR210229.1+                                           AAAAATACCAGAGAGCGAGAAAA
DR210241.1+                                                ACCAGAGAGCGAGAAAA
DR210230.1+                                                ACCAGAGAGCGAGAAAA
DR210232.1+                                             AAAAAATACCAGAGAGCGAGAAAA
DR210233.1+                                             AAAAAATACCAGAGAGCGAGAAAA
AV828031.1+                                             AAAAAATACCAGAGAGCGAGAAAA
DR210226.1+                                              AAAAATACCAGAGAGCGAGAAAA
DR210224.1+                                              AAAAATACCAGAGAGCGAGAAAA
DR210223.1+                                             AAAAAATACCAGAGAGCGAGAAAA
DR210239.1+                                             AAAAAATACCAGAGAGCGAGAAAA
DR210227.1+                                              AAAAATACCAGAGAGCGAGAAAA
CB264192.1+                                                        GAAAA
AV825943.1+                                                ACCAGAGAGCGAGAAAA
BP817990.1+                                                 CCAGAGAGCGAGAAAA
DR210234.1+                                                ACCAGAGAGCGAGAAAA
BP840043.1+                                                 CCAGAGAGCGAGAAAA
BP841249.1+                                                 CCAGAGAGCGAGAAAA
BE525173.1+                                                           A
DR210240.1+                                                ACCAGAGAGCGAGAAAA
DR210236.1+                                                ACCAGAGAGCGAGAAAA
DR281355.1+                                                ACCAGAGAGCGAGAAAA

consensus            AGAATAGTGTAAAAACCACTCTTGAAACCTCCTCTTAAAAAATACCAGAGAGCGAGAAAA


                           .    :    .    :    .    :    .    :    .    :    .    :
BP820279.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR281354.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
BP857405.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAG
DR281353.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210228.1+            ATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210229.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210241.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210230.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210232.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210233.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
AV828031.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210226.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210224.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210223.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210239.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210227.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210237.1+                            CCAATGTCGAGAGGACCAGGTBGTTTAATACA-GAAC
CB264192.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
AV825943.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
BP817990.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAG
BP868479.1+                  CTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210234.1+          GCTCTTATCTTCTTCATSTTC-TCSAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
BP840043.1+          GCTCTTATCTTCTTCATCTTCCTCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
BP841249.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
BE525173.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
EL138437.1-                                                CGTTTAATACA-GAAC
DR210225.1+                ATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATAYTWGAAC
DR210240.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
DR210236.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
CB074586.1-                                       GAGGACCAGGTCGTTTAATACA-GAAC
DR281355.1+          GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC

consensus            GCTCTTATCTTCTTCATCTTC-TCCAATGTCGAGAGGACCAGGTCGTTTAATACA-GAAC
```

```
                    .    :    .    :    .    :    .    :    .    :    .    :
BP820279.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR281354.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
BP857405.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR281353.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210228.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210229.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210241.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210230.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACTGCGTTATGGACAACA
DR210232.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210233.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
AV828031.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210226.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210224.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210223.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210239.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210227.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210237.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TSTACT-CGTTATGGACAACA
CB264192.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
AV825943.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
BP817990.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
BP868479.1+         GTGACACAGTTCGGCTGATGCTCAACTTCAAACAATTC-TCTACTCGGTTATGGACAACA
DR210234.1+         GTGACACAGTTSG-CTGATGCTCAA-TTCAAACAATTC-TMTACT-CGTTATGGACAACA
BP840043.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
BP841249.1+         GTGACACAGTTCG-GTGATGCTCAA-TTCAAACAATTCCTCTACT-CGTTATGGACAACA
BE525173.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
EL138437.1-         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210225.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210240.1+         GTGACACAGTTCG-CTGATGSTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
DR210236.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
CB074586.1-         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
BE525348.1+               GTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
AV524787.1+                     GCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACTACA
DR281355.1+         GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA
                    ____________________________________________________________
consensus           GTGACACAGTTCG-CTGATGCTCAA-TTCAAACAATTC-TCTACT-CGTTATGGACAACA


                    .    :    .    :    .    :    .    :    .    :    .    :
BP820279.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR281354.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
BP857405.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR281353.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210228.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210229.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210241.1+         AGTTATAGACATTCTCGATATTTCCAATCAAGSWCGWCTTGTCTCCTTWCACTCTCGCCT
DR210230.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210232.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210233.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTSACTCTCGCCT
AV828031.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210226.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210224.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210223.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210239.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210227.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210237.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTSTCCTTTCANTCTCGSCT
CB264192.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
AV825943.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
BP817990.1+         AGTTATAGACATTCTCGAT-TTTC-AATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
BP868479.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210234.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTMTTGTMTCCTTTCACTCTCGCCT
BP840043.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
BP841249.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
BE525173.1+         AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
EL087060.1-                           CCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
EL138437.1-         AGTTATAGACATTCTCGAT-TTTCCAATCAAGC
```

```
ES134717.1-                                                    CCTTTCACTCTCGCCT
DR210225.1+          AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210240.1+          AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
DR210236.1+          AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
CB074586.1-          AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
BE525348.1+          AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
AV524787.1+          AGCTATAGACATTCTCGAT-TCTCCAATCAAGCTAGTGTTGGCTACTTGCACTCTCGCCT
DR281355.1+          AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT
                     ____________________________________________________________
consensus            AGTTATAGACATTCTCGAT-TTTCCAATCAAGCTCGTCTTGTCTCCTTTCACTCTCGCCT


                         .    :    .    :    .    :    .    :    .    :    .    :
BP820279.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTGGGCATCCCTGAG-T-CATCTCCAAAA
DR281354.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
BP857405.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTGGGCATCCCTGAG-TTCATCTCCAAAA
DR281353.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
CF651819.1+                                                CCTGAG-TTCATCTCCAAAA
DR210228.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210229.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210241.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210230.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210232.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210233.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
AV828031.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210226.1+          TTGATATCGC-TGGCTCTGGCACCTCGNGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210224.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210223.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210239.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210227.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210237.1+          TTGATATCGS-TGGMTCTG-CACCTCGTGGTTTTGGVATCCCTGAG-TTCATSTCCAAAA
CB264192.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
AV825943.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
EH922670.1-                                                              CTCCAAAA
BP817990.1+          TTGATATCGCCTGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAGCTTCATCTCCAAAA
BP868479.1+          TTGATATCGG-TGGCTCTG-CACCTCGTGGTTTGGGCATCCCTGAG-TTCATCTCCAAAA
DR210234.1+          TTGATATCGC-TGGCTCTG-CACCTMGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
BP840043.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTGGGCATCCCTGAG-T-CATCTCCAAAA
BP841249.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
BE525173.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
EL087060.1-          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
ES134717.1-          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210225.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210240.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
DR210236.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
CB074586.1-          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
BE525348.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAA-TTCATCTCCAAAA
AV524787.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGACATCCCTGAG-TTCATCTCCAAAA
DR281355.1+          TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA
CF651820.1+                                                CCTGAG-TTCATCTCCAAAA
                     ____________________________________________________________
consensus            TTGATATCGC-TGGCTCTG-CACCTCGTGGTTTTGGCATCCCTGAG-TTCATCTCCAAAA


                         .    :    .    :    .    :    .    :    .    :    .    :
BP820279.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR281354.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
BP857405.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR281353.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
CF651819.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210228.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210229.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210241.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210230.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210232.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210233.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
```

```
AV828031.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210226.1+          TTTCGTATCTCTCTGTTTTTGCGGNTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210224.1+          TTTCGTATCTCTCTGTTTTTGCGGNTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210223.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTAAGACATTGCATTGGATT
DR210239.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210227.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210237.1+          TTTCGTATYTNTCTGTTTTTGSGGTTGCTACGCTTGGAACTTATGAMATTGCATTGGATT
CB264192.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
AV825943.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
EH922670.1-          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
EL280686.1+             CGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
BP817990.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
BP868479.1+          TTTCGTATCTCTCTGTTTTGGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210234.1+          TTTCGTATCTCTMTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
BP840043.1+          TTTCGTATCTCTCTGTTTTGGCGGTTGCTACGCTTGGAACTTATGACATTGGATTGGATT
BP841249.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
BE525173.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGACT
EL087060.1-          TTTCGT
ES134717.1-          TTTCGTATCTCTCTGTTTTTGCGGTTGCTAC
DR210225.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210240.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR210236.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGASATTGCATTGGATT
CB074586.1-          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
BE525348.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
AV524787.1+          TTGCGTATCTCTCTGCATTTGCGGCTGCTACGCTTGGAGCTTATGACATTGCATTGGATT
DR281355.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
CF651820.1+          TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
DR281356.1+               TATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT
                     ____________________________________________________________
consensus            TTTCGTATCTCTCTGTTTTTGCGGTTGCTACGCTTGGAACTTATGACATTGCATTGGATT


                              .    :    .    :    .    :    .    :    .    :
BP820279.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGC
DR281354.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
BP857405.1+          -GGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGT
DR281353.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
CF651819.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210228.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210229.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210241.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210230.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210232.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210233.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
AV828031.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210226.1+          TGG
DR210224.1+          TGGGNAAAAAAAGCATATGCCAAAGGGATTGC
DR210223.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCNTTA
DR210239.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210227.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210237.1+          TGGGAAAGAA-GTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
CB264192.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
AV825943.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
EH922670.1-          TGGGAAAGAAAGTCNTATGCCAAAGGGATTGCAAAAC-TTGT
EL280686.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGG
BP817990.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAACCTTGTAATGGGTGGCAGGCATTA
BP868479.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210234.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
BP840043.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
BP841249.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
BE525173.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210225.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210240.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR210236.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
CB074586.1-          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
BE525348.1+          TGGGAAAGAAAGTCATATGCCAAAGG
```

```
AV524787.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGACAAGCATTA
DR281355.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
CF651820.1+          TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
DR281356.1+          TGGGAAAGAAAGTCATATGCCAAAGGGAATGCAAAAC-TTGTAATGGGTGGCAGGCATTA
CK117674.1+                AGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA
                     ____________________________________________________________
consensus            TGGGAAAGAAAGTCATATGCCAAAGGGATTGCAAAAC-TTGTAATGGGTGGCAGGCATTA


                         .    :    .    :    .    :    .    :    .    :    .    :
DR281354.1+          CGGTGCACTATGTSRAA-GGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR281353.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
CF651819.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210228.1+          CGGTGCACTATGTS-AAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210229.1+          CGGTGCACTATGTS-AAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210241.1+          CGGTGCACTATGTS-AAAGGAACAGGGAGTGTTCATTACCAAATCAMAGATTACAACTTG
DR210230.1+          CGGTGCACTATGTN-AAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210232.1+          CGGTGCACTATGTS-AAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210233.1+          CGGTGCACTATGTS-AAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
AV828031.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210223.1+          C
DR210239.1+          CGGTGCACTATGTSAAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTAC
DR210227.1+          CGGTGCACTATGTNAAAGGGAACAGGGAGTGTTCATTACCAAATCAAAG
DR210237.1+          CGGTGCACTATGTCAAAGG
CB264192.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
AV825943.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
BP817990.1+          CGGTGCACTATGTGCAAAGG
BP868479.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTAC
DR210234.1+          CGGTGCACTATGT
BP840043.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAAT
BP841249.1+          CGGTGCACTATGTGCAAAGGAACAG
BE525173.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210225.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210240.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR210236.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
CB074586.1-          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
AV524787.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR281355.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
CF651820.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
DR281356.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
CK117674.1+          CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG
                     ____________________________________________________________
consensus            CGGTGCACTATGTGCAAAGGAACAGGGAGTGTTCATTACCAAATCAAAGATTACAACTTG


                         .    :    .    :    .    :    .    :    .    :    .    :
DR281354.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGA
DR281353.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
CF651819.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
DR210228.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
DR210229.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
DR210241.1+          AGAAGTGGAGAGAAACCANCAGCAGATTGTGTCGCAGATGCTATAGTTG
DR210230.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGAT
DR210232.1+          AGAAGTGGAGAGAAACCACCAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGC
DR210233.1+          AGAAGTGGAGAGAAACC
AV828031.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCANATGCTATAGTTGAGAATCGAGCT
CB264192.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
AV825943.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
BE525173.1+          AGAAGTGGAGAGAAAC
DR210225.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGT
DR210240.1+          AGAAGTGGAGAGAAAC
DR210236.1+          AGAAGTGGAGAGAAACCANCAGCAGATTGTGTCGCAGATGCTATAGTTGAG
CB074586.1-          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
AV524787.1+          AGAAGTGGAGAGAAACCAACAGCAGATCGTGTCGCAGATGCTAT
DR281355.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
CF651820.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
```

```
DR281356.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT
CK117674.1+          AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT

consensus            AGAAGTGGAGAGAAACCAACAGCAGATTGTGTCGCAGATGCTATAGTTGAGAATCGAGCT


                        .    :    .    :    .    :    .    :    .    :    .    :
DR281353.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
CF651819.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
DR210228.1+          GAGTTAGTTCATCTTCCCTCC
DR210229.1+          GAGTTAGTTCATCTTCC
AV828031.1+          GAGTTAGTTCATCTTCCCTCCTNCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
CB264192.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
AV825943.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
CB074586.1-          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
DR281355.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
CF651820.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
BP815201.1+                GTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
BP818510.1+                GTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGG
DR281356.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC
CK117674.1+          GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC

consensus            GAGTTAGTTCATCTTCCCTCCTCCTTCAACCATAGTGCTCCATTGCCATCAAAAGACTGC


                           .    :    .    :    .    :    .    :    .    :    .    :
DR281353.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
CF651819.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
AV828031.1+          CCAACCTTGTGATGGAACGGGTGCGATGAGC
CB264192.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
AV825943.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
CB074586.1-          CCAACTTG
DR281355.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
CF651820.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
BP815201.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
BP818510.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
DR281356.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT
CK117674.1+          CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT

consensus            CCAAC-TTGTGATGGAACGGGTGCGATGAGCTGTACTGAGTGCAAGAACAAGCTGCAAGT


                           .    :    .    :    .    :    .    :    .    :    .    :
DR281353.1+          CAGAATCTCAGCTGATGATATCATGGGACCTCCATGGAAAGCCTATAATGTGCTGAAGAA
CF651819.1+          CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA
CB264192.1+          CAGAATCTCAGCTGATGATATCATGGGACCT
AV825943.1+          CAGAATCTCAGCTGATGATATCATGGA
DR281355.1+          CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA
CF651820.1+          CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA
BP815201.1+          CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA
BP818510.1+          CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA
DR281356.1+          CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA
CK117674.1+          CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA

consensus            CAGAATCTCAGCTGATGATATCATGGAACCTCCATGGAAAGCCTATAATGTGCTGAAGAA


                           .    :    .    :    .    :    .    :    .    :    .    :
DR281353.1+          GATGGATTATCCGTACGAGCACATTGTTCACAGCATGAAAGATCCCAGCATTGCAAAT
CF651819.1+          GATGGATTATCCGTACGAGCACATTGT
DR281355.1+          GATGGATTATCCGTACGAGCACATTGTTC
CF651820.1+          GATGGATTATCCGTACGAGCACATTGTTCACAGCATGAAAGATCCCAGCATTGCAAATTT
BP815201.1+          GATGGATTATCCGTACGAGCACATTGTTCACAGCATGAAAGATCCCAGCATTGCAAATTC
BP818510.1+          GATGGATTATCCGTACGAGCACATTGTTCACAGCATGAAAGATCCCAGCATTGCAAATTT
CB074418.1+                                        CATGAAAGATCCCAGCATTGCAAATTT
DR281356.1+          GATGGATTATCCGTACGAGCACATTGTTCACAGCATGAAAGATCCCAGCATTGCAAATTT
```

```
CK117674.1+          GATGGATTATCCGTACGAGCACATTGTTCACAGCATGAAAGATCCCAGCATTGCAAATTT
                     ____________________________________________________________
consensus            GATGGATTATCCGTACGAGCACATTGTTCACAGCATGAAAGATCCCAGCATTGCAAATTT


                        .    :    .    :    .    :    .    :    .    :    .    :
CF651820.1+          CTGGTTGATCACTTTGCCTCAGATTGTGGGTGGGTTTGATTATGATGAAGATGTCAAGAA
BP815201.1+          CTGGTTGATCACTTTGCCTCAGATTGTGGGTGGGTTTGATTATGATGAAGATGTCAAGAA
BP818510.1+          CTGGTTGATCACTTTGCCTCAGATTGTGGGTGGGTTTGATTATGATGAAGATGTCAAGAA
CB074418.1+          CTGGTTGATCACTTTGCCTCAGATTGTGGGTGGGTTTGATTATGATGAAGATGTCAAGAA
DR281356.1+          CTGGTTG
CK117674.1+          CTGGTTGATCACTTTGCCTCAGATTGTGGGTGGGTTTGATTATGATGAAGATGTCAAGAA
                     ____________________________________________________________
consensus            CTGGTTGATCACTTTGCCTCAGATTGTGGGTGGGTTTGATTATGATGAAGATGTCAAGAA


                        .    :    .    :    .    :    .    :    .    :    .    :
CF651820.1+          GAAAATATGGTGGCAATATGAGGAATCTATGCGATATGATCAACTTCGGGACTTGGTGGC
BP815201.1+          GAAAATATGGTGGCAATATGAGGAATCTATGCGATATGATCAACTTCGGGACTTGGTGGC
BP818510.1+          GAAAATATGGTGGCAATATGAGGAATCTATGCGATATGATCAACTTCGGGACTTGGTGGC
CB074418.1+          GAAAATATGGTGGCAATATGAGGAATCTATGCGATATGATCAACTTCGGGACTTGGTGGC
CK117674.1+          GAAAATATGGTGGCAATATGAGGAATCTATGCGATATGATCAACTTCGGGACTTGGTGGC
DR373600.1+                GCAATATGAGGAATCTATGCGATATGATCAACTTCGG-ACTTGGTG-C
                     ____________________________________________________________
consensus            GAAAATATGGTGGCAATATGAGGAATCTATGCGATATGATCAACTTCGGGACTTGGTGGC


                        .    :    .    :    .    :    .    :    .    :    .    :
CF651820.1+          TAAACGCAATCCTGGCTGGGAATATTTGCAGGATGCCTTGGTCTCCATTGATCCTGTCCG
BP815201.1+          TAAACGCAATCCTGGCTG
BP818510.1+          TAAACGCAATCCTGGCTGGGAATATTTG
CB074418.1+          TAAACGCAATCCTGGCTGGGAATATTTGCAGGATGCCTTGGTCTCCATTGATCCTGTCCG
ES157359.1+                ATCCTGGCTGGGAATATTTGCAGGANGCCTTGGTCTCCATTGATCCTGTCCG
CK117674.1+          TAAACGCAATCCTGGCTGGGAATATTTGCAGGATGCCTTGGTCTCCATTGATCCTGTCCG
DR373600.1+          TAAACGCAATCCTGGCTGGGAATATTTGCAGGATGCCTTGGTCTCCATTGATCCTGTCCG
                     ____________________________________________________________
consensus            TAAACGCAATCCTGGCTGGGAATATTTGCAGGATGCCTTGGTCTCCATTGATCCTGTCCG


                        .    :    .    :    .    :    .    :    .    :    .    :
CF651820.1+          TGCTCGGGAAGACCCCGTCATAGTGAAGAATGTCCCTTATTACAAGGCTAAGAAATCTCT
CB074418.1+          TGCTCGGGAAGACCCCGTCATAGTGAAGAATGTCCCTTATTACAAGGCTAAGAAATCTCT
ES157359.1+          TGCTCGGGAAGACCCCGTCATAGTGAAGAATGTCCCTTATTACAAGGCTAA
CK117674.1+          TGCTCGGGAAGACCCCGTCATAGTGAAGAATGTCCCTTATTACAAGCCTAAGAAATCTCT
DR373600.1+          TGCTCGGGAAGACCCCGTCATAGTGAAGAATGTCCCTTATTACAAAGCTAAGAAATCTCT
                     ____________________________________________________________
consensus            TGCTCGGGAAGACCCCGTCATAGTGAAGAATGTCCCTTATTACAAGGCTAAGAAATCTCT


                        .    :    .    :    .    :    .    :    .    :    .    :
CF651820.1+          GGAGGCTGAAGTCACGAAACTCAATCCACCTCCCCGTCCTCAGAATTGGGGTGAACTGAA
ES122212.1-             GGCTGAAGTCACGAAACTCAATCCACCTCCCCGTCCTCAGAATTGGG-TGAACTGAA
CB074418.1+          GGAGGCTGAAGTCACGAAACTCAATCCGCCTCCCCGTCCTCAGAATTGGGGTGAACTGAA
CK117674.1+          GGAGGCTGAAGTCACGAAACTCAATCCACCTCCCCGTCCTCAGAATTGGGGTGAACTGAA
DR373600.1+          GGAGGCTGAAGTCACGAAACTCAATCCACCTCCCCGTCCTCAGAATTGGGGTGAACTGAA
DR210235.1+                           AATCCACCTCCCCGTCCTCAGAATTGGGGTGAACTGAA
AV530947.1-                                 CCCCGTCCTCAGAATTGGGGTGAACTGAA
                     ____________________________________________________________
consensus            GGAGGCTGAAGTCACGAAACTCAATCCACCTCCCCGTCCTCAGAATTGGGGTGAACTGAA


                        .    :    .    :    .    :    .    :    .    :    .    :
CF651820.1+          TCTCCCATTGAATACTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGCAAAACTGTA
ES122212.1-          TCTCCCATTGAATACTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGCAAAACT
CB074418.1+          TCTCCCATTGAATATTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGCAAAACTGT
CK117674.1+          TCTCCCATTGAATA
```

```
DR373600.1+          TCTCCCATTAAATACTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGCAAAACTGTA
DR210235.1+          TCTCCCATTAAATACTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGCAAAACTGTA
AV530947.1-          TCTCCCATTGAATATTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGCAAAACTGTA
EL199196.1+                  TTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGNAAAACTGTA
ES185124.1+                                                AAAACCCGGCAAAACTGTA
EL189532.1+                                      GAAGAGGATCTCAAAAACCCGGCAAAACTGTA
BP644655.1-        ATTGAATATTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGCCAAAACTGTA
DR354773.1+                                      AAGAGGATCTCAAAAACCCGGCAAAACTGTA
DR354772.1+                                      AAGAGGATCTCAAAAACCCGGCAAAACTGTA
                     ____________________________________________________________
consensus            TCTCCCATTGAATACTTCCTCTTGGAGTGAAGAGGATCTCAAAAACCCGGCAAAACTGTA


                         .    :    .    :    .    :    .    :    .    :    .    :
CF651820.1+          CGAGAAGACAGTTCTT
DR373600.1+          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
DR210235.1+          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
AV530947.1-          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
EL199196.1+          CGAGAAGACAGTTCTT
EL261543.1-                                                              CACA
ES185124.1+          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
ES094189.1-                                                 GACAAAATCTTGGATGCACA
EL189532.1+          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
EL305669.1-                                CAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
CB074314.1+                GTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
AV786339.1-                                AAGAGAAATCGCAGACAAAATCTTGGATGCACA
BP778443.1-                                                           GGATGCACA
AV787051.1-                                                            ATGCACA
BP644655.1-          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCCCAGACAAAATCTTGGATGCACA
AV793206.1-                GTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
AV793011.1-             AAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
AV533263.1-               GACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
DR354773.1+          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
DR354772.1+          CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA
AV794205.1-                              AAAGAGAAATCGCAGACAAAATCTTGGATGCACA
EH852490.1+                                                   AAATCTTGGATGCACA
                     ____________________________________________________________
consensus            CGAGAAGACAGTTCTTCTCAATGCTCAAAGAGAAATCGCAGACAAAATCTTGGATGCACA


                         .    :    .    :    .    :    .    :    .    :    .    :
DR373600.1+          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
DR210235.1+          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
AV530947.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
EL261543.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
ES185124.1+          ATGGGAA
EL025397.1+                                                   AGCAGAAAGTGAGAC
ES094189.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
EL189532.1+          ATGGGAAG
EL305669.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
BP649293.1-                                      GGAGGAAATGTTGGAGCAGAAAGTGAGAC
CB074314.1+          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
BP781232.1-               AAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
AV786339.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
BP781929.1-                  GGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
BP778443.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
AV787051.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
BP785897.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
BP644655.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
AV793206.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
AV793011.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
AV533263.1-          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
DR354773.1+          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
DR354772.1+          ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
AV794205.1-          ATGGGAAGGCTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
EH852490.1+          ATGGGAAGCTTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC
```

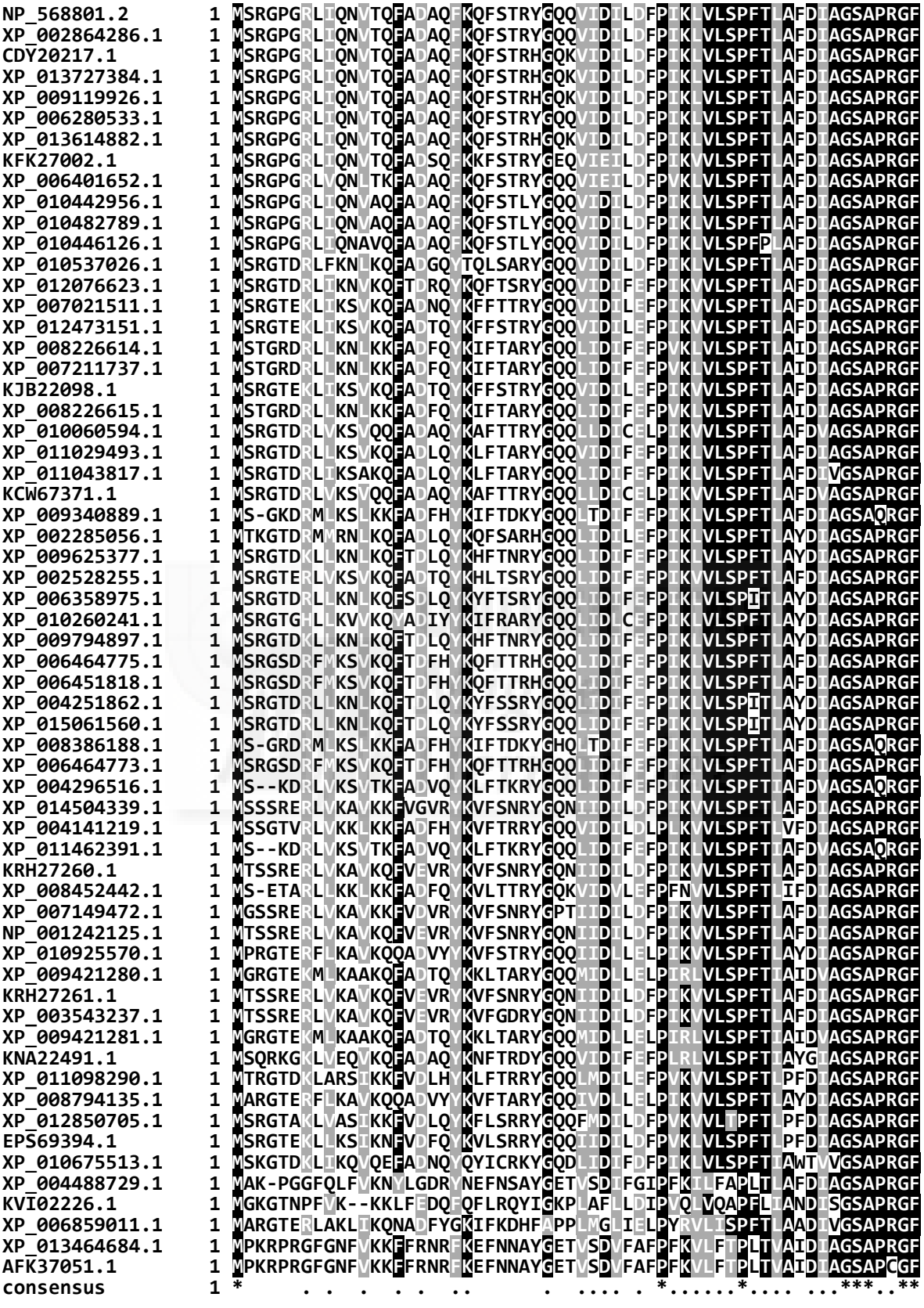
```
consensus             ATGGGAAG-CTAAATGGCGTCAAGAGAAGGTGGAGGAAATGTTGGAGCAGAAAGTGAGAC


                          .    :    .    :    .    :    .    :    .    :    .    :
DR373600.1+           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
DR210235.1+           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
AV530947.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
EL261543.1-           CTTACATCCAAGAACTCAAGCATGGCTGTTCTTCCACAACC
EL025397.1+           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
ES094189.1-           CTTACATCCAAGA-CTCAAGC
EL305669.1-           CTTACATC
DR210238.1+                                                                      G
BP649293.1-           CTTACATCCAAGA-CTCAAGCATGGCTTTTCTTCCACAACCCATCTTGTTAAAGTCACAG
CB074314.1+           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
BP781232.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
AV786339.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
BP781929.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
BP778443.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
AV787051.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
BP785897.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
BP644655.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
AV793206.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
AV793011.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
AV533263.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
DR354773.1+           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
DR354772.1+           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
AV794205.1-           CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG
EH852490.1+           CTTACATCCAAGA-CTCAAGCATGG

consensus             CTTACATCCAAGA-CTCAAGCATGGCTGTTCTTCCACAACCCATCTTGTTAAAGTCACAG


                          .    :    .    :    .    :    .    :    .    :    .    :
DR373600.1+           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
DR210235.1+           AAGAAAGCTCAAAAGGGAAGCCGNNNAAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
AV530947.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
EL025397.1+           AAGAAAGCTCAAAA
EL180847.1-                                                                CCAAGTG
EH952171.1-                  CTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
DR210238.1+           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
BP648872.1-              AAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
BP649293.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTCCCAAGTG
CB074314.1+           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
BP781232.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
AV786339.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTG-TTCTTCTAACCTGCCAAGTG
BP781929.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
BP778443.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
AV787051.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
BP785897.1-           AAGAAAGCTCAAGAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
BP644655.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
AV793206.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
AV793011.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
AV533263.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
DR354773.1+           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
DR354772.1+           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG
AV794205.1-           AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG

consensus             AAGAAAGCTCAAAAGGGAAGCCGGC-AAAGGAAGTGGTGGTTCTTCTAACCTGCCAAGTG


                          .    :    .    :    .    :    .    :    .    :    .    :
DR373600.1+           TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
DR210235.1+           TGTCGAAACGAATCTATATAAAATGGATTTTGAATNNNAGTTTTTTTGTGCTTCTTTGAT
AV530947.1-           TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
ES043916.1+                                AATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
```

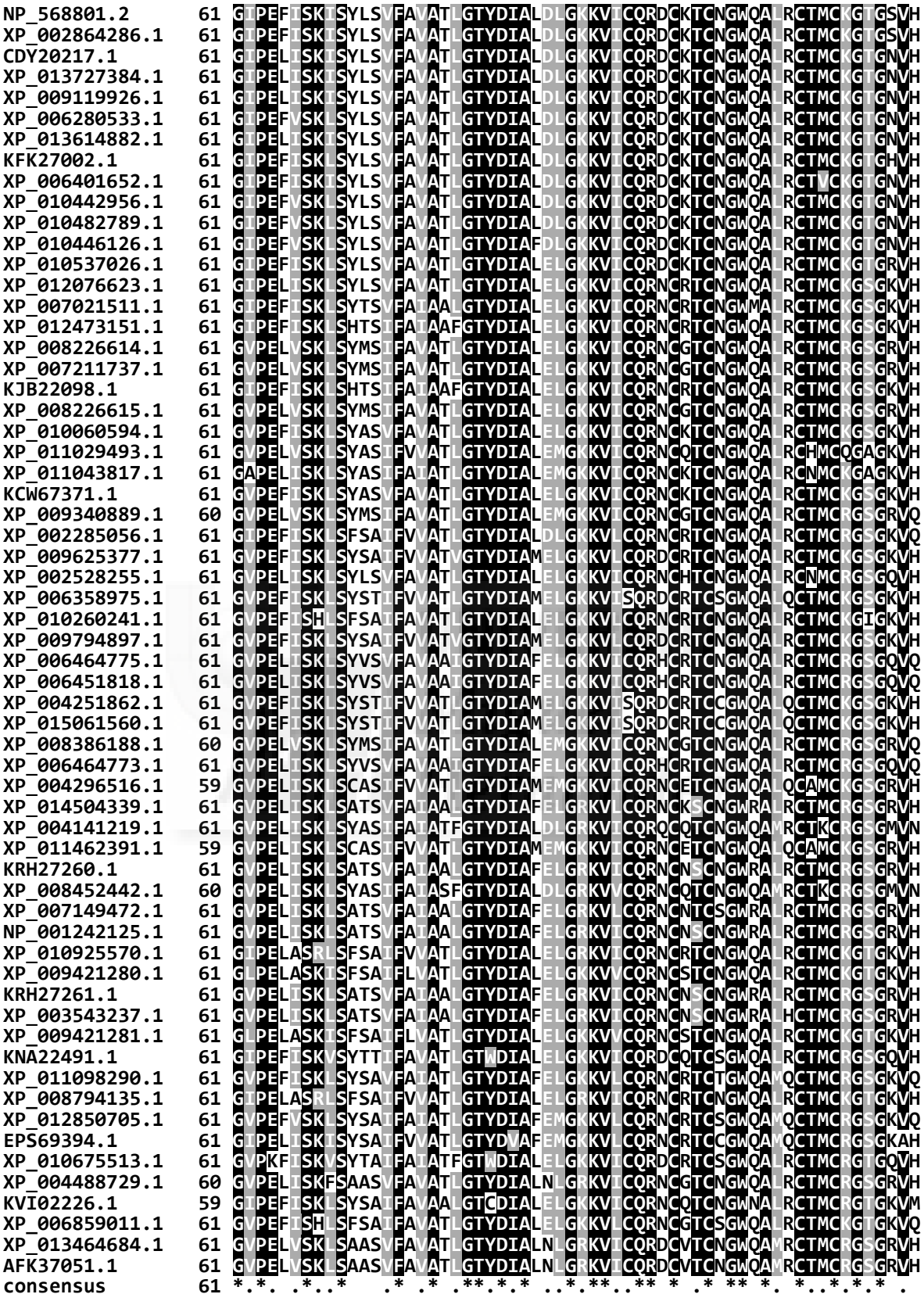
```
EL180847.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
EH952171.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
DR210238.1+          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
BP648872.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
BP649293.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
CB074314.1+          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
BP781232.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
AV786339.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
BP781929.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
BP778443.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
AV787051.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
BP785897.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
BP644655.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
AV793206.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
AV793011.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
AV533263.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
DR354773.1+          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
DR354772.1+          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
AV794205.1-          TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT
AA721892.1+                                        AATTG-AGTTTTTTTGTGCTTCTTTGAT

consensus            TGTCGAAACGAATCTATATAAAATGGATTTTGAATTG-AGTTTTTTTGTGCTTCTTTGAT


                          .    :    .    :    .    :    .    :    .    :    .    :
DR373600.1+          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATC
DR210235.1+          GAAAATATGCTTTGATATGAAATCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AV530947.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
ES043916.1+          GAAA-TATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
EL180847.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAAT
EH952171.1-          GAAAAT
DR210238.1+          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
BP648872.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
BP649293.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
CB074314.1+          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
BP781232.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AV786339.1-          GAAAATATGCTTTGATATGAA-TCTTGATGG-TGTCAAATATGATTACACTTCAAAATCA
BP781929.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
BP778443.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AV787051.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
BP785897.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
BP644655.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AV793206.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AV793011.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AV533263.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
DR354773.1+          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
DR354772.1+          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AV794205.1-          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA
AA721892.1+          GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA

consensus            GAAAATATGCTTTGATATGAA-TCTTGATGGGTGTCAAATATGATTACACTTCAAAATCA


                            .    :    .    :    .    :    .    :    .    :    .    :
DR210235.1+          G-TTGTAATGTTTTAAAAGAAA
AV530947.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
ES043916.1+          G-TTG
DR210238.1+          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGAC
BP648872.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGCCAGATTCGTTGTGTAC
BP649293.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
CB074314.1+          G-TTGTAATGTTTTAAAAAAAATGTTGTTTTTGC
BP781232.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
AV786339.1-          GGTTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
BP781929.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
BP778443.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
AV787051.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
```

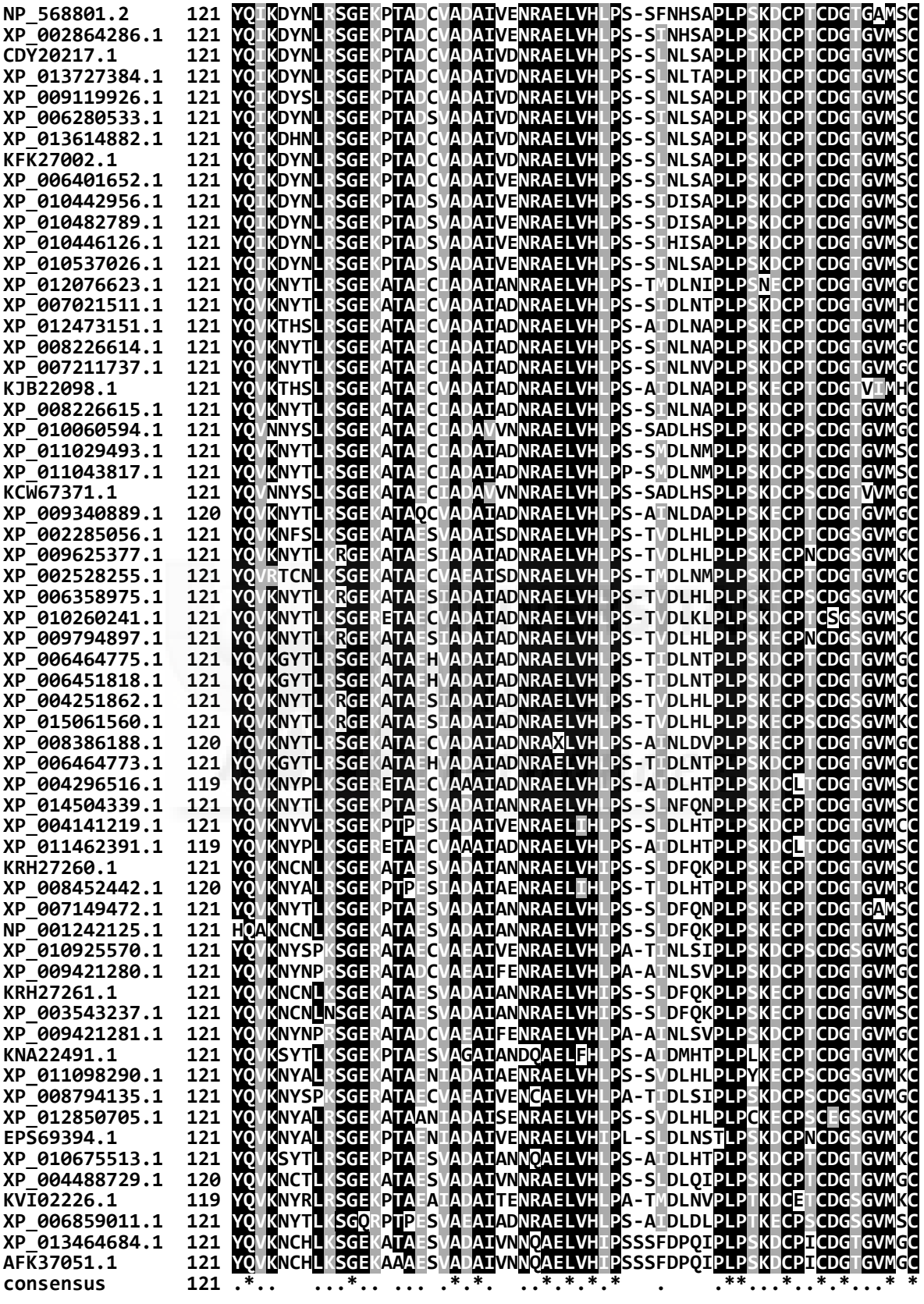
```
BP785897.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
BP644655.1-          G-TTGTAATGTTTT
AV793206.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
AV793011.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
AV533263.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
DR354773.1+          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
DR354772.1+          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
AV794205.1-          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
AA721892.1+          G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC
                     ____________________________________________________________
consensus            G-TTGTAATGTTTTAAAAGAAATGTTGTTTTTGCAGACACTTGGCAGATTCGTTGTGTAC


                        .    :    .    :    .    :    .    :    .    :    .    :
AV530947.1-          CGCCACGTGTATACGATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTCTTTCTCCT
BP648872.1-          CGCCACGTGTATACGATGATCTTTCTT
BP649293.1-          CCCCACGTGTATACCATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTCTTTCTCC
BP781232.1-          CGCCACGTGTATACGATGATCTTTCTTTC
AV786339.1-          CGCCACGTGTATACGATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTC
BP781929.1-          CCCCACGTGTATACGATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTC
BP778443.1-          CGCCACGTGTATACGATGATCTTTCTT
AV787051.1-          CGCCACGTGTATACGATGATCTTTCTTTC
BP785897.1-          CGCCACGTGTATACGATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTCTTTCTCCT
AV793206.1-          CGCCACGTGTATACGATGATCTTTCTTTC
AV793011.1-          CGCCACGTGTATACGATGATCTTTCTT
AV533263.1-          CGCCACGTGTATACGATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTCT
DR354773.1+          CGCCACGTGTATACGATGATCTTTCTTT
DR354772.1+          CGCCACGTGTATACGATGATCTTTCTTTC
AV794205.1-          CGCCACGTGTATACGATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTCTTTCTCCT
AA721892.1+          CGCCACGTGTATACGATGATCTTNCTTTCATGCTTTGAACTTAGTTTTCGTCTTTCTCCT
                     ____________________________________________________________
consensus            CGCCACGTGTATACGATGATCTTTCTTTCATGCTTTGAACTTAGTTTTCGTCTTTCTCCT


                        .    :    .    :    .    :    .    :    .    :    .    :
AA721892.1+          ATCAATCACTCTCTCTAT
                     ____________________________________________________________
consensus            ATCAATCACTCTCTCTAT
```

Figure S4. CAP3 assembly of *ANU7* expressed sequence tags.

(a)

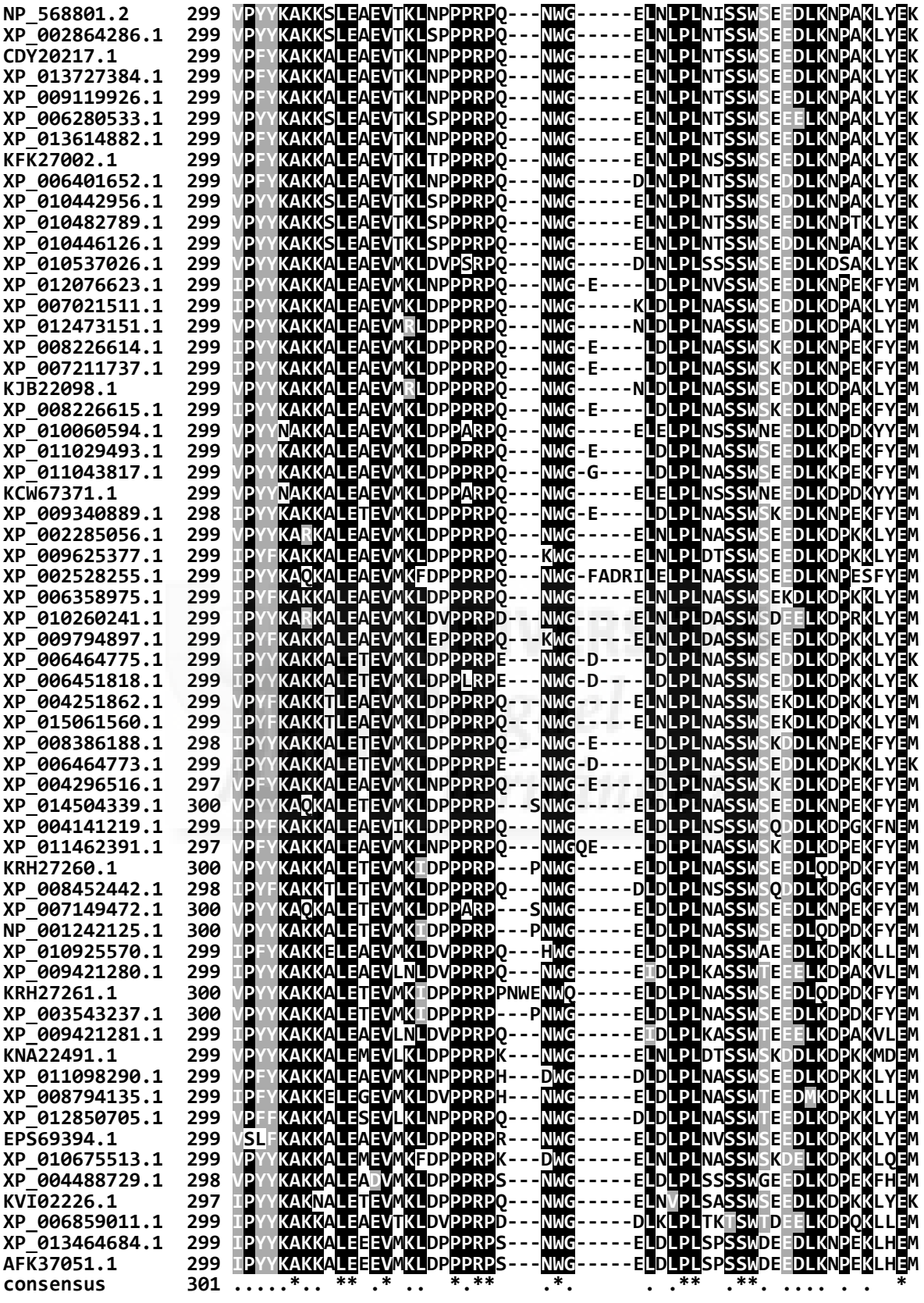

NP_568801.2 1 MSRGPGRLIQNVTQFADAQFKQFSTRYGQQVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_002864286.1 1 MSRGPGRLIQNVTQFADAQFKQFSTRYGQQVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
CDY20217.1 1 MSRGPGRLIQNVTQFADAQFKQFSTRHGQKVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_013727384.1 1 MSRGPGRLIQNVTQFADAQFKQFSTRHGQKVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_009119926.1 1 MSRGPGRLIQNVTQFADAQFKQFSTRHGQKVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_006280533.1 1 MSRGPGRLIQNVTQFADAQFKQFSTRYGQQVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_013614882.1 1 MSRGPGRLIQNVTQFADAQFKQFSTRHGQKVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
KFK27002.1 1 MSRGPGRLIQNVTQFADSQFKKFSTRYGEQVIEILDFPIKVVLSPFTLAFDIAGSAPRGF
XP_006401652.1 1 MSRGPGRLVQNLTKFADAQFKQFSTRYGQQVIEILDFPVKLVLSPFTLAFDIAGSAPRGF
XP_010442956.1 1 MSRGPGRLIQNVAQFADAQFKQFSTLYGQQVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_010482789.1 1 MSRGPGRLIQNVAQFADAQFKQFSTLYGQQVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_010446126.1 1 MSRGPGRLIQNAVQFADAQFKQFSTLYGQQVIDILDFPIKLVLSPFPLAFDIAGSAPRGF
XP_010537026.1 1 MSRGTDRLFKNLKQFADGQYTQLSARYGQQVIDILDFPIKLVLSPFTLAFDIAGSAPRGF
XP_012076623.1 1 MSRGTDRLIKNVKQFTDRQYKQFTSRYGQQVIDIFEFPIKVVLSPFTLAFDIAGSAPRGF
XP_007021511.1 1 MSRGTEKLIKSVKQFADNQYKFFTTRYGQQVIDILEFPIKVVLSPFTLAFDIAGSAPRGF
XP_012473151.1 1 MSRGTEKLIKSVKQFADTQYKFFSTRYGQQVIDILEFPIKVVLSPFTLAFDIAGSAPRGF
XP_008226614.1 1 MSTGRDRLLKNLKKFADFQYKIFTARYGQQLIDIFEFPVKLVLSPFTLAIDIAGSAPRGF
XP_007211737.1 1 MSTGRDRLLKNLKKFADFQYKIFTARYGQQLIDIFEFPVKLVLSPFTLAIDIAGSAPRGF
KJB22098.1 1 MSRGTEKLIKSVKQFADTQYKFFSTRYGQQVIDILEFPIKVVLSPFTLAFDIAGSAPRGF
XP_008226615.1 1 MSTGRDRLLKNLKKFADFQYKIFTARYGQQLIDIFEFPVKLVLSPFTLAIDIAGSAPRGF
XP_010060594.1 1 MSRGTDRLVKSVQQFADAQYKAFTTRYGQQLLDICELPIKVVLSPFTLAFDVAGSAPRGF
XP_011029493.1 1 MSRGTDRLLKSVKQFADLQYKLFTARYGQQVIDIFEFPIKLVLSPFTLAFDIAGSAPRGF
XP_011043817.1 1 MSRGTDRLIKSAKQFADLQYKLFTARYGQQLIDIFEFPIKLVLSPFTLAFDIVGSAPRGF
KCW67371.1 1 MSRGTDRLVKSVQQFADAQYKAFTTRYGQQLLDICELPIKVVLSPFTLAFDVAGSAPRGF
XP_009340889.1 1 MS-GKDRMLKSLKKFADFHYKIFTDKYGQQLTDIFEFPIKLVLSPFTLAFDIAGSAQRGF
XP_002285056.1 1 MTKGTDRMMRNLKQFADLQYKQFSARHGQQLIDILEFPIKLVLSPFTLAYDIAGSAPRGF
XP_009625377.1 1 MSRGTDKLLKNLKQFTDLQYKHFTNRYGQQLIDIFEFPIKLVLSPFTLAYDIAGSAPRGF
XP_002528255.1 1 MSRGTERLVKSVKQFADTQYKHLTSRYGQQLIDIFEFPIKVVLSPFTLAFDIAGSAPRGF
XP_006358975.1 1 MSRGTDRLLKNLKQFSDLQYKYFTSRYGQQLIDIFEFPIKLVLSPITLAYDIAGSAPRGF
XP_010260241.1 1 MSRGTGHLLKVVKQYADIYYKIFRARYGQQLIDLCEFPIKLVLSPFTLAYDIAGSAPRGF
XP_009794897.1 1 MSRGTDKLLKNLKQFTDLQYKHFTNRYGQQLIDIFEFPIKLVLSPFTLAYDIAGSAPRGF
XP_006464775.1 1 MSRGSDRFMKSVKQFTDFHYKQFTTRHGQQLIDIFEFPIKLVLSPFTLAFDIAGSAPRGF
XP_006451818.1 1 MSRGSDRFMKSVKQFTDFHYKQFTTRHGQQLIDIFEFPIKLVLSPFTLAFDIAGSAPRGF
XP_004251862.1 1 MSRGTDRLLKNLKQFTDLQYKYFSSRYGQQLIDIFEFPIKLVLSPITLAYDIAGSAPRGF
XP_015061560.1 1 MSRGTDRLLKNLKQFTDLQYKYFSSRYGQQLIDIFEFPIKLVLSPITLAYDIAGSAPRGF
XP_008386188.1 1 MS-GRDRMLKSLKKFADFHYKIFTDKYGHQLTDIFEFPIKLVLSPFTLAFDIAGSAQRGF
XP_006464773.1 1 MSRGSDRFMKSVKQFTDFHYKQFTTRHGQQLIDIFEFPIKLVLSPFTLAFDIAGSAPRGF
XP_004296516.1 1 MS--KDRLVKSVTKFADVQYKLFTKRYGQQLIDIFEFPIKLVLSPFTIAFDVAGSAQRGF
XP_014504339.1 1 MSSSRERLVKAVKKFVGVRYKVFSNRYGQNIIDILDFPIKVVLSPFTLAFDIAGSAPRGF
XP_004141219.1 1 MSSGTVRLVKKLKKFADFHYKVFTRRYGQQVIDILDLPLKVVLSPFTLVFDIAGSAPRGF
XP_011462391.1 1 MS--KDRLVKSVTKFADVQYKLFTKRYGQQLIDIFEFPIKLVLSPFTIAFDVAGSAQRGF
KRH27260.1 1 MTSSRERLVKAVKQFVEVRYKVFSNRYGQNIIDILDFPIKVVLSPFTLAFDIAGSAPRGF
XP_008452442.1 1 MS-ETARLLKKLKKFADFQYKVLTTRYGQKVIDVLEFPFNVVLSPFTLIFDIAGSAPRGF
XP_007149472.1 1 MGSSRERLVKAVKKFVDVRYKVFSNRYGPTIIDILDFPIKVVLSPFTLAFDIAGSAPRGF
NP_001242125.1 1 MTSSRERLVKAVKQFVEVRYKVFSNRYGQNIIDILDFPIKVVLSPFTLAFDIAGSAPRGF
XP_010925570.1 1 MPRGTERFLKAVKQQADVYYKVFSTRYGQQIIDLLELPIKVVLSPFTLAYDIAGSAPRGF
XP_009421280.1 1 MGRGTEKMLKAAKQFADTQYKKLTARYGQQMIDLLELPIRLVLSPFTIAIDVAGSAPRGF
KRH27261.1 1 MTSSRERLVKAVKQFVEVRYKVFSNRYGQNIIDILDFPIKVVLSPFTLAFDIAGSAPRGF
XP_003543237.1 1 MTSSRERLVKAVKQFVEVRYKVFGDRYGQNIIDILDFPIKVVLSPFTLAFDIAGSAPRGF
XP_009421281.1 1 MGRGTEKMLKAAKQFADTQYKKLTARYGQQMIDLLELPIRLVLSPFTIAIDVAGSAPRGF
KNA22491.1 1 MSQRKGKLVEQVKQFADAQYKNFTRDYGQQVIDIFEFPLRLVLSPFTIAYGIAGSAPRGF
XP_011098290.1 1 MTRGTDKLARSIKKFVDLHYKLFTRRYGQQLMDILEFPVKVVLSPFTLPFDIAGSAPRGF
XP_008794135.1 1 MARGTERFLKAVKQQADVYYKVFTARYGQQIVDLLELPIKVVLSPFTLAYDIAGSAPRGF
XP_012850705.1 1 MSRGTAKLVASIKKFVDLQYKFLSRRYGQQFMDILDFPVKVVLIPFTLPFDIAGSAPRGF
EPS69394.1 1 MSRGTEKLLKSIKNFVDFQYKVLSRRYGQQIIDILDFPVKLVLSPFTLPFDIAGSAPRGF
XP_010675513.1 1 MSKGTDKLIKQVQEFADNQYQYICRKYGQDLIDIFDFPIKLVLSPFTIAWTVVGSAPRGF
XP_004488729.1 1 MAK-PGGFQLFVKNYLGDRYNEFNSAYGETVSDIFGIPFKILFAPLTLAFDIAGSAPRGF
KVI02226.1 1 MGKGTNPFVK--KKLFEDQFQFLRQYIGKPLAFLLDIPVQLVQAPFLIANDISGSAPRGF
XP_006859011.1 1 MARGTERLAKLIKQNADFYGKIFKDHFAPPLMGLIELPYRVLISPFTLAADIVGSAPRGF
XP_013464684.1 1 MPKRPRGFGNFVKKFFRNRFKEFNNAYGETVSDVFAFPFKVLFTPLTVAIDIAGSAPRGF
AFK37051.1 1 MPKRPRGFGNFVKKFFRNRFKEFNNAYGETVSDVFAFPFKVLFTPLTVAIDIAGSAPCGF
consensus 1 * . . . . *.....*.... ...***..**


NP_568801.2 61 GIPEFISKISYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGSVH
XP_002864286.1 61 GIPEFISKISYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGSVH
CDY20217.1 61 GIPELISKISYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
XP_013727384.1 61 GIPELISKISYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
XP_009119926.1 61 GIPELISKISYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
XP_006280533.1 61 GIPEFVSKLSYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
XP_013614882.1 61 GIPELISKISYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
KFK27002.1 61 GIPEFISKLSYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGHVH
XP_006401652.1 61 GIPEFISKISYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTVCKGTGNVH
XP_010442956.1 61 GIPEFVSKLSYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
XP_010482789.1 61 GIPEFVSKLSYLSVFAVATLGTYDIALDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
XP_010446126.1 61 GIPEFVSKLSYLSVFAVATLGTYDIAFDLGKKVICQRDCKTCNGWQALRCTMCKGTGNVH
XP_010537026.1 61 GIPEFISKLSYLSVFAVATLGTYDIALELGKKVICQRDCKTCNGWQALRCTMCKGTGRVH
XP_012076623.1 61 GIPELISKLSYLSVFAVATLGTYDIALELGKKVICQRNCRTCNGWQALRCTMCKGSGKVH
XP_007021511.1 61 GIPEFISKLSYTSVFAIAALGTYDIALELGKKVICQRNCRTCNGWMALRCTMCKGSGKVH
XP_012473151.1 61 GIPEFISKLSHTSIFAIAAFGTYDIALELGKKVICQRNCRTCNGWQALRCTMCKGSGKVH
XP_008226614.1 61 GVPELVSKLSYMSIFAVATLGTYDIALELGKKVICQRNCGTCNGWQALRCTMCRGSGRVH
XP_007211737.1 61 GVPELVSKLSYMSIFAVATLGTYDIALELGKKVICQRNCGTCNGWQALRCTMCRGSGRVH
KJB22098.1 61 GIPEFISKLSHTSIFAIAAFGTYDIALELGKKVICQRNCRTCNGWQALRCTMCKGSGKVH
XP_008226615.1 61 GVPELVSKLSYMSIFAVATLGTYDIALELGKKVICQRNCGTCNGWQALRCTMCRGSGRVH
XP_010060594.1 61 GVPEFISKLSYASVFAVATLGTYDIALELGKKVICQRNCKTCNGWQALRCTMCKGSGKVH
XP_011029493.1 61 GVPELVSKLSYASIFVVATLGTYDIALEMGKKVICQRNCQTCNGWQALRCHMCQGAGKVH
XP_011043817.1 61 GAPELISKLSYASIFAIATLGTYDIALEMGKKVICQRNCKTCNGWQALRCNMCKGAGKVH
KCW67371.1 61 GVPEFISKLSYASVFAVATLGTYDIALELGKKVICQRNCKTCNGWQALRCTMCKGSGKVH
XP_009340889.1 60 GVPELVSKLSYMSIFAVATLGTYDIALEMGKKVICQRNCGTCNGWQALRCTMCRGSGRVQ
XP_002285056.1 61 GIPEFISKLSFSAIFVVATLGTYDIALDLGKKVLCQRNCRTCNGWQALRCTMCRGSGKVQ
XP_009625377.1 61 GVPEFISKLSYSAIFVVATVGTYDIAMELGKKVLCQRDCRTCNGWQALRCTMCKGSGKVH
XP_002528255.1 61 GVPELISKLSYLSVFAVATLGTYDIALELGKKVICQRNCHTCNGWQALRCNMCRGSGQVH
XP_006358975.1 61 GVPEFISKLSYSTIFVVATLGTYDIAMELGKKVISQRDCRTCSGWQALQCTMCKGSGKVH
XP_010260241.1 61 GVPEFISHLSFSAIFAVATLGTYDIALELGKKVLCQRNCRTCNGWQALRCTMCKGIGKVH
XP_009794897.1 61 GVPEFISKLSYSAIFVVATVGTYDIAMELGKKVLCQRDCRTCNGWQALRCTMCKGSGKVH
XP_006464775.1 61 GVPELISKLSYVSVFAVAAIGTYDIAFELGKKVICQRHCRTCNGWQALRCTMCRGSGQVQ
XP_006451818.1 61 GVPELISKLSYVSVFAVAAIGTYDIAFELGKKVICQRHCRTCNGWQALRCTMCRGSGQVQ
XP_004251862.1 61 GVPEFISKLSYSTIFVVATLGTYDIAMELGKKVISQRDCRTCCGWQALQCTMCKGSGKVH
XP_015061560.1 61 GVPEFISKLSYSTIFVVATLGTYDIAMELGKKVISQRDCRTCCGWQALQCTMCKGSGKVH
XP_008386188.1 60 GVPELVSKLSYMSIFAVATLGTYDIALEMGKKVICQRNCGTCNGWQALRCTMCRGSGRVQ
XP_006464773.1 61 GVPELISKLSYVSVFAVAAIGTYDIAFELGKKVICQRHCRTCNGWQALRCTMCRGSGQVQ
XP_004296516.1 59 GVPELISKLSCASIFVVATLGTYDIAMEMGKKVICQRNCETCNGWQALQCAMCKGSGRVH
XP_014504339.1 61 GVPELISKLSATSVFAIAALGTYDIAFELGRKVLCQRNCKSCNGWRALRCTMCRGSGRVH
XP_004141219.1 61 GVPELISKLSYASIFAIATFGTYDIALDLGRKVICQRQCQTCNGWQAMRCTKCRGSGMVN
XP_011462391.1 59 GVPELISKLSCASIFVVATLGTYDIAMEMGKKVICQRNCETCNGWQALQCAMCKGSGRVH
KRH27260.1 61 GVPELISKLSATSVFAIAALGTYDIAFELGRKVICQRNCNSCNGWRALRCTMCRGSGRVH
XP_008452442.1 60 GVPELISKLSYASIFAIASFGTYDIALDLGRKVVCQRNCQTCNGWQAMRCTKCRGSGMVN
XP_007149472.1 61 GVPELISKLSATSVFAIAALGTYDIAFELGRKVLCQRNCNTCSGWRALRCTMCRGSGRVH
NP_001242125.1 61 GVPELISKLSATSVFAIAALGTYDIAFELGRKVICQRNCNSCNGWRALRCTMCRGSGRVH
XP_010925570.1 61 GIPELASRLSFSAIFVVATLGTYDIALELGRKVICQRNCRTCNGWQALRCTMCKGTGKVH
XP_009421280.1 61 GLPELASKISFSAIFLVATLGTYDIALELGKKVVCQRNCSTCNGWQALRCTMCKGTGKVH
KRH27261.1 61 GVPELISKLSATSVFAIAALGTYDIAFELGRKVICQRNCNSCNGWRALRCTMCRGSGRVH
XP_003543237.1 61 GVPELISKLSATSVFAIAALGTYDIAFELGRKVICQRNCNSCNGWRALHCTMCRGSGRVH
XP_009421281.1 61 GLPELASKISFSAIFLVATLGTYDIALELGKKVVCQRNGSTCNGWQALRCTMCKGTGKVH
KNA22491.1 61 GIPEFTSKVSYTTIFAVATLGTWDIALELGKKVICQRDCQTCSGWQALRCTMCRGSGQVH
XP_011098290.1 61 GVPEFISKLSYSAVFAIATLGTYDIAFELGKKVLCQRNCRTCTGWQAMQCTMCRGSGKVQ
XP_008794135.1 61 GIPELASRLSFSAIFVVATLGTYDIALELGRKVICQRNCRTCNGWQALRCTMCKGTGKVH
XP_012850705.1 61 GVPEFVSKLSYSAIFAIATLGTYDIAFEMGKKVLCQRNCRTCSGWQAMQCTMCRGSGKVQ
EPS69394.1 61 GIPELISKISYSAIFVVATLGTYDVAFEMGKKVLCQRNCRTCCGWQAMQCTMCRGSGKAH
XP_010675513.1 61 GVPKFISKVSYTAIFAIATFGTWDIALELGKKVICQRDCRTCSGWQALRCTMCRGTGQVH
XP_004488729.1 60 GVPELISKFSAASVFAVATLGTYDIALNLGRKVICQRNCGTCNGWQALRCTMCRGSGRVH
KVI02226.1 59 GIPEFISKLSYSAIFAVAALGTCDIALELGKKVICQRNCQTCNGWNALRCTMCRGTGKVM
XP_006859011.1 61 GVPEFISHLSFSAIFAVATLGTYDIALELGKKVLCQRNCGTCSGWQALRCTMCKGTGKVQ
XP_013464684.1 61 GVPELVSKLSAASVFAVATLGTYDIALNLGRKVICQRDCVTCNGWQAMRCTMCRGSGRVH
AFK37051.1 61 GVPELVSKLSAASVFAVATLGTYDIALNLGRKVICQRDCVTCNGWQAMRCTMCRGSGRVH
consensus 61 *.*. .*..* .* .* .**.*.* ..*.**..** * .* ** *. *..*.*.* .


NP_568801.2 121 YQIKDYNLRSGEKPTADCVADAIVENRAELVHLPS-SFNHSAPLPSKDCPTCDGTGAMSC
XP_002864286.1 121 YQIKDYNLRSGEKPTADCVADAIVENRAELVHLPS-SINHSAPLPSKDCPTCDGTGVMSC
CDY20217.1 121 YQIKDYNLRSGEKPTADCVADAIVDNRAELVHLPS-SLNLSAPLPTKDCPTCDGTGVMSC
XP_013727384.1 121 YQIKDYNLRSGEKPTADCVADAIVDNRAELVHLPS-SLNLTAPLPTKDCPTCDGTGVMSC
XP_009119926.1 121 YQIKDYSLRSGEKPTADCVADAIVDNRAELVHLPS-SLNLSAPLPTKDCPTCDGTGVMSC
XP_006280533.1 121 YQIKDYNLRSGEKPTADSVADAIVDNRAELVHLPS-SINLSAPLPSKDCPTCDGTGVMSC
XP_013614882.1 121 YQIKDHNLRSGEKPTADCVADAIVDNRAELVHLPS-SLNLSAPLPSKDCPTCDGTGVMSC
KFK27002.1 121 YQIKDYNLRSGEKPTADCVADAIVDNRAELVHLPS-SLNLSAPLPSKDCPTCDGTGVMSC
XP_006401652.1 121 YQIKDYNLRSGEKPTADCVADAIVENRAELVHLPS-SINLSAPLPSKDCPTCDGTGVMSC
XP_010442956.1 121 YQIKDYNLRSGEKPTADSVADAIVENRAELVHLPS-SIDISAPLPSKDCPTCDGTGVMSC
XP_010482789.1 121 YQIKDYNLRSGEKPTADSVADAIVENRAELVHLPS-SIDISAPLPSKDCPTCDGTGVMSC
XP_010446126.1 121 YQIKDYNLRSGEKPTADSVADAIVENRAELVHLPS-SIHISAPLPSKDCPTCDGTGVMSC
XP_010537026.1 121 YQIKDYNLRSGEKPTADSVADAIVENRAELVHLPS-SINLSAPLPSKDCPTCDGTGVMSC
XP_012076623.1 121 YQVKNYTLRSGEKATAECIADAIANNRAELVHLPS-TMDLNIPLPSNECPTCDGTGVMGC
XP_007021511.1 121 YQVKNYTLRSGEKATAECVADAIADNRAELVHLPS-SIDLNTPLPSKDCPTCDGTGVMHC
XP_012473151.1 121 YQVKTHSLRSGEKATAECVADAIADNRAELVHLPS-AIDLNAPLPSKECPTCDGTGVMHC
XP_008226614.1 121 YQVKNYTLKSGEKATAECIADAIADNRAELVHLPS-SINLNAPLPSKDCPTCDGTGVMGC
XP_007211737.1 121 YQVKNYTLKSGEKATAECIADAIADNRAELVHLPS-SINLNVPLPSKDCPTCDGTGVMGC
KJB22098.1 121 YQVKTHSLRSGEKATAECVADAIADNRAELVHLPS-AIDLNAPLPSKECPTCDGTVIMHC
XP_008226615.1 121 YQVKNYTLKSGEKATAECIADAIADNRAELVHLPS-SINLNAPLPSKDCPTCDGTGVMGC
XP_010060594.1 121 YQVNNYSLKSGEKATAECIADAVVNNRAELVHLPS-SADLHSPLPSKDCPSCDGTGVMGC
XP_011029493.1 121 YQVKNYTLRSGEKATAECIADAIADNRAELVHLPS-SMDLNMPLPSKDCPTCDGTGVMSC
XP_011043817.1 121 YQVKNYTLRSGEKATAECIADAIADNRAELVHLPP-SMDLNMPLPSKDCPSCDGTGVMSC
KCW67371.1 121 YQVNNYSLKSGEKATAECIADAVVNNRAELVHLPS-SADLHSPLPSKDCPSCDGTVVMGC
XP_009340889.1 120 YQVKNYTLRSGEKATAQCVADAIADNRAELVHLPS-ATNLDAPLPSKECPTCDGTGVMGC
XP_002285056.1 121 YQVKNFSLKSGEKATAESVADAISDNRAELVHLPS-TVDLHLPLPSKDCPTCDGSGVMGC
XP_009625377.1 121 YQVKNYTLKRGEKATAESIADAIADNRAELVHLPS-TVDLHLPLPSKECPNCDGSGVMKC
XP_002528255.1 121 YQVRTCNLKSGEKATAECVAEAISDNRAELVHLPS-TMDLNMPLPSKDCPTCDGTGVMGC
XP_006358975.1 121 YQVKNYTLKRGEKATAESIADAIADNRAELVHLPS-TVDLHLPLPSKECPSCDGSGVMKC
XP_010260241.1 121 YQVKNYTLKSGERETAECVADAIADNRAELVHLPS-TVDLKLPLPSKDCPTCSGSGVMSC
XP_009794897.1 121 YQVKNYTLKRGEKATAESIADAIADNRAELVHLPS-TVDLHLPLPSKECPNCDGSGVMKC
XP_006464775.1 121 YQVKGYTLRSGEKATAEHVADAIADNRAELVHLPS-TIDLNTPLPSKDCPTCDGTGVMGC
XP_006451818.1 121 YQVKGYTLRSGEKATAEHVADAIADNRAELVHLPS-TIDLNTPLPSKDCPTCDGTGVMGC
XP_004251862.1 121 YQVKNYTLKRGEKATAESIADAIADNRAELVHLPS-TVDLHLPLPSKECPSCDGSGVMKC
XP_015061560.1 121 YQVKNYTLKRGEKATAESIADAIADNRAELVHLPS-TVDLHLPLPSKECPSCDGSGVMKC
XP_008386188.1 120 YQVKNYTLRSGEKATAECVADAIADNRAXLVHLPS-ATNLDVPLPSKECPTCDGTGVMGC
XP_006464773.1 121 YQVKGYTLRSGEKATAEHVADAIADNRAELVHLPS-TIDLNTPLPSKDCPTCDGTGVMGC
XP_004296516.1 119 YQVKNYPLKSGERETAECVAAAIADNRAELVHLPS-AIDLHTPLPSKDCLTCDGTGVMSC
XP_014504339.1 121 YQVKNYTLKSGEKPTAESVADAIANNRAELVHLPS-SLNFQNPLPSKECPTCDGTGVMSC
XP_004141219.1 121 YQVKNYVLRSGEKPTPESIADAIVENRAELIHLPS-SLDLHTPLPSKDCPTCDGTGVMCC
XP_011462391.1 119 YQVKNYPLKSGERETAECVAAAIADNRAELVHLPS-AIDLHTPLPSKDCLTCDGTGVMSC
KRH27260.1 121 YQVKNCNLKSGEKATAESVADAIANNRAELVHIPS-SLDFQKPLPSKECPTCDGTGVMSC
XP_008452442.1 120 YQVKNYALRSGEKPTPESIADAIAENRAELIHLPS-TLDLHTPLPSKDCPTCDGTGVMRC
XP_007149472.1 121 YQVKNYTLKSGEKPTAESVADAIANNRAELVHLPS-SLDFQNPLPSKECPTCDGTGAMSC
NP_001242125.1 121 HQAKNCNLKSGEKATAESVADAIANNRAELVHIPS-SLDFQKPLPSKECPTCDGTGVMSC
XP_010925570.1 121 YQVKNYSPKSGERATAECVAEAIVENRAELVHLPA-TINLSIPLPSKDCPSCDGSGVMGC
XP_009421280.1 121 YQVKNYNPRSGERATADCVAEAIFENRAELVHLPA-AINLSVPLPSKDCPTCDGTGVMGC
KRH27261.1 121 YQVKNCNLKSGEKATAESVADAIANNRAELVHIPS-SLDFQKPLPSKECPTCDGTGVMSC
XP_003543237.1 121 YQVKNCNLNSGEKATAESVADAIANNRAELVHIPS-SLDFQKPLPSKECPTCDGTGVMSC
XP_009421281.1 121 YQVKNYNPRSGERATADCVAEAIFENRAELVHLPA-AINLSVPLPSKDCPTCDGTGVMGC
KNA22491.1 121 YQVKSYTLKSGEKPTAESVAGAIANDQAELFHLPS-AIDMHTPLPLKECPTCDGTGVMKC
XP_011098290.1 121 YQVKNYALRSGEKATAENIADAIAENRAELVHLPS-SVDLHLPLPYKECPSCDGSGVMKC
XP_008794135.1 121 YQVKNYSPKSGERATAECVAEAIVENCAELVHLPA-TIDLSIPLPSKDCPSCDGSGVMGC
XP_012850705.1 121 YQVKNYALRSGEKATAANIADAISENRAELVHLPS-SVDLHLPLPCKECPSCEGSGVMKC
EPS69394.1 121 YQVKNYALRSGEKPTAENIADAIVENRAELVHIPL-SLDLNSTLPSKDCPNCDGSGVMKC
XP_010675513.1 121 YQVKSYTLRSGEKPTAESVADAIANNQAELVHLPS-AIDLHTPLPSKDCPTCDGTGVMKC
XP_004488729.1 120 YQVKNCTLKSGEKATAESVADAIVNNRAELVHLPS-SLDLQIPLPSKDCPTCDGTGVMGC
KVI02226.1 119 YQVKNYRLRSGEKPTAEATADAITENRAELVHLPA-TMDLNVPLPTKDCETCDGSGVMKC
XP_006859011.1 121 YQVKNYTLKSGQRPTPESVAEAIADNRAELVHLPS-AIDLDLPLPTKECPSCDGSGVMSC
XP_013464684.1 121 YQVKNCHLKSGEKATAESVADAIVNNQAELVHIPSSSFDPQIPLPSKDCPICDGTGVMGC
AFK37051.1 121 YQVKNCHLKSGEKAAAESVADAIVNNQAELVHIPSSSFDPQIPLPSKDCPICDGTGVMGC
consensus 121 .*.. ...*.. ... .*.*. ..*.*.*.* . .**...*..*.*...* *


NP_568801.2 180 TECKNKL-QVRISADDIMEPPWKAYNVLKKMDYPYEHIVHSMKDPSIANFWLITLPQIVG
XP_002864286.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLKKMDYPYEHIVHSMKDPSIANFWLITLPQIVG
CDY20217.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITLPQIVG
XP_013727384.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITLPQIVG
XP_009119926.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITLPQIVG
XP_006280533.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITMPQIMG
XP_013614882.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITLPQIVG
KFK27002.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITLPQIMG
XP_006401652.1 180 PECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITLPQIMG
XP_010442956.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITMPQIVG
XP_010482789.1 180 TECNNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITMPQIVG
XP_010446126.1 180 TECKNKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIANFWLITMPQIVG
XP_010537026.1 180 PECKNKL-QVKISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPNIATFWLITLPQVVG
XP_012076623.1 180 PECKDKL-QVRISADDIMEPPWKAYNIMRKMDYPYEHIVESMKDPSIAAFWLFSLPQIVG
XP_007021511.1 180 PECKHKL-QVRISADDIMEPPWKAYNILRKMDYPYEHIVHSMKDPSIAAFWLLTLPQIVG
XP_012473151.1 180 PECKRKI-QVRISADDIMEPPWKAYNILRKMDYPYEHIVHSMKDPSIAAFWLLTLPQIVG
XP_008226614.1 180 HECKDKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIAAFWLFTLPQVAG
XP_007211737.1 180 HECKDKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIAAFWLFTLPQVAG
KJB22098.1 180 PECKRKI-QVRISADDIMEPPWKAYNILRKMDYPYEHIVHSMKDPSIAAFWLLTLPQIVG
XP_008226615.1 180 HECKDKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIAAFWLFTLPQVAG
XP_010060594.1 180 PECKHKL-QVRISADDIMEPPWLAYNVMRKMDYPYEHIVHSMKDPSIAAFWLITFPQIVG
XP_011029493.1 180 PECKNKL-QVRISADDIMEPPWQAYNVLSKMEYPYEHIVDSMKDPSIAAFWLFSFPEIMG
XP_011043817.1 180 PECKNKL-QVRISADDIMEPPWQAYNVLRKMEYPYEHIVDSMRDPSIAAFWLFSFPEIMG
KCW67371.1 180 PECKHKL-QVRISADDIMEPPWLAYNVMRKMDYPYEHIVHSMKDPSIAAFWLITFPQIVG
XP_009340889.1 179 PECKDKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSIAAFWLFTLPQVVG
XP_002285056.1 180 PECKNKL-QIRISADDIMEPPWTAYNILRKMHYPYEHIVHSMKDPSIAAFWLLTMPQIVG
XP_009625377.1 180 PECKDKL-QIRISADDIMEPPWKAYNIMRKMDYPYEHIVDSMKDPSIAAFWLITLPQIVG
XP_002528255.1 180 PECKHKL-QVRISADDIMEPPWKAYNILKKMDYPYEHIVDSMKDPSIAAFWLFSLPQIVG
XP_006358975.1 180 PECKDKL-QIRISADDIMEPPWKAYNIMRKMDYPYEHIVDSMKDPSIAAFWLLTLPQIVG
XP_010260241.1 180 PECKHKL-QVRISADDIMEPPWKAYNVLRKMDYPYEHIVHSMKDPSVAAFWLITMPQILG
XP_009794897.1 180 PDCKDKL-QIRISADDIMEPPWKAYNIMRKMDYPYEHIVDSMKDPSIAAFWLITLPQIVG
XP_006464775.1 180 PECKNKL-QVTISADDIMEPPWKAYNVLRKMDYPYEHIVESMRDPSIAAFWLMTFPQIVG
XP_006451818.1 180 PECKNKL-QVTISADDIMEPPWKAYNVLRKMDYPYEHIVESMRDPSIAAFWLMTFPQIVG
XP_004251862.1 180 PECKDKL-QIKISADDIMEPPWKAYNIMRKMDYPYEHIVDSMKDPSIAAFWLLTLPQIVG
XP_015061560.1 180 PECKDKL-QIKISADDIMEPPWKAYNIMRKMDYPYEHIVDSMKDPSIAAFWLLTLPQIVG
XP_008386188.1 179 PECKDKL-QVRISADDIMEPPWKAYNVLRNMDYPYEXIVHGMKDPSIAAFWLFTLPQVAG
XP_006464773.1 180 PECKNKL-QVTISADDIMEPPWKAYNVLRKMDYPYEHIVESMRDPSIAAFWLMTFPQIVG
XP_004296516.1 178 PECKNKL-QVRISADDIMEPPWKAYNVMKKMDYPYEHIVDSMKDPSIAAFWLFTLPRVVG
XP_014504339.1 180 PECKHRLIHVRISADDIMEPPWQAYNVLKKMDYPYEHIIHSMKDPSIAAFWLITFPQIMG
XP_004141219.1 180 PECKNKL-PLRISADDIMEPPWKAYNVMKKMEYPYEHLAHSMKDPSIAAFWLLTFPQIVG
XP_011462391.1 178 PECKNKL-QVRISADDIMEPPWKAYNVMKKMDYPYEHIVDSMKDPSIAAFWLFTLPRVVG
KRH27260.1 180 PECKHRLLHVRISADDIMEPPWQAYNVLKKIDYPYEHILHSMKDPSIAAFWLITFPQIMG
XP_008452442.1 179 PECKNKL-PLRISADDIMEPPWKAYNVLRKMEYPYEHLVQSMKDPSIATFWLITFPQIVG
XP_007149472.1 180 PECKHRLVHVRISADDIMEPPWQAYNVLKKMDYPYEHIVHSMKDPSIAAFWLITFPQIMG
NP_001242125.1 180 PECKHRLLHVRISADDIMEPPWQAYNVLKKIDYPYEHILHSMKDPSIAAFWLITFPQIMG
XP_010925570.1 180 PDCKHRL-QVRISADDIMEPPWKAYNVMRKMDYPYENIVHSMRDPKIAAFWLITMPQVLG
XP_009421280.1 180 PECKHKL-QVRISADEIMEPPWKAYNVMRKMDYPYENIVHSMRDPKIAAFWLITMPQIMG
KRH27261.1 180 PECKHRLLHVRISADDIMEPPWQAYNVLKKIDYPYEHILHSMKDPSIAAFWLITFPQIMG
XP_003543237.1 180 PECKHRLLHVRISADDIMEPPWQACNVLKKIDYPYEHILHSMKDPSIAAFWLITFPQIMG
XP_009421281.1 180 PECKHKL-QVRISADEIMEPPWKAYNVMRKMDYPYENIVHSMRDPKIAAFWLITMPQIMG
KNA22491.1 180 TECKNKL-QIRISADDIMEPPWKAYNVLKKMDYPYEHIINSMKDPSIAAFWLLTMPQIVG
XP_011098290.1 180 PECKGKF-PLRISADDIMEPPWKACDILRKMDYPYEHIVHSMKDPSIAAFWLITLPQIVG
XP_008794135.1 180 PECKHRL-QVRISADDIMEPPWKAYNVMRKMDYPYENIVHSMRDPKIAAFWLITMPQVLG
XP_012850705.1 180 PECKGKL-PIRISADDIMEPPWKAYNILRKMDYPYEHVIHSMKDPSIAAFWLITLPQIVG
EPS69394.1 180 PECNGKF-PLRISADDIMEPPWKAYNIMRKMDYPYEHIIHSMKDPSIAAFWLITFPQIVG
XP_010675513.1 180 PECKGKL-HVRISADDIMEPPWKAYNVLKKMDYPYEHIVHSMKDPSIAEFWLLTMPQIVG
XP_004488729.1 179 TECKNKL-QLRISADGIMEPPWQAYSVLEKMDYPYEHIRHSMKDPSIAAFWLITFPQIMG
KVI02226.1 178 YECKDKL-QVRISADDITEPPWKAYNIMKKMDYPYEHIIDSMKDPSIAAFWLITMPQIMG
XP_006859011.1 180 PECKNKL-QVRISADDIMEPPWKAYNVFRKMEYPYEHIADSMRDPNIAAFWLISMPEIMG
XP_013464684.1 181 PECKNKL-PFKISADGIMEPHWQARNVLEKMDYPYEHIRDSMKDPDIAEFWLIAYPQIMY
AFK37051.1 181 PECKNKL-PFKISADGIMEPHWRARNVLEKMDYPYEHIRDSMKDPDIAEFWLIAYPQIMY
consensus 181 .*. .. ..****.*.**.* *........****... .*.** .* *** . *....


NP_568801.2 239 GFDYDEDVKKKIWWQYEESMRYDQLRDLVAKRNPGWEYLQDALVSIDPVRAREDPVIVKN
XP_002864286.1 239 GFDYDEDVKKKIWWQYEESMRYDQLRDLVAKRNPGWEYLQDALVSIDPVRAREDPVIVKN
CDY20217.1 239 GFDYDEDVKKKIWLQYEESMRYDQLRDLVAKRNPGWEYLQDALFSIDPVRAQEDPVIVKN
XP_013727384.1 239 GFDYDEDVKKKIWLQYEESMRYDQLRDLVAKRNPGWEYLQDALFSIDPVRAQEDPVIVKN
XP_009119926.1 239 GFDYDEDVKKKIWLQYEESMRYDQLRDLVAKRNPGWEYLQDALFSIDPVRAQEDPVIVKN
XP_006280533.1 239 GFEYDEDVKKKIWWQYEESMRYDQLRDLVAKRNPGWEYLQNALISIDPVRARDDPVIVKN
XP_013614882.1 239 GFDYDEDVKKKIWLQYEESMRYDQLRDLVAKRNPGWEYLQDALFSIDPVRAQEDTVIVKN
KFK27002.1 239 GFEYDEDVKKKIWWQYEESMKYDQLRDLVAKRNPGWEYLQDALFSIDPVRARDDPVIVKN
XP_006401652.1 239 GFDYDEDVKRKIWLQYDESMRYDQLRDLVAKRNPGWEYLQDALISIDPDRAREDPVIVRN
XP_010442956.1 239 GFEYDEDVKKKIWWQYEESTRYDQLRDLVAKRNPGWEYLQDALVSIDPFRAREDPIIVKN
XP_010482789.1 239 GFEYDEDVKKKIWWQYEESTRYDQLRDLVAKRNPGWEYLQDALVSIDPVRAREDPVIVKN
XP_010446126.1 239 GFEYDEDVKKKIWWQYEESTRYDQLRDLVAKRNPGWEYLQDALVSIDPVRAREDPVIVKN
XP_010537026.1 239 GFEYDEDVKRKIWWQYEESMRYDQLRDLVAKRKPGWEYLQEALISIDPVRAREDPLIVKN
XP_012076623.1 239 GFNYDDDVKQKIWWQYKESMRYDQLRDVVAKRKPGWEHLQEALISIDPIRAREDPVIVKN
XP_007021511.1 239 GFDYDDDVKQKIWWQYKESMRYDQLRDVVAKRNPGWEHLQEALIAIDPVRARDDPVVVKN
XP_012473151.1 239 GLEYDDDVKQKIWWQYKESMRYDQLRDVVAKRNPGWEHLQEGLITLDPVRARDDPVVVKN
XP_008226614.1 239 GFDFDDDVKKKIWWQYKESMRYDQLRDAVAKRKPGWENLQEALISIDPARAREDPVIVKN
XP_007211737.1 239 GFDFDDDVKKKIWWQYKESMRYDQLRDVVAKRKPGWENLQEALISIDPARAREDPVIVKN
KJB22098.1 239 GLEYDDDVKQKIWWQYKESMRYDQLRDVVAKRNPGWEHLQEGLITLDPVRARDDPVVVKN
XP_008226615.1 239 GFDFDDDVKKKIWWQYKESMRYDQLRDAVAKRKPGWENLQEALISIDPARAREDPVIVKN
XP_010060594.1 239 GFEYDDEVKQKIWWQYKESMRYDQIRDVVANRKPGWEHMQEALISIDPVRARDDPVVVKN
XP_011029493.1 239 GFNYDGDIKQKIWWQYKESMRYDQLRDMVAKRKPGWEHLQEALISIDPVRAREDPIVVKN
XP_011043817.1 239 GFNYDGDVKQRIWWQYKESMRYDQLRDMVAKRNPGWEHLQEALISIDPVRARDDPVVVKN
KCW67371.1 239 GFEYDDEVKQKIWWQYKESMRYDQIRDVVANRKPGWEHMQEALISIDPVRARDDPVVVKN
XP_009340889.1 238 GFDFDDDIKKKIWWQYKESMRYDQLRDVVAKRKPGWEQLQEALISIDPARAKEDPVIVKN
XP_002285056.1 239 GFNYDDEVKQKIWWQYKESMRYDQLRDVVAKRIPGWEHLQEALISIDPVRAREDPVVVKN
XP_009625377.1 239 GFEYDDDVKQKIWWQYKESMRYDQLRDVVAKRKPGWEYLQDALISIDPSRAREDPVIVKN
XP_002528255.1 239 GFDYDDDVKQKIWWQYKESMRYDQLRDVVAKRKPGWEYLQEALISIDPVRAREDPVIVKN
XP_006358975.1 239 GFEYDEDVKKKIWWQYKESMRYDQLRDVVAKRKPGWEYLQEALVSIDPARAREDPVIVKN
XP_010260241.1 239 GFNYDEDVKQKIWWQYKESMRYDQIRDVVAKRKPGWEHLQEALISIDPVRARDDPVVVKN
XP_009794897.1 239 GFEYDDDVKQKIWWQYKESMRYDQLRDVVAKKKPGWEYLQDALISIDPSRAREDPVIVKN
XP_006464775.1 239 GFDFDDEVKQKIWWQYKESMRYDQLRELVAKRSPGWEHLQEALIPIDPVRAREDPVIVKN
XP_006451818.1 239 GFDFDDEVKQKIWWQYKESMRYDQLRELVAKRSPGWEHLQEALISIDPVRAREDPVIVKN
XP_004251862.1 239 GFEYDDDVKKKIWWQYKESMRYDQLRDVVAKRKPGWEYLQEALVSIDPARAREDPIIVKN
XP_015061560.1 239 GFEYDDDVKKKIWWQYKESMRYDQLRDVVAKRKPGWEYLQEALVSIDPARAREDPIIVKN
XP_008386188.1 238 GFDFDDDIKKKIWWQYKESMRYDQLRDVVAKRKPGWEZLQEALISIDPARAKEDPVIVKN
XP_006464773.1 239 GFDFDDEVKQKIWWQYKESMRYDQLRELVAKRSPGWEHLQEALIPIDPVRAREDPVIVKN
XP_004296516.1 237 GYEYDDDVKQKLWWQYKESMRYDQLRDVVAKRKPGWEHLQEALISIDPARARDDPVIVKN
XP_014504339.1 240 GFTYDDDVKQKIWWQYKESRRYDQLRDVVAKRKPGWEYLQEALISIDPDRARDDPVIVKN
XP_004141219.1 239 GFNFDEDVKQKIWWQYKESMRYDQLRDAVAERKPGWEYLQKALISLDPVRARDDPVVVKN
XP_011462391.1 237 GYEYDDDVKQKLWWQYKESMRYDQLRDVVAKRKPGWEHLQEALISIDPARARDDPVIVKN
KRH27260.1 240 GFTYDDDVKQKIWWQYKENMRYDQLRDVVAKRKPGWEYLQDALISIDPARARDDPVIVKN
XP_008452442.1 238 GFNFDDDIKQKIWWQYKESMRYDQLRDAVAKRKPGWEYLQEALISIDPVRARDDPVVVKN
XP_007149472.1 240 GFTYDDDVKQKIWWQYKESRRYDQLRDVVAKRKPGWEYLQEALISIDPDRARDDPVIVKN
NP_001242125.1 240 GFTYDDDVKQKIWWQYKENMRYDQLRDVVAKRKPGWEYLQDALISIDPARARDDPVIVKN
XP_010925570.1 239 GFNYDDDVRQKIWWAYKESMRYDQLRDAVAKRKPGWEHLQDALISLDPIRARDDPVIVKN
XP_009421280.1 239 GFNFDEDVKQKIWWSYEESMRYDQLRDLVAERKPGWEHLQNALISVDPVRARDDPVIVKN
KRH27261.1 240 GFTYDDDVKQKIWWQYKENMRYDQLRDVVAKRKPGWEYLQDALISIDPARARDDPVIVKN
XP_003543237.1 240 GFTYDDDVKQKIWWQYKENMRYDQLRDVVAKRKPGWEYLQDALISIDPARARDDPVIVKN
XP_009421281.1 239 GFNFDEDVKQKIWWSYEESMRYDQLRDLVAERKPGWEHLQNALISVDPVRARDDPVIVKN
KNA22491.1 239 GFNFDDETKQKIWWQYKESMRYDQLRDVVAQRQPGWEHLQQALVTLDPARARDDPVIVKN
XP_011098290.1 239 GMDYDDDVKQKIWWQYKESMRYDQLRNEVAKREPGWEYLQEALISIDPSRAKDDPVVVKN
XP_008794135.1 239 GFNYDDDVKQKIWWAYKESMRYDQLRDAVAKRKPGWEHLQDALISLDPIRARDDPVIVKN
XP_012850705.1 239 GVDFDDEIKQKIWWEYKESRRYDQLRDEVAKRKPGWEYLQEALISIDPNRAKEDPVIVKN
EPS69394.1 239 GFEYDDDVKQKIWWQYKESIRYDQLRDEISKRKPGWEYLQQALISIDPARAHDDPVVVKN
XP_010675513.1 239 GFNFDDEIKQKIWWQYKESMRYDQLRDVVAQRQPGWEDLQQALITLDPARARDDPVVVKN
XP_004488729.1 238 GFTYGDNAKQAIWSQYKESRRYDQLRDEVAKRKPGWEYLQNALISIDPDRARDDPVIVKN
KVI02226.1 237 GFEYDDEVKQKIWWQYKENMRYDQLRFAVAKRKPGWEYLQEGLISMDPERARSDPVIVKN
XP_006859011.1 239 GFTFDDDAKKKIWWAYKESMRYDQLRDIVAKRKPGWEHLQQALVSIDPERARDDPVIVKN
XP_013464684.1 240 GFTYGDNAKKTILSEYEES-RWDQLRDEVATRKPGWEFLNDALISIDPDRARDDPVIGKN
AFK37051.1 240 GFTYGDNAKKTILSEYEES-RWDQLRDEVATRKPGWEFLNDALISIDPDRARDDPVIGKN
consensus 241 *. .... * *. ..**.*. .. . **** .. .*...** **..*.....*


NP_568801.2 299 VPYYKAKKSLEAEVTKLNPPPRPQ---NWG-----ELNLPLNISSWSEEDLKNPAKLYEK
XP_002864286.1 299 VPYYKAKKSLEAEVTKLSPPPRPQ---NWG-----ELNLPLNTSSWSEEDLKNPAKLYEK
CDY20217.1 299 VPFYKAKKALEAEVTKLNPPPRPQ---NWG-----ELNLPLNTSSWSEEDLKNPAKLYEK
XP_013727384.1 299 VPFYKAKKALEAEVTKLNPPPRPQ---NWG-----ELNLPLNTSSWSEEDLKNPAKLYEK
XP_009119926.1 299 VPFYKAKKALEAEVTKLNPPPRPQ---NWG-----ELNLPLNTSSWSEEDLKNPAKLYEK
XP_006280533.1 299 VPYYKAKKSLEAEVTKLSPPPRPQ---NWG-----ELNLPLNTSSWSEEELKNPAKLYEK
XP_013614882.1 299 VPFYKAKKALEAEVTKLNPPPRPQ---NWG-----ELNLPLNTSSWSEEDLKNPAKLYEK
KFK27002.1 299 VPFYKAKKALEAEVTKLTPPPRPQ---NWG-----ELNLPLNSSSWSEEDLKNPAKLYEK
XP_006401652.1 299 VPFYKAKKALEAEVTKLNPPPRPQ---NWG-----DLNLPLNTSSWSEDDLKNPAKLYEK
XP_010442956.1 299 VPYYKAKKSLEAEVTKLSPPPRPQ---NWG-----ELNLPLNTSSWSEDDLKNPAKLYEK
XP_010482789.1 299 VPYYKAKKSLEAEVTKLSPPPRPQ---NWG-----ELNLPLNTSSWSEEDLKNPTKLYEK
XP_010446126.1 299 VPYYKAKKSLEAEVTKLSPPPRPQ---NWG-----ELNLPLNTSSWSEDDLKNPAKLYEK
XP_010537026.1 299 VPYYKAKKALEAEVMKLDVPSRPQ---NWG-----DLNLPLSSSSWSEEDLKDSAKLYEK
XP_012076623.1 299 IPYYKAKKALEAEVMKLNPPPRPQ---NWG-E----LDLPLNVSSWSEEDLKNPEKFYEM
XP_007021511.1 299 IPYYKAKKALEAEVMKLDPPPRPQ---NWG-----KLDLPLNASSWSEDDLKDPAKLYEM
XP_012473151.1 299 VPYYKAKKALEAEVMRLDPPPRPQ---NWG-----NLDLPLNASSWSEDDLKDPAKLYEM
XP_008226614.1 299 IPYYKAKKALEAEVMKLDPPPRPQ---NWG-E----LDLPLNASSWSKEDLKNPEKFYEM
XP_007211737.1 299 IPYYKAKKALEAEVMKLDPPPRPQ---NWG-E----LDLPLNASSWSKEDLKNPEKFYEM
KJB22098.1 299 VPYYKAKKALEAEVMRLDPPPRPQ---NWG-----NLDLPLNASSWSEDDLKDPAKLYEM
XP_008226615.1 299 IPYYKAKKALEAEVMKLDPPPRPQ---NWG-E----LDLPLNASSWSKEDLKNPEKFYEM
XP_010060594.1 299 VPYYNAKKALEAEVMKLDPPARPQ---NWG-----ELELPLNSSSWNEEDLKDPDKYYEM
XP_011029493.1 299 VPYYKAKKALEAEVMKLDPPPRPQ---NWG-E----LDLPLNASSWSEEDLKKPEKFYEM
XP_011043817.1 299 VPYYKAKKALEAEVMKLDPPPRPQ---NWG-G----LDLPLNASSWSEEDLKKPEKFYEM
KCW67371.1 299 VPYYNAKKALEAEVMKLDPPARPQ---NWG-----ELELPLNSSSWNEEDLKDPDKYYEM
XP_009340889.1 298 IPYYKAKKALETEVMKLDPPPRPQ---NWG-E----LDLPLNASSWSKEDLKNPEKFYEM
XP_002285056.1 299 VPYYKARKALEAEVMKLDPPPRPQ---NWG-----ELNLPLNASSWSEEDLKDPKKLYEM
XP_009625377.1 299 IPYFKAKKALEAEVMKLDPPPRPQ---KWG-----ELNLPLDTSSWSEEDLKDPKKLYEM
XP_002528255.1 299 IPYYKAQKALEAEVMKFDPPPRPQ---NWG-FADRILELPLNASSWSEEDLKNPESFYEM
XP_006358975.1 299 IPYFKAKKALEAEVMKLDPPPRPQ---NWG-----ELNLPLNASSWSEKDLKDPKKLYEM
XP_010260241.1 299 IPYYKARKALEAEVMKLDVPPRPD---NWG-----ELNLPLDASSWSDEELKDPRKLYEM
XP_009794897.1 299 IPYFKAKKALEAEVMKLEPPPRPQ---KWG-----ELNLPLDASSWSEEDLKDPKKLYEM
XP_006464775.1 299 IPYYKAKKALETEVMKLDPPPRPE---NWG-D----LDLPLNASSWSEDDLKDPKKLYEK
XP_006451818.1 299 IPYYKAKKALETEVMKLDPPLRPE---NWG-D----LDLPLNASSWSEDDLKDPKKLYEK
XP_004251862.1 299 VPYFKAKKTLEAEVMKLDPPPRPQ---NWG-----ELNLPLNASSWSEKDLKDPKKLYEM
XP_015061560.1 299 IPYFKAKKTLEAEVMKLDPPPRPQ---NWG-----ELNLPLNASSWSEKDLKDPKKLYEM
XP_008386188.1 298 IPYYKAKKALETEVMKLDPPPRPQ---NWG-E----LDLPLNASSWSKDDLKNPEKFYEM
XP_006464773.1 299 IPYYKAKKALETEVMKLDPPPRPE---NWG-D----LDLPLNASSWSEDDLKDPKKLYEK
XP_004296516.1 297 VPFYKAKKALEAEVMKLNPPPRPQ---NWG-E----LDLPLNASSWSKEDLKDPEKFYEM
XP_014504339.1 300 VPYYKAQKALETEVMKLDPPPRP---SNWG-----ELDLPLNASSWSEEDLKNPEKFYEM
XP_004141219.1 299 IPYFKAKKALEAEVIKLDPPPRPQ---NWG-----ELDLPLNSSSWSQDDLKDPGKFNEM
XP_011462391.1 297 VPFYKAKKALEAEVMKLNPPPRPQ---NWGQE----LDLPLNASSWSKEDLKDPEKFYEM
KRH27260.1 300 VPYYKAKKALETEVMKIDPPPRP---PNWG-----ELDLPLNASSWSEEDLQDPDKFYEM
XP_008452442.1 298 IPYFKAKKTLETEVMKLDPPPRPQ---NWG-----DLDLPLNSSSWSQDDLKDPGKFYEM
XP_007149472.1 300 VPYYKAQKALETEVMKLDPPARP---SNWG-----ELDLPLNASSWSEEDLKNPEKFYEM
NP_001242125.1 300 VPYYKAKKALETEVMKIDPPPRP---PNWG-----ELDLPLNASSWSEEDLQDPDKFYEM
XP_010925570.1 299 IPFYKAKKELEAEVMKLDVPPRPQ---HWG-----ELDLPLNASSWAEEDLKDPKKLLEM
XP_009421280.1 299 IPYYKAKKALEAEVLNLDVPPRPQ---NWG-----EIDLPLKASSWTEEELKDPAKVLEM
KRH27261.1 300 VPYYKAKKALETEVMKIDPPPRPPNWENWQ-----ELDLPLNASSWSEEDLQDPDKFYEM
XP_003543237.1 300 VPYYKAKKALETEVMKIDPPPRP---PNWG-----ELDLPLNASSWSEEDLKDPDKFYEM
XP_009421281.1 299 IPYYKAKKALEAEVLNLDVPPRPQ---NWG-----EIDLPLKASSWTEEELKDPAKVLEM
KNA22491.1 299 VPYYKAKKALEMEVLKLDPPPRPK---NWG-----ELNLPLDTSSWSKDDLKDPKKMDEM
XP_011098290.1 299 VPFYKAKKALEAEVMKLNPPPRPH---DWG-----DLDLPLNASSWSEEDLKDPKKLYEM
XP_008794135.1 299 IPFYKAKKELEGEVMKLDVPPRPH---NWG-----ELDLPLNASSWTEEDMKDPKKLLEM
XP_012850705.1 299 VPFFKAKKALESEVLKLNPPPRPQ---NWG-----DLDLPLNASSWTEEDLKDPKKLYEM
EPS69394.1 299 VSLFKAKKALEAEVMKLDPPPRPR---NWG-----ELDLPLNVSSWSEEDLKDPKKLYEM
XP_010675513.1 299 VPYYKAKKALEMEVMKFDPPPRPK---DWG-----ELNLPLNASSWSKDELKDPKKLQEM
XP_004488729.1 298 VPYYKAKKALEADVMKLDPPPRPS---NWG-----ELDLPLSSSSWGEEDLKDPEKFHEM
KVI02226.1 297 IPYYKAKNALETEVMKLDPPPRPQ---NWG-----ELNVPLSASSWSEEDLKDPKKLYEK
XP_006859011.1 299 IPYYKAKKALEAEVTKLDVPPRPD---NWG-----DLKLPLTKTSWTDEELKDPQKLLEM
XP_013464684.1 299 IPYYKAKKALEEEVMKLDPPPRPS---NWG-----ELDLPLSPSSWDEEDLKNPEKLHEM
AFK37051.1 299 IPYYKAKKALEEEVMKLDPPPRPS---NWG-----ELDLPLSPSSWDEEDLKNPEKLHEM
consensus 301 .....*.. ** .* .. *.** .*. . .** .**. .... . . *


NP_568801.2 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEQKVRPYIQDSSMAVLPQPILLKSQ-K
XP_002864286.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQNSSMAVLPQPILLKSQ-K
CDY20217.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSNMAVLPQPILLKSQ-K
XP_013727384.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSNMAVLPQPILLKSQ-K
XP_009119926.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSNMAVLPQPILLKSQ-K
XP_006280533.1 351 TILLNAQREISDKILDAQWEAKWRQEK-VEEMLEEKVRPFTQDSSMAVLPQPILLKSQ-K
XP_013614882.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSNTAVLPQPILLKSQ-K
KFK27002.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSSMAVLPQHILLKSH-K
XP_006401652.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSSMAVLPQPILLKSQ-K
XP_010442956.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSSTAALPQPILLKSQ-K
XP_010482789.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFIQDSSTAVLQQPILLKSQ-K
XP_010446126.1 351 TVLLNAQREIADKILDAQWEAKWRQEK-VEEMLEEKVRPFVQDSNTAVLQQPILLKSQ-K
XP_010537026.1 351 TVLLNAQREIADKILDAQWESKWRQEK-LEEMLEEKVRPYMQDPSTAVLPQPILLKSH-K
XP_012076623.1 351 TVLLNAQREIADKILDAQWETKWRQEK-LNEILEEKVRPYIQNINNAVLSRPIIMQSQ-N
XP_007021511.1 351 TVLLNAQREIADKLLDTQWETKWRQEK-LNEMLEEKVRPYIQNIDNSLLPQPIVIQSQ-N
XP_012473151.1 351 TVLLNAQREIADKLLDAQWETKWRQEK-LNEMLEEKVRPYIQNIDNGVLPQPIVIPSQ-T
XP_008226614.1 351 TVLLNAQREIADKILDSQWENRWRQQR-LNEMLEEKVRPYMQNIDSGILSQPIILPSQ-D
XP_007211737.1 351 TVLLNAQREIADKILDSQWENRWRQQR-LNEMLEEKVRPYMQNIDSGILSQPIILPSQ-D
KJB22098.1 351 TVLLNAQREIADKLLDAQWETKWRQEK-LNEMLEEKVRPYIQNIDNGVLPQPIVIPSQ-T
XP_008226615.1 351 TVLLNAQREIADKILDSQWENRWRQQR-LNEMLEEKVRPYMQNIDSGILSQPIILPSQ-D
XP_010060594.1 351 TVLLNAQREIADKILDTQWQAKWRQEK-LNEMLEEKVRPYIQDIDSSVLSRPIVLQSK-A
XP_011029493.1 351 TVLLNAQREIADKILDAQWETKWRQEK-LNELLEEKVRPYIQNINSGALPRPIIIQPQ-N
XP_011043817.1 351 TVLLNAQREIADKILDSQWETKWRQEK-LNDMLEEKVRPYIQNISNGALPRPILIQSQ-K
KCW67371.1 351 TVLLNAQREIADKILDTQWQAKWRQEK-LNEMLEEKVRPYIQDIDSSVLSRPIVLQSK-A
XP_009340889.1 350 TVLLNAQRELADKILDAQWETRWRQER-LNEMLEEKVRPYMQSIDSGTLSQPIILQLQ-D
XP_002285056.1 351 TVLLNAQREIADQVLDAQWQTKWRQEK-LNETLKEKVQPYIQSIDNGVLSQPIIIQSP-N
XP_009625377.1 351 TVLLNAQREIAEKMLDAQWEAKWREEK-LNKMMEEKVRPYIQNIDNTVLPQPIIMESERN
XP_002528255.1 355 TVLLNAQREIADKILDAQWESKWRQEK-LNELLEEKVRPYIQNSNTGVLSRPIILQSQ-D
XP_006358975.1 351 TVLLNAQREISEKMLDAQWEAKWREEK-LNKMLEEKVRPYIQNIDNAVLPQPIVMESERN
XP_010260241.1 351 TVLLNAQREMADKILDSQWEEKWRQEK-LNELLEEKVRPYIQRLDTGVLPRPIIIRSQ-H
XP_009794897.1 351 TVLLNAQREIAEKMLDAQWEAKWREEK-LNKMLEEKVRPYIQNIDNTVLPQPIVMESERN
XP_006464775.1 351 TVLLNAQREIADKMLDAQWETKWRQEK-LNEMLEEKVRPFMQNIDNAVLSRPILMHPQ-N
XP_006451818.1 351 TVLLNAQREIADKMLDAQWETKWRQEK-LNEMLEEKVRPFVQNIDNAVLSRPILMHPQ-N
XP_004251862.1 351 TVLLNAQREISEKMLDAQWEAKWREEK-LNKMLEEKVRPYIQNIDNTVLPQPIVMESERN
XP_015061560.1 351 TVLLNAQREISEKMLDAQWEAKWREEK-LNKMLEEKVRPYIQNIDNTVLPQPIVMESERN
XP_008386188.1 350 TVLLNAQRELADKILDAQWETRWRQER-LNEMLEEKVRPYMQSIDSGTLSQPIILXSQ-D
XP_006464773.1 351 TVLLNAQREIADKMLDAQWETKWRQEKQLNEMLEEKVRPFMQNIDNAVLSRPILMHPQ-N
XP_004296516.1 349 TVLLNAQREIADKMLDAQWETKWRQEK-LNEMLAEKVRPYMQKIDNGTLSQPIILPQQ-D
XP_014504339.1 352 TVLLNAQREISDKILDAQWETKWRQDK-LNEMLEAKVKPYIQDIDNAVLPEPILLKPQ-V
XP_004141219.1 351 TVLLNAQREIADKILDAQWETKWRQEK-LNELLEEKLRPYVTSANSHVLTAPIVLKSK-D
XP_011462391.1 350 TVLLNAQREIADKMLDAQWETKWRQEK-LNEMLAEKVRPYMQKIDNGTLSQPIILPQQ-D
KRH27260.1 352 TVLLNAQREISDKILDAQWETKWRQDK-LNEMLEAKVKPYIQDIDNAVLPEPILLKPQ-N
XP_008452442.1 350 TVLLNAQREIADKILDAQWETKWRQEK-LNELLEEKLRPYVTSANSHVLTAPIVLKSQ-D
XP_007149472.1 352 TVLLNAQREISDKILDAQWETKWRQDK-LNEMLEAKVKPFIQDIDNAVLPEPILIKPQ-N
NP_001242125.1 352 TVLLNAQREISDKILDAQWETKWRQDK-LNEMLEAKVKPYIQDIDNAVLPEPILLKPQ-N
XP_010925570.1 351 TTLLNAQREIADKILDAQWEAKWRQDK-LNDLLKEKVEPYVRSIDKGVLPQSIVIKSH-D
XP_009421280.1 351 TSLLTAQREIADNILDAQWEAKWRQDK-LNDLLKEKLQPYLRSVDKAGLPQSIVTRSH-D
KRH27261.1 355 TVLLNAQREISDKILDAQWETKWRQDK-LNEMLEAKVKPYIQDIDNAVLPEPILLKPQ-N
XP_003543237.1 352 TVLLNAQREISDKILDAQWETKWRQDK-LNEMLEAKVKPYIQDIDNAVLPEPILLKPQ-N
XP_009421281.1 351 TSLLTAQREIADNILDAQWEAKWRQDK-LNDLLKEKLQPYLRSVDKAGLPQSIVTRSH-D
KNA22491.1 351 KLLLNAQKEIAEKILDAQWEAKWRHNK-LNEMLEEKVQPYIQNVDT--LSQPIIMHPP-A
XP_011098290.1 351 TILLNAQREIADKILDTQWETKWREEK-LNEMLKEKVEPYLQNIDNGVLPQPIVLQSQ-N
XP_008794135.1 351 TTLLNAQREIADKILDAQWEAKWRQDK-LNDLLKEKVEPYVRSIDKGVLPQSIIIRSH-D
XP_012850705.1 351 TILLNAQREIADQILDAQWETRWREEK-MNEMLKEKVEPYLKNVDNSVLPQPIVLPSQRN
EPS69394.1 351 TILLNAQRDIADRMLDAQWEMQWREDK-VNEMLKEKLKPYLQSDGGGILSEPIFVQQQ-V
XP_010675513.1 351 TVLLNAQREIAEKVLDAQWEAKWRQNK-LNEMMEEKVQPYIQNVET--LSQPIIMHPP-A
XP_004488729.1 350 TVLLNAQREISEKILDAQWETKWRQEK-MNEMLEEKMQPYIQNIDKEVLSEPILLKPK-K
KVI02226.1 349 TVLLNAQREIAEKILDDQWETTWRQEK-LNKMLEEKLQPFIQEIDNGVLPHPILAGSQ-K
XP_006859011.1 351 TTLLTAQREMADKFLDTQWETKWRQDK-LNELLEEKVRPYIKSVDTGVLPQPIIMKPQ-D
XP_013464684.1 351 NVLLNAQREISENILDVQWETKWRQEK-VNEMLEEKLQPYIQNIDNEVLSEPILLKSQ-K
AFK37051.1 351 NVLLNAQREISENILDVQWETKWRQEK-VNEMLEEKLQPYIQNIDNEVLSEPILLKSQ-K
consensus 361 ..**.**... ...** **. .** .. . ... *. *.. * .*..

```
NP_568801.2     409 ----KA-QKG-S-RQRKWWFF
XP_002864286.1  409 ----KA-QKG-S-RQRKWWFF
CDY20217.1      409 ----TN-QKG-N-RQKKWWFF
XP_013727384.1  409 ----TN-QKG-N-RQKKWWFF
XP_009119926.1  409 ----TN-QKG-N-RQKKWWFF
XP_006280533.1  409 ----KD-QKG-N-RQKKWWIF
XP_013614882.1  409 ----TN-QKG-N-RQKKWWFF
KFK27002.1      409 ----KD-QKG-N-RQRKWWFF
XP_006401652.1  409 ----KN-QKG-N-RQRKWWFF
XP_010442956.1  409 ----KD-QKG-N-RQRKWWIF
XP_010482789.1  409 ----KD-QKG-N-RQRKWWIF
XP_010446126.1  409 ----KD-QKG-N-RQRKWWIF
XP_010537026.1  409 ----KD-QKG-N-KQRRWWFF
XP_012076623.1  409 ----KD-KKR-S-RQRRWWFF
XP_007021511.1  409 ----RN-QKR-A-RRRRWLFF
XP_012473151.1  409 ----GN-QKR-G-RRKRWLFF
XP_008226614.1  409 ----QN-QKK-T-RQRRWWFF
XP_007211737.1  409 ----QN-QKK-T-RQRRWWFF
KJB22098.1      409 ----GN-QKR-G-RRKRWLFF
XP_008226615.1  409 ----QN-QK--T-RQRRWWFF
XP_010060594.1  409 ----QD-QKR-KPRQRRWWFF
XP_011029493.1  409 ----QD-KKR---RQRRWWFF
XP_011043817.1  409 ----QD-KKR---RQRRWWFF
KCW67371.1      409 ----QD-QKR-KPRQRRWWFF
XP_009340889.1  408 ----QN-QKR-T-RQRRWWFF
XP_002285056.1  409 ----QD-QKGKT-RKRRWWLF
XP_009625377.1  410 ----QD-QKKKH-GRRRWWLF
XP_002528255.1  413 ----KD-KKV-----------
XP_006358975.1  410 ----QD-QKRKH-GRRRWWLF
XP_010260241.1  409 ----QE-NKK-R-TKRRWWLF
XP_009794897.1  410 ----QD-QKKKH-GRRRWWLF
XP_006464775.1  409 ENKNKN-QKK-A-RQKRWWFF
XP_006451818.1  409 ENKNKN-QKK-A-RQKRWWFF
XP_004251862.1  410 ----QD-QKRKH-GRRRWWLF
XP_015061560.1  410 ----QD-QKRKH-GRRRWWLF
XP_008386188.1  408 ----QN-QKR-T-RQRRWWFF
XP_006464773.1  410 ENKNKN-QKK-A-RQKRWWFF
XP_004296516.1  407 ----QN-QKR-N-RLRRWWFF
XP_014504339.1  410 -------QEKR--QRRRWFFF
XP_004141219.1  409 ----PE-QKR-N-RRRWFWFF
XP_011462391.1  408 ----QN-QKR-N-RLRRWWFF
KRH27260.1      410 -------QEKR--QRRRWFFF
XP_008452442.1  408 ----PV-QKR-N-RRRWFWFF
XP_007149472.1  410 -------QEKR--QRRRWFFF
NP_001242125.1  410 -------QEKR--QRRRWFFF
XP_010925570.1  409 ----QENKKKAT-TRRRWWFF
XP_009421280.1  409 ----QG-NKK-T-TRRRWFFF
KRH27261.1      413 -------QEKR--QRRRWFFF
XP_003543237.1  410 -------QEKR--QRRRWFFF
XP_009421281.1  409 ----QG-N-K-T-TRRRWFFF
KNA22491.1      407 E----D-KKRNS-RQRRWWFF
XP_011098290.1  409 ----HD-QKKNQ-GRRRWWLF
XP_008794135.1  409 ----QENKKK-T-TRRRWWFF
XP_012850705.1  410 ----QD-QKKNQ-GRRRWWLF
EPS69394.1      409 ----NH-HHSK--RKRRWWLF
XP_010675513.1  407 E----N-R-K-R-KSKRWWLF
XP_004488729.1  408 ----QD-K-----NRRRWWFF
KVI02226.1      407 ----ND-QKRSH-RSRRWWLF
XP_006859011.1  409 ----QE-AKESR-RRRRWLFF
XP_013464684.1  409 ----QE-K-K-H-RQRRWFIF
AFK37051.1      409 ----QE-K-K-H-RQRRWFIF
consensus       421                .... .
```

(b)

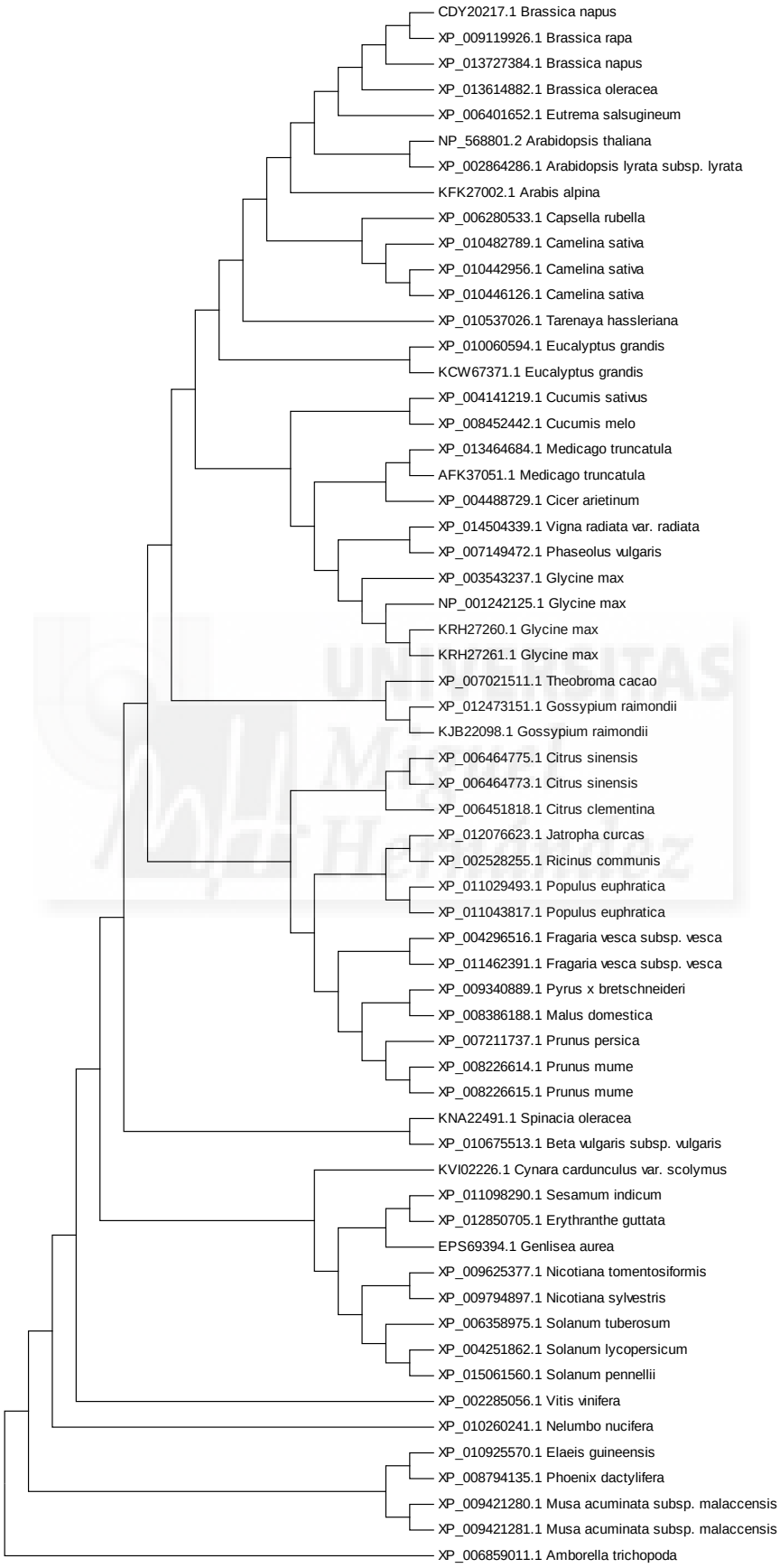

CDY20217.1 Brassica napus
XP_009119926.1 Brassica rapa
XP_013727384.1 Brassica napus
XP_013614882.1 Brassica oleracea
XP_006401652.1 Eutrema salsugineum
NP_568801.2 Arabidopsis thaliana
XP_002864286.1 Arabidopsis lyrata subsp. lyrata
KFK27002.1 Arabis alpina
XP_006280533.1 Capsella rubella
XP_010482789.1 Camelina sativa
XP_010442956.1 Camelina sativa
XP_010446126.1 Camelina sativa
XP_010537026.1 Tarenaya hassleriana
XP_010060594.1 Eucalyptus grandis
KCW67371.1 Eucalyptus grandis
XP_004141219.1 Cucumis sativus
XP_008452442.1 Cucumis melo
XP_013464684.1 Medicago truncatula
AFK37051.1 Medicago truncatula
XP_004488729.1 Cicer arietinum
XP_014504339.1 Vigna radiata var. radiata
XP_007149472.1 Phaseolus vulgaris
XP_003543237.1 Glycine max
NP_001242125.1 Glycine max
KRH27260.1 Glycine max
KRH27261.1 Glycine max
XP_007021511.1 Theobroma cacao
XP_012473151.1 Gossypium raimondii
KJB22098.1 Gossypium raimondii
XP_006464775.1 Citrus sinensis
XP_006464773.1 Citrus sinensis
XP_006451818.1 Citrus clementina
XP_012076623.1 Jatropha curcas
XP_002528255.1 Ricinus communis
XP_011029493.1 Populus euphratica
XP_011043817.1 Populus euphratica
XP_004296516.1 Fragaria vesca subsp. vesca
XP_011462391.1 Fragaria vesca subsp. vesca
XP_009340889.1 Pyrus x bretschneideri
XP_008386188.1 Malus domestica
XP_007211737.1 Prunus persica
XP_008226614.1 Prunus mume
XP_008226615.1 Prunus mume
KNA22491.1 Spinacia oleracea
XP_010675513.1 Beta vulgaris subsp. vulgaris
KVI02226.1 Cynara cardunculus var. scolymus
XP_011098290.1 Sesamum indicum
XP_012850705.1 Erythranthe guttata
EPS69394.1 Genlisea aurea
XP_009625377.1 Nicotiana tomentosiformis
XP_009794897.1 Nicotiana sylvestris
XP_006358975.1 Solanum tuberosum
XP_004251862.1 Solanum lycopersicum
XP_015061560.1 Solanum pennellii
XP_002285056.1 Vitis vinifera
XP_010260241.1 Nelumbo nucifera
XP_010925570.1 Elaeis guineensis
XP_008794135.1 Phoenix dactylifera
XP_009421280.1 Musa acuminata subsp. malaccensis
XP_009421281.1 Musa acuminata subsp. malaccensis
XP_006859011.1 Amborella trichopoda

Figure S5. Multiple sequence alignment and phylogenic tree of ANU7 and its orthologs. (a) Amino acid sequences were aligned using the consistency-based method implemented in T-Coffee, and colored with BOXSHADE version 3.21 (http://www.ch.embnet.org/software/BOX_form.html), by selecting 0.9 as the fraction of sequences that must agree for shading. (b) Phylogenetic analysis of ANU7 and its orthologs performed according to the Maximum Likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained automatically by applying the Maximum Parsimony method. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 1.0089)]. The rate variation model allowed for some sites to be evolutionarily invariable (+I, 17.4334% sites). The analysis involved 61 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 396 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. The tree was rooted using the sequence from the basal angiosperm *Amborella trichopoda* as the outgroup.

SUPPLEMENTAL REFERENCES

Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39,** 783-791.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8,** 275-282.

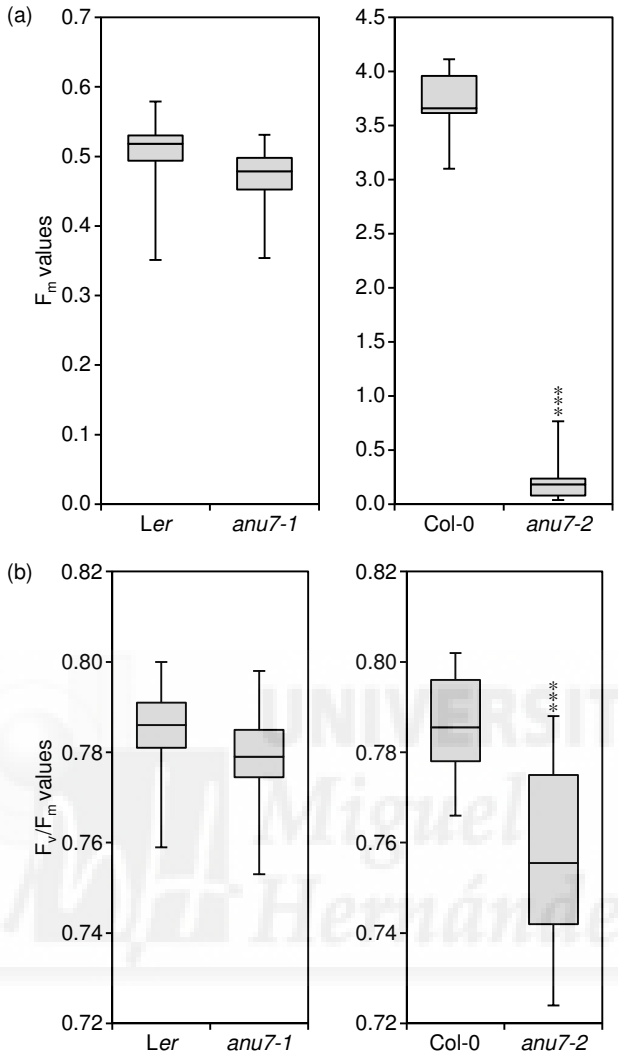


Figure S6. Photosynthetic parameters in L*er*, *anu7-1*, Col-0 and *anu7-2* plants. Box plot distribution of (a) F_m, and (b) F_v/F_m values in plants of the genotypes shown. Boxes are delimited by the first (Q1, lower hinge) and third (Q3, upper hinge) quartiles. Whiskers represent Q1-1.5·IQ (lower) and Q3+1.5·IQ (upper), where IQ=Q3-Q1. Outliers are not shown. Measurements were performed 16 das in L*er* and *anu7-1*, and 38 das in Col-0 and *anu7-2*. Asterisks indicate values significantly different from the corresponding wild type in a Student's *t*-test (n≥10; ****p*<0.001).

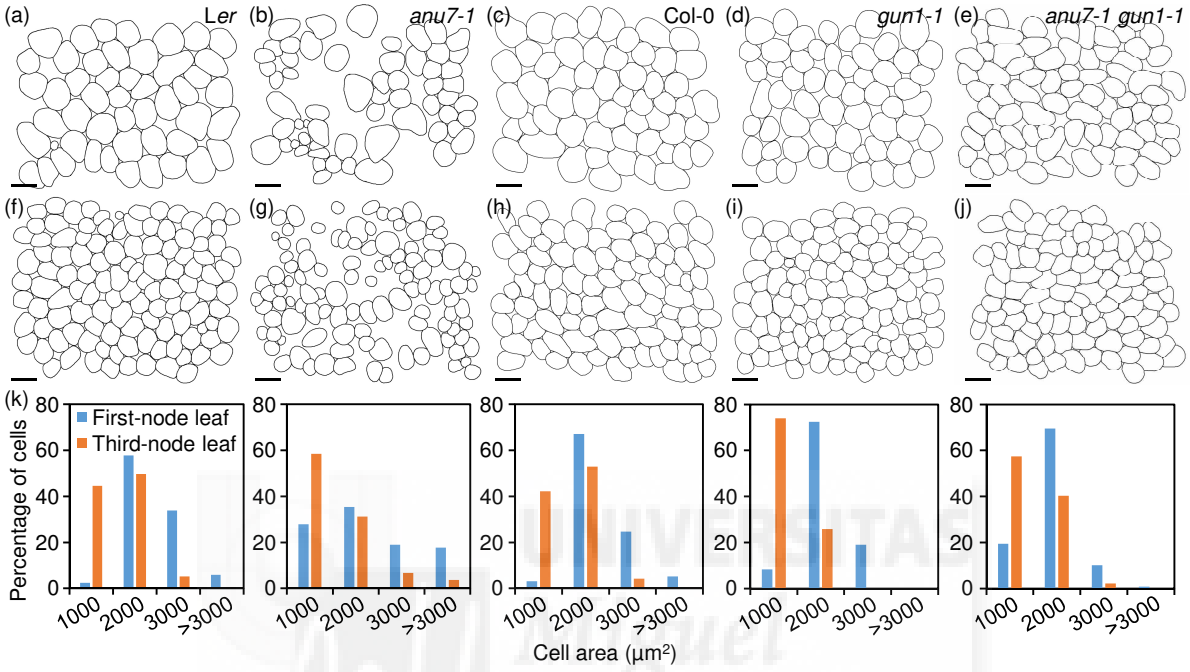


Figure S7. Morphometry of palisade mesophyll cells from L*er*, *anu7-1,* Col-0, *gun1-1* and *anu7-1 gun1-1* plants. (a-j) Diagrams of the subepidermal layer of palisade mesophyll cells from (a-e) first- and (f-j) third-node leaves. (k) Distribution of palisade mesophyll cell sizes (n≥600) in first- and third-node leaves. Plants were collected 21 das. Scale bars indicate 50 µm.

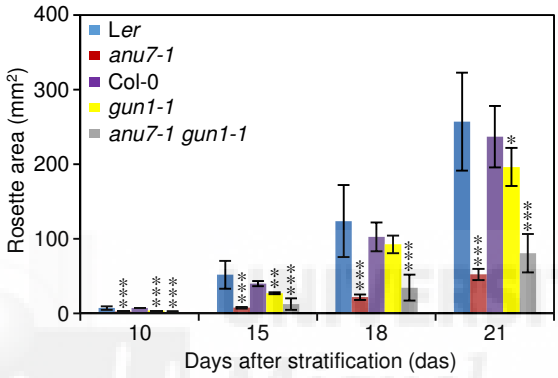


Figure S8. Rosette area of *anu7-1*, *gun1-1* and *anu7-1 gun1-1* plants. Error bars indicate standard deviations. Asterisks indicate values significantly different from the corresponding wild type in a Student's *t*-test ($^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, n≥16). The *anu7-1* and *gun1-1* single mutants were compared with L*er* and Col-0, respectively. The *anu7-1 gun1-1* double mutant was compared with Col-0.

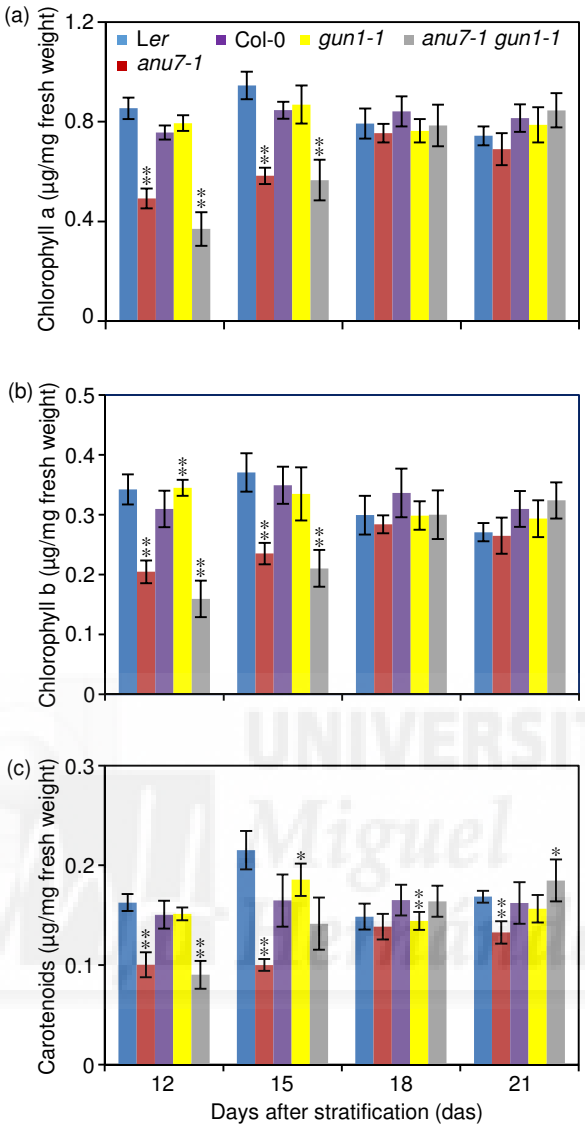


Figure S9. Pigment content in L*er*, *anu7-1*, Col-0, *gun1-1* and *anu7-1 gun1-1*. (a) Chlorophyll a, (b) chlorophyll b, and (c) carotenoids content. Plants were collected 12 ,15, 18 and 21 das. Error bars indicate standard deviations. Asterisks indicate values significantly different from the corresponding wild type in a Mann-Whitney U-test ($^{*}p<0.05$, $^{**}p<0.01$, n=6). The *anu7-1* and *gun1-1* single mutants were compared with L*er* and Col-0, respectively. The *anu7-1 gun1-1* double mutant was compared with Col-0.

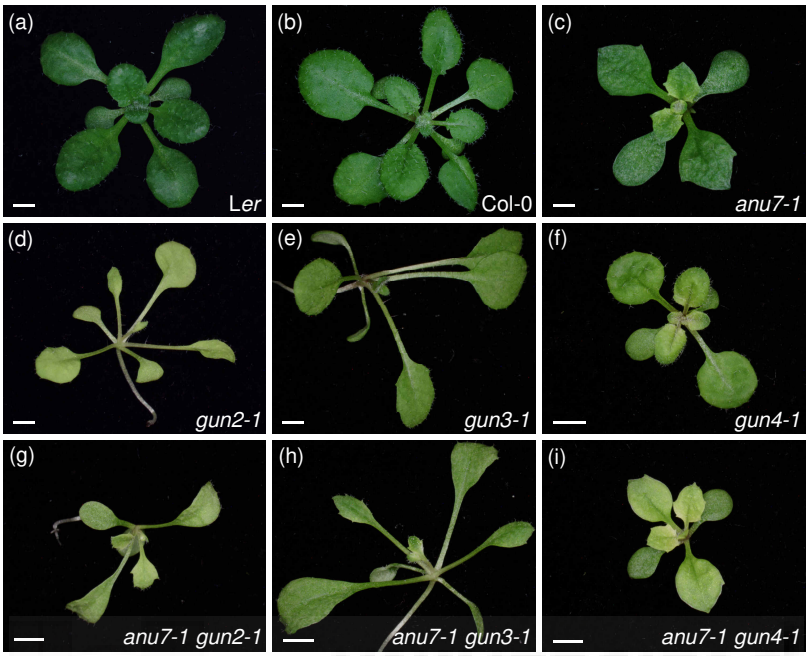


Figure S10. Vegetative phenotype of the *anu7-1 gun2-1*, *anu7-1 gun3-1* and *anu7-1 gun4-1* double mutants. Pictures were taken 16 das. Scale bars indicate 2 mm.

Table S1. Identity and similarity of ANU7 and its orthologs

Accession number	Organism	Protein length (aa)	Identity (%)	Similarity (%)
NP_568801.2	*Arabidopsis thaliana*	422	100.00	100.00
XP_002864286.1	*Arabidopsis lyrata*	422	98.34	99.29
CDY20217.1	*Brassica napus*	422	94.55	96.45
XP_013727384.1	*Brassica napus*	422	94.31	96.45
XP_009119926.1	*Brassica rapa*	422	94.31	96.21
XP_010482789.1	*Camelina sativa*	422	94.08	95.97
XP_013614882.1	*Brassica oleracea*	422	94.08	95.73
XP_010442956.1	*Camelina sativa*	422	93.84	96.45
XP_006280533.1	*Capsella rubella*	422	93.60	96.92
KFK27002.1	*Arabis alpina*	422	93.60	96.68
XP_006401652.1	*Eutrema salsugineum*	422	93.36	96.92
XP_010446126.1	*Camelina sativa*	422	93.13	95.50
XP_010537026.1	*Tarenaya hassleriana*	422	87.91	92.89
XP_012076623.1	*Jatropha curcas*	422	80.33	89.81
XP_007021511.1	*Theobroma cacao*	422	79.86	90.05
XP_012473151.1	*Gossypium raimondii*	422	78.20	89.81
KJB22098.1	*Gossypium raimondii*	422	77.96	89.81
XP_008226614.1	*Prunus mume*	422	77.73	89.34
XP_008226615.1	*Prunus mume*	421	77.30	88.89
XP_010060594.1	*Eucalyptus grandis*	423	77.30	87.47
XP_011043817.1	*Populus euphratica*	421	77.25	88.15
XP_007211737.1	*Prunus persica*	422	77.07	88.89
KCW67371.1	*Eucalyptus grandis*	423	77.07	87.47
XP_011029493.1	*Populus euphratica*	421	77.01	86.97
XP_002528255.1	*Ricinus communis*	417	76.29	85.68

Table S1 (continuation). Identity and similarity of ANU7 and its orthologs

Accession number	Organism	Protein length (aa)	Identity (%)	Similarity (%)
XP_009340889.1	*Pyrus x bretschneideri*	421	75.83	86.97
XP_006451818.1	*Citrus clementina*	426	75.82	86.15
XP_006464775.1	*Citrus sinensis*	426	75.82	85.92
XP_006464773.1	*Citrus sinensis*	427	75.64	85.71
XP_002285056.1	*Vitis vinifera*	423	75.41	87.94
XP_006358975.1	*Solanum tuberosum*	424	75.24	87.97
XP_014504339.1	*Vigna radiata*	421	75.18	85.11
XP_004251862.1	*Solanum lycopersicum*	424	74.76	87.97
XP_009625377.1	*Nicotiana tomentosiformis*	424	74.76	87.26
XP_010260241.1	*Nelumbo nucifera*	422	74.64	88.63
XP_009794897.1	*Nicotiana sylvestris*	424	74.53	87.74
XP_007149472.1	*Phaseolus vulgaris*	421	74.29	83.96
XP_011462391.1	*Fragaria vesca*	421	74.23	85.11
XP_004296516.1	*Fragaria vesca*	420	74.23	84.87
XP_008386188.1	*Malus domestica*	421	74.17	86.02
XP_015061560.1	*Solanum pennellii*	424	74.06	88.21
KRH27260.1	*Glycine max*	421	74.00	85.11
NP_001242125.1	*Glycine max*	421	73.82	84.91
XP_003543237.1	*Glycine max*	421	73.58	84.67
XP_004141219.1	*Cucumis sativus*	422	73.29	85.34
KRH27261.1	*Glycine max*	424	73.24	84.27
XP_008452442.1	*Cucumis melo*	421	72.27	86.49
KNA22491.1	*Spinacia oleracea*	425	72.10	85.58
XP_011098290.1	*Sesamum indicum*	423	71.63	86.05
XP_010925570.1	*Elaeis guineensis*	424	71.23	85.38

Table S1 (continuation). Identity and similarity of ANU7 and its orthologs

Accession number	Organism	Protein length (aa)	Identity (%)	Similarity (%)
XP_009421280.1	*Musa acuminata*	422	70.85	84.36
XP_009421281.1	*Musa acuminata*	421	70.62	84.12
XP_010675513.1	*Beta vulgaris*	419	70.21	84.40
XP_004488729.1	*Cicer arietinum*	418	70.14	80.09
XP_012850705.1	*Erythranthe guttata*	424	70.05	84.20
XP_008794135.1	*Phoenix dactylifera*	423	69.98	85.11
EPS69394.1	*Gensilea aurea*	422	69.03	83.22
KVI02226.1	*Cynara cardunculus*	421	68.32	80.38
XP_013464684.1	*Medicago truncatula*	421	66.67	78.72
AFK37051.1	*Medicago truncatula*	421	66.43	78.72
XP_006859011.1	*Amborella trichopoda*	423	66.19	81.80

Table S3. Gene Ontology terms significantly enriched in genes expressed at higher levels

Ontology	GO term	Description	Input	Reference	*p*-value	FDR
P	GO:0032774	RNA biosynthetic process	44	814	1.4e-06	0.00063
P	GO:0006350	Transcription	67	1468	8.7e-07	0.00063
P	GO:0006351	Transcription, DNA-dependent	44	812	1.3e-06	0.00063
P	GO:0080090	Regulation of primary metabolic process	66	1503	3.7e-06	0.00088
P	GO:0006355	Regulation of transcription, DNA-dependent	41	772	4.8e-06	0.00088
P	GO:0009719	Response to endogenous stimulus	40	748	5.4e-06	0.00088
P	GO:0051252	Regulation of RNA metabolic process	41	778	5.8e-06	0.00088
P	GO:0045449	Regulation of transcription	62	1385	4.2e-06	0.00088
P	GO:0019219	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	63	1418	4.4e-06	0.00088
P	GO:0009889	Regulation of biosynthetic process	63	1450	8.7e-06	0.00095
P	GO:0010556	Regulation of macromolecule biosynthetic process	62	1421	9.1e-06	0.00095
P	GO:0031326	Regulation of cellular biosynthetic process	63	1450	8.7e-06	0.00095
P	GO:0051171	Regulation of nitrogen compound metabolic process	63	1446	8e-06	0.00095
P	GO:0009791	Post-embryonic development	30	501	1.1e-05	0.0011
P	GO:0060255	Regulation of macromolecule metabolic process	65	1540	1.5e-05	0.0013
P	GO:0016070	RNA metabolic process	53	1174	1.8e-05	0.0015
P	GO:0042221	Response to chemical stimulus	69	1684	2e-05	0.0015
P	GO:0010468	Regulation of gene expression	63	1489	1.9e-05	0.0015
P	GO:0010033	Response to organic substance	46	974	2.3e-05	0.0016
P	GO:0009725	Response to hormone stimulus	36	687	2.4e-05	0.0017
P	GO:0019222	Regulation of metabolic process	67	1646	3.2e-05	0.002
P	GO:0031323	Regulation of cellular metabolic process	64	1547	3.1e-05	0.002
P	GO:0009414	Response to water deprivation	14	155	5.1e-05	0.003
P	GO:0009415	Response to water	14	164	9e-05	0.0051
P	GO:0010467	Gene expression	84	2304	0.00013	0.0072

Table S3 (continuation). Gene Ontology terms significantly enriched in genes expressed at higher levels

Ontology	GO term	Description	Input	Reference	*p*-value	FDR
P	GO:0006970	Response to osmotic stress	22	372	0.00019	0.01
P	GO:0009737	Response to abscisic acid stimulus	17	255	0.00028	0.014
P	GO:0050789	Regulation of biological process	98	2864	0.00032	0.016
P	GO:0050896	Response to stimulus	104	3107	0.00044	0.021
P	GO:0009651	Response to salt stress	20	344	0.00046	0.021
P	GO:0050794	Regulation of cellular process	91	2665	0.00058	0.026
P	GO:0006139	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	89	2598	0.00061	0.026
P	GO:0006950	Response to stress	65	1766	0.00064	0.027
P	GO:0009886	Post-embryonic morphogenesis	5	26	0.00074	0.03
P	GO:0006807	Nitrogen compound metabolic process	101	3055	0.00082	0.031
P	GO:0009628	Response to abiotic stimulus	44	1083	0.00081	0.031
P	GO:0009966	Regulation of signal transduction	9	103	0.0014	0.049
P	GO:0010646	Regulation of cell communication	9	103	0.0014	0.049
F	GO:0003700	Transcription factor activity	63	1448	8.3e-06	0.0036
F	GO:0030528	Transcription regulator activity	67	1628	2.3e-05	0.0051
F	GO:0003677	DNA binding	85	2283	6e-05	0.0088
F	GO:0016874	Ligase activity	23	405	0.00024	0.027
F	GO:0003676	Nucleic acid binding	110	3338	0.00051	0.045
C	GO:0009295	Nucleoid	8	33	4.4e-06	0.0015
C	GO:0016602	CCAAT-binding factor complex	5	12	3.5e-05	0.0061
C	GO:0009507	Chloroplast	98	2789	0.00013	0.015
C	GO:0009536	Plastid	98	2859	0.0003	0.026

P: Biological process; F: Molecular function; C: Cellular component.

Table S4. Gene Ontology terms significantly enriched in genes expressed at lower levels

Ontology	GO term	Description	Input	Reference	*p*-value	FDR
P	GO:0009725	Response to hormone stimulus	36	687	1.1e-12	7.3e-10
P	GO:0009719	Response to endogenous stimulus	37	748	2.6e-12	8.8e-10
P	GO:0010033	Response to organic substance	40	974	7.5e-11	1.7e-08
P	GO:0009733	Response to auxin stimulus	20	250	1.7e-10	2.9e-08
P	GO:0042221	Response to chemical stimulus	49	1684	3.5e-08	4.7e-06
P	GO:0050896	Response to stimulus	72	3107	1.2e-07	1.4e-05
P	GO:0009735	Response to cytokinin stimulus	8	68	4.3e-06	0.00042
P	GO:0009664	Plant-type cell wall organization	8	74	7.7e-06	0.00065
P	GO:0009416	Response to light stimulus	16	416	9.9e-05	0.0074
P	GO:0009314	Response to radiation	16	427	0.00013	0.0089
F	GO:0005199	Structural constituent of cell wall	7	44	2.9e-06	0.00086
F	GO:0016641	Oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	6	37	1.4e-05	0.002
F	GO:0016638	Oxidoreductase activity, acting on the CH-NH2 group of donors	6	48	5.2e-05	0.0051
F	GO:0016491	Oxidoreductase activity	33	1302	0.00011	0.0082
F	GO:0015035	Protein disulfide oxidoreductase activity	5	38	0.00018	0.011
F	GO:0015036	Disulfide oxidoreductase activity	5	48	0.00048	0.024
F	GO:0016667	Oxidoreductase activity, acting on sulfur group of donors	6	88	0.0011	0.043
F	GO:0005507	Copper ion binding	7	123	0.0012	0.043

P: Biological process; F: Molecular function.

Table S5. Primer sets used in this work

Purpose	Oligonucleotide name(s)	Oligonucleotide sequences (5'→3')	
		Forward primer (F)	Reverse primer (R)
Linkage analysis	MCL19_F/R	GGTAATGAGTAAAGTTTCAATATCA	GATGTTGATTTCTCATAAAGTCATA
	cer457664_F/R	CACTAGACCACATTCACTTTTCG	GAAAGAAACTAATGATGTAGTATTC
	cer454240_F/R	GACTATAGTCTCTAACGGTGAC	AGAGAAGCTAAGCTATTAAATCC
	cer455938_F/R	TTGTACTATGGTCAGCTTTCG	CTTAACTTATGGGATCGTTTATGC
	cer457042_F/R	GAATGGACGTACGTGAGTAGG	GTCTATTGGACATGGATAAAAGC
	cer452893_F/R	GACTACCTCGACTAACTTCTCC	CAATACTTGTCGAATCTAACGC
	MNB8_F/R	TCCTGCACATTTCATAGCGGAA	CTCTACACGTGTATATATCAACA
	cer456411_F/R	CTCAGCCCAATAGGCCCGTA	CAGGTTCTGTTGGGCCTATTG
	cer454427_F/R	CCTTGCGTGAGATCTACACTT	CACTAAAACGGAAAATTGAACAG
	cer456100_F/R	CAGCCTCGGCAGGCTTAGCT	GCCTCGGCCTCGGCCGGTTT
	JV65/66_F/R	CCATTACAAACAATAGAACTCTTT	GTGTTAATTAAACCTTTGTAGATC
	cer455494_F/R	CATGTAACATTCATGTACCA	AACATGGAAAGTGAAGGATC
	cer466785_F/R	GGCATGTACGAGCAAATGAATC	AGATTCCAATATTCAGTTAATGTG
	cer455324_F/R	GTGAAAGACAACTAAGACAGAAA	GCTATAGTGGATGAACTTAAATG
	JV61/62_F/R	GGATTTTTAAGATCACATGGATG	GAGATGAAGAGAGTAATATACTTG
	MUA2_F/R	CCGTTATTTCTCAAATGATCTTTC	GCTATGATGAAGAAGAAAGAAGG
Sequencing of candidate mutation	At5g53860_F1/R1	CAAAGAAACGACACGTAGTCTA	AACAAGAGATAATCCCACCTTC
	At5g53860_F2/R2	GGTTCTACTGTTTGAGTTCATC	TGCTGTGAACAATGTGCTGAAA
	At5g53860_F3/R3	CCTTTACTGACGATTCATTTTGC	GAAGAAAGCTTACACCCCAATT
	At5g53860_F4/R4	GAAGAATGTCCCTTATTACAAGG	CATAACATGAAAGACACACGTAA

Table S5 (continuation). Primer sets used in this work

Purpose	Oligonucleotide name(s)	Oligonucleotide sequences (5'→3')	
		Forward primer (F)	Reverse primer (R)
Gateway cloning	ANU7cds_F1/R1	GGGGACAAGTTTGTACAAAAAAGCAGGCT ATGTCGAGAGGACCAGGTCGT	GGGGACCACTTTGTACAAGAAAGCTGG GTTTAGAAGAACCACCACTTCCT
	ANU7cds_F2/R2	GGGGACAAGTTTGTACAAAAAAGCAGGCT CATCTTCTCCAATGTCGAGAG	GGGGACCACTTTGTACAAGAAAGCTGG GTAGAAGAACCACCACTTCCTTT
	ANU7cds_F3	GGGGACAAGTTTGTACAAAAAAGCAGGCT ATCTTCTCCAATGTCGAGAGG	
	ANU7pro_F/R	GGGGACAAGTTTGTACAAAAAAGCAGGCT GAGATAGAGAAAGAGGTATAG	GGGGACCACTTTGTACAAGAAAGCTGG GTTCGACATTGGAGAAGATGAAG
Genotyping of bacterial clones and transgenic plants	pGEMT221_F/R	GTTGTAAAACGACGGCCAGTG	GGAAACAGCTATGACCATGATT
	pMDC32_F/R	CCTTCGCAAGACCCTTCCTC	ACATGCTTAACGTAATTCAACAG
	pMDC83_F/R	GTAAGGGATGACGCACAATC	CGTATGTTGCATCACCTTCAC
	pMDC107_F/R	TCTTCGCTATTACGCCAGCTG	CGTATGTTGCATCACCTTCAC
	pMDC163_F/R	CTCTAGCATTCGCCATTCAGG	ATTGCCCGGCTTTCTTGTAAC
	ANU7_Int200_F/R	GAGACCTTACATCCAAGACTCA	GAGATGAACTCAGGGATGCCA
Genotyping of point mutations	anu7-1_F	CTTAAGCTATCCTAGTGATATTC	
	gun1-1_F/R	TCGTAATGAGTGCGATAAAGC	GCGTCAATCACAGCATTGTAGG
Genotyping of T-DNA lines	emb2737_1_LP/RP	ATGTAAGATCGCCTATTTGTGG	ACGAATCTGCCAAGTGTCTGC
	emb2737_2_LP/RP	CGCTTCGAATGTTTGCTAAAG	TGCACTCAGTACAGCTCATCG
qRT-PCR	q_FSD2_F/R	TATAGTGAAGACGCCCAATGC	ACAGTTTCCCATGACACAAGC
	q_FSD3_F/R	GGATGTGTGGGAGCACTCTTA	GGGTTCACCAAGATTCACAAA
	q_FLN1_F/R	TGTGACTGATGGAACACTTCG	GCTTTCTCATTATCCCTGCAA
	q_FLN2_F/R	ACGTTGATGTGCTCAAAGAGG	GCTGGAGTGCCATAGTGGTAA
	q_pTAC2_F/R	ACTTCTGGATGCTTTGTGGTG	AGCCAAACCGATAGTGCTGTA
	q_pTAC10_F/R	GAGGAACCAGGGAGGTACAAG	CATACCACACATTGCCTTTCC

Table S5 (continuation). Primer sets used in this work

Purpose	Oligonucleotide name(s)	Oligonucleotide sequences (5'→3')	
		Forward primer (F)	Reverse primer (R)
qRT-PCR	q_pTAC14_F/R	TGGAAAGATACGGCTTCTCAA	TCGATAACTCGCTGTCATGGT

UNIVERSITAS
Miguel
Hernández

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Deficient glutamate biosynthesis triggers a concerted upregulation of ribosomal protein genes in Arabidopsis

Received: 19 May 2017
Accepted: 29 June 2017
Published: xx xx xxxx

Tamara Muñoz-Nortes, José Manuel Pérez-Pérez, Raquel Sarmiento-Mañús, Héctor Candela & José Luis Micol

Biomass production requires the coordination between growth and metabolism. In a large-scale screen for mutants affected in leaf morphology, we isolated the *orbiculata1* (*orb1*) mutants, which exhibit a pale green phenotype and reduced growth. The combination of map-based cloning and next-generation sequencing allowed us to establish that *ORB1* encodes the GLUTAMATE SYNTHASE 1 (GLU1) enzyme, also known as FERREDOXIN-DEPENDENT GLUTAMINE OXOGLUTARATE AMINOTRANSFERASE 1 (Fd-GOGAT1). We performed an RNA-seq analysis to identify global gene expression changes in the *orb1–3* mutant. We found altered expression levels of genes encoding enzymes involved in nitrogen assimilation and amino acid biosynthesis, such as glutamine synthetases, asparagine synthetases and glutamate dehydrogenases, showing that the expression of these genes depends on the levels of glutamine and/or glutamate. In addition, we observed a concerted upregulation of genes encoding subunits of the cytosolic ribosome. A gene ontology (GO) analysis of the differentially expressed genes between L*er* and *orb1–3* showed that the most enriched GO terms were 'translation', 'cytosolic ribosome' and 'structural constituent of ribosome'. The upregulation of ribosome-related functions might reflect an attempt to keep protein synthesis at optimal levels even when the pool of glutamate is reduced.

The final shape and size of leaves depends on a complex sequence of developmental events, which include the recruitment of cells to the incipient leaf primordium, the control of cell proliferation, the transition from cell proliferation to cell expansion, and cell expansion and differentiation[1–3]. Leaf size also depends on the availability, absorption and assimilation of nutrients, mainly nitrogen, whose metabolism has to be tightly coordinated with carbon metabolism to promote biomass accumulation[4–7]. Because plant biomass is the outcome of interactions between metabolism and growth, understanding how metabolic pathways supply the building blocks for the growth of developing plant organs is a fundamental step towards the goal of engineering more productive crops with increased biomass[8]. Previous studies in *Arabidopsis thaliana* (hereafter, Arabidopsis) and other plant species have shown that the production of biomass can be enhanced by manipulating the expression of genes that encode positive and negative regulators of cell proliferation and expansion[9–16]. In Arabidopsis, a positive effect on plant growth and biomass accumulation occurs when certain genes are overexpressed. Two examples are *GROWTH-REGULATING FACTOR5* (*GRF5*) and *EXPANSIN10* (*EXP10*), which respectively encode a transcription factor and an expansin[9–11]. By contrast, other genes function as negative regulators of plant growth. Loss of function mutations of the *BIG BROTHER* (*BB*) and *DA2* genes, which encode E3 ubiquitin ligases, enhance cell proliferation[15,16]. As an alternative approach, modifying the expression levels of key enzymes in primary carbon and nitrogen metabolic pathways might also lead to increased biomass production[17].

In the leaves, light harvesting and carbon fixation occur in the chloroplasts, where the biosynthesis of essential metabolites required for the rapid growth of developing tissues takes place[18]. The dual role of leaves as both active metabolic sources and sinks makes them very sensitive to mutations that damage central biosynthetic pathways, often leading to plants with reduced growth. Indeed, many genes identified in an ethyl methanesulfonate (EMS) screen for leaf developmental mutants have been found to encode enzymes that catalyze steps of diverse metabolic pathways and other housekeeping functions[19–26]. The *VENOSA3* (*VEN3*) and *VEN6* genes of Arabidopsis encode the two subunits of the carbamoyl phosphate synthetase, which catalyzes the conversion of glutamine and bicarbonate into carbamoyl phosphate and glutamate in the arginine biosynthesis pathway[21]. *VEN1*, also

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202, Elche, Spain. Correspondence and requests for materials should be addressed to J.L.M. (email: jlmicol@umh.es)

known as *WEAK ETHYLENE INSENSITIVE2/ANTHRANILATE SYNTHASE α1* (*WEI2/ASA1*), encodes the α subunit of anthranilate synthase, which catalyzes the conversion of chorismate to anthranilate, the rate-limiting step in the tryptophan biosynthesis pathway[24]. *RUGOSA1* (*RUG1*) encodes the porphobilinogen deaminase, also known as hydroxymethylbilane synthase, which catalyzes the deamination and polymerization of four molecules of porphobilinogen into the linear tetrapyrrole 1-hydroxymethylbilane, the fifth step of tetrapyrrole biosynthesis[22]. The *EXIGUA1* (*EXI1*), *EXI2* and *EXI5* genes encode the CELLULOSE SYNTHASE 8 (CESA8), CESA7 and CESA4 catalytic subunits of the cellulose synthase complex, which is required for secondary cell wall synthesis[23]. In addition, many other genes identified in this screen, such as *APICULATA2* (*API2*), *ANGUSTA3* (*ANG3*), *DENTICULATA5* (*DEN5*), *DEN12*, *DEN29* and *DEN30*, encode different subunits of the cytosolic ribosome, highlighting the close association between active cell proliferation and protein synthesis during leaf development[19,20,25].

In this work, we focus on the *orbiculata1* (*orb1*) mutants, which were isolated in the screen mentioned above. *orb1* mutants have small, round, pale green leaves with no apparent patterning defects. We have found *ORB1* to be the same gene as At5g04140, also known as *GLUTAMATE SYNTHASE 1* (*GLU1*) and *FERREDOXIN-DEPENDENT GLUTAMINE OXOGLUTARATE AMINOTRANSFERASE 1* (*Fd-GOGAT1*). The Fd-GOGAT1 enzyme catalyzes the synthesis of glutamate from glutamine and α-ketoglutarate. Together with the conversion of glutamate and ammonium into glutamine using ATP, which is catalyzed by GLUTAMINE SYNTHETASE 2 (GS2), this reaction is an essential component of the GS/GOGAT cycle. The GS/GOGAT cycle plays a key role in the primary assimilation of exogenous ammonium[27], and in the re-assimilation of the ammonium released during photorespiration[28]. Glutamate is the major amino-group donor for the biosynthesis of many different amino acids and other nitrogen-containing compounds[29]. Tobacco mutants lacking Fd-GOGAT1, which do not produce glutamate, have been reported to have altered levels of other amino acids[28]. Previous studies have found a significant correlation between glutamate content and shoot biomass in barley, as well as between glutamate content and high productivity in rice, as expected if this amino acid acts as a metabolic hub linking numerous biosynthetic pathways with growth and development[30–32]. Our RNA-seq profiling of *orb1* mutant leaves has given insight on how nitrogen-related metabolic pathways are regulated during the vegetative phase, and shows some differences with the results of a microarray analysis previously performed using a different allele of the same gene[33]. In addition, our analysis of RNA-seq results has unveiled a concerted transcriptional increase of genes encoding components of the translational machinery in *orb1* mutants.

Results and Discussion

orb1 mutants have small, pale green leaves. In order to identify genes involved in leaf development, a large-scale screen for EMS-induced mutants of Arabidopsis with abnormal leaf growth or pigmentation was performed in the laboratory of J.L. Micol, which led to the isolation of the allelic *orbiculata1–1* (*orb1-1*), *orb1-2* and *orb1-3* mutants, among many others[26]. An additional, loss-of-function allele (*orb1-4*) was identified in the SALK collection of T-DNA insertional mutants (SALK_011035 C). The *orb1* mutations are recessive and cause a similar phenotype, including reduced leaf growth and pale green pigmentation (Fig. 1a–e). We have characterized the *orb1-1*, *orb1-3* and *orb1-4* mutants in more detail. Our measurements of first- and third-node leaves collected 21 days after stratification (das) uncovered a significant reduction in the area of the leaf lamina of the *orb1-1*, *orb1-3* and *orb1-4* mutants ($p < 0.05$; Fig. 2a). In line with this reduction, the basal rosettes of *orb1-1*, *orb1-3* and *orb1-4* were significantly smaller than those of their wild types, the Landsberg *erecta* (L*er*) and Columbia-0 (Col-0) accessions. This growth defect was most severe in the *orb1-4* mutant, whose rosette area was significantly reduced compared to Col-0 ($p < 0.001$; $n = 14$-30; Fig. 2b). The projected area of *orb1-1* and *orb1-3* rosettes was also significantly smaller than that of L*er* rosettes ($p < 0.01$; $n = 14$–30; Fig. 2b). In addition, the *orb1–1*, *orb1-3* and *orb1-4* mutants exhibited reductions in their fresh and dry weights, which were statistically significant throughout plant development ($p < 0.001$; $n = 8$; Fig. 2c,d).

The pale green phenotype of *orb1* mutants was associated with significant reductions in the levels of chlorophyll a, chlorophyll b and carotenoids ($p < 0.05$; $n = 5$; Fig. 3), which were apparent in rosettes harvested between 12 and 24 das. The levels of chlorophyll a in *orb1-1*, *orb1-3* and *orb1-4* rosettes were, respectively, 47.7%, 63.4%, and 52.8% of those of the wild-type control plants. As regards chlorophyll b, its levels were, respectively, 45.8%, 57.9%, and 59.3% of those of the wild type. The carotenoid levels in *orb1-1*, *orb1-3* and *orb1-4* rosettes were, respectively, 55.9%, 74.4%, and 58.5% of those of the wild type (Fig. 4k).

We studied the internal tissues of *orb1* mutants using differential interference contrast (DIC) microscopy (Fig. 4). Rosette leaves from the first and third nodes were cleared using chloral hydrate. Two pictures per leaf were taken halfway along the primary vein and the leaf margin, using 6 leaves per genotype. Differences in the area of palisade mesophyll cells were tested using measurements from $n \geq 600$ cells per genotype. A significant reduction of cell area was observed in palisade mesophyll cells using paradermal sections of leaves from the three mutants ($p < 0.001$; Fig. 4). The ratio of the area of leaf lamina to the area of palisade mesophyll cells, however, was similar in the mutants and the wild type, suggesting that the reduction in leaf size mainly results from the observed reduction in cell area.

ORB1 is the same gene as _Fd-GOGAT1_. To clone the *ORB1* gene, we followed the approach outlined in Mateo-Bonmatí *et al.*[34], with an initial high-resolution linkage mapping step followed by the resequencing of the complete genome. The *ORB1* gene was first mapped between the AthCTR1 marker and the telomere of chromosome 5 by analyzing simple sequence length polymorphic (SSLP) markers in an F_2 mapping population derived from a cross involving the *orb1-1* mutant and wild-type Col-0 plants[35,36]. By genotyping 324 F_2 plants for additional SSLP and insertion/deletion (indel) markers using primers listed in Supplementary Table S1, we placed the mutation between the AthCTR1 and nga225 markers (Fig. 5a). The analysis of the cer455551, cer479319 and cer457348 markers, which are located between AthCTR1 and nga225, allowed us to map the position of the

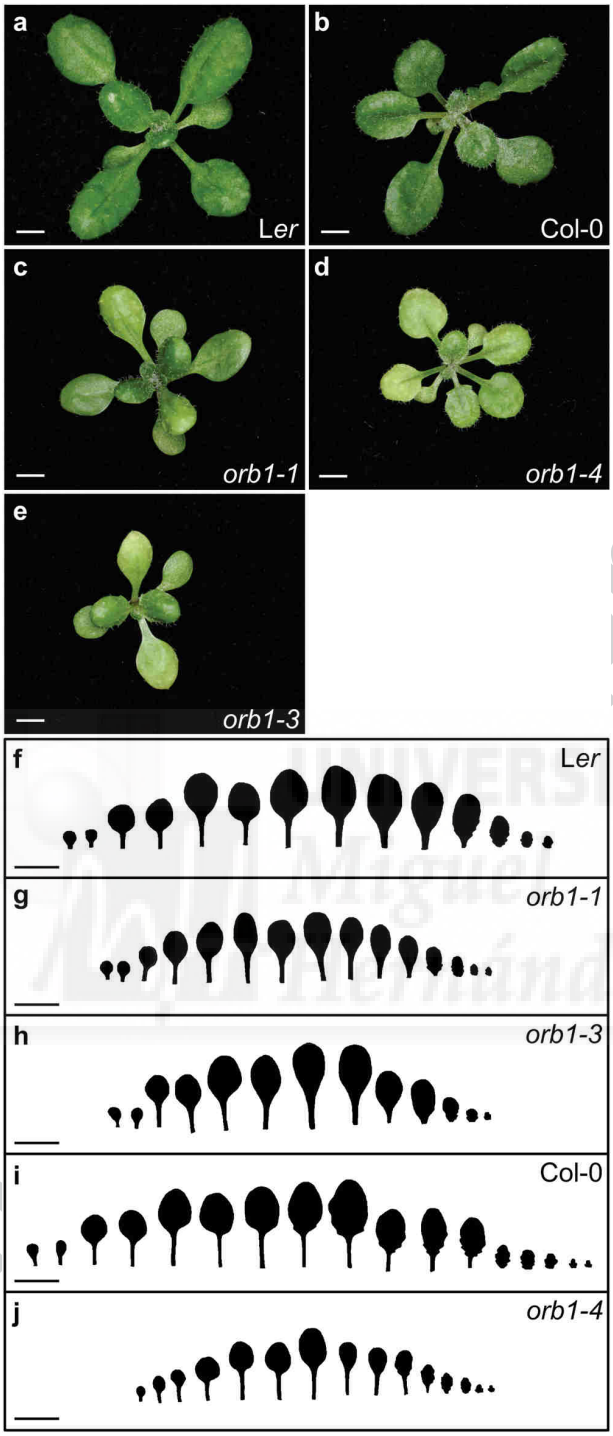


Figure 1. Rosette and leaf shape of *orb1* mutants. The *orb1-1* and *orb1-3* mutants are in a L*er* genetic background, and that of *orb1-4* is Col-0. (**a–e**) Rosette pictures from *orb1* mutants. (**f–j**) Drawings of leaves from *orb1* mutants. Plants were collected **(a–e)** 16 and (f-j) 21 days after stratification (das). Scale bars indicate (**a–e**) 2 mm, and **(f–j)** 1 cm.

orb1-1 mutation between cer455551 and cer479319, in a 222-kb candidate interval that encompasses 64 genes and is roughly delimited by the At5g03870 and At5g04440 genes (Fig. 5a). To identify the causal mutation within this candidate interval, we sequenced the complete *orb1-1* genome with the Illumina HiSeq2000 platform, using 25,864,186 read pairs. Of these, only 18,552,905 read pairs (71.73%) were concordantly aligned to the most recent

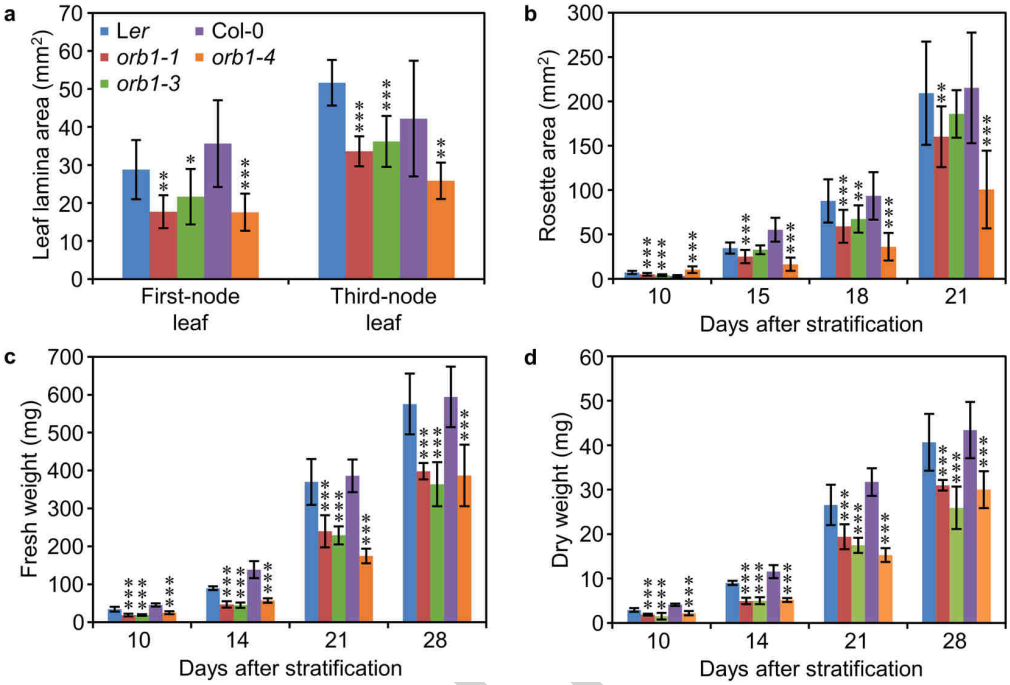


Figure 2. Size and mass of *orb1* mutants. (**a**) Area of the lamina of first- and third-node leaves, (**b**) rosette area, (**c,d**) whole plant (**c**) fresh weight and (**d**) dry weight of *Ler*, *orb1-1*, *orb1-3*, Col-0 and *orb1-4* plants. Error bars indicate standard deviations. Asterisks indicate values significantly different from the corresponding wild type in (**a,c,d**) a Mann-Whitney U-test (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, n = 8–10), and (**b**) a Student's *t*-test (***$p < 0.001$, **$p < 0.01$, n = 14–30).

version of the Arabidopsis nuclear genome then available (TAIR10)[37] using Bowtie 2 (version 2.1.0)[38], yielding a sequencing depth of 28.03×. After discarding the *Ler*/Col-0 polymorphisms, we found that the *orb1-1* mutant carries a G→A transition mutation in the coding region of the At5g04140 gene (Fig. 5c). This mutation was confirmed by conventional Sanger sequencing (Fig. 5b). The At5g04140 gene encodes the chloroplast-localized FERREDOXIN-DEPENDENT GLUTAMINE OXOGLUTARATE AMINOTRANSFERASE 1 (Fd-GOGAT1) protein, also known as GLUTAMATE SYNTHASE 1 (GLU1)[39–41]. Although the TAIR10 annotation includes two different splice forms for this gene, At5g04140.1 (accession number NM_120496.3) and At5g04140.2 (NM_180432.2), we only obtained experimental evidence supporting the At5g04140.1 isoform, as shown by the alignment of RNA sequencing (RNA-seq) read pairs to the reference genome (see below; Fig. 5d). This isoform encompasses 33 exons, encoding a protein that is 1622 amino acids long and has a molecular mass of 176.9 kDa (Fig. 5d).

The ORB1 protein contains a glutamine amidotransferase type 2 domain (Class-II or type 2 GATase domain; IPR017932)[42] and a glutamate synthase domain (IPR002932), which are connected by a glutamate synthase central-N domain (IPR006982). The glutamate synthase domain contains a putative flavin mononucleotide (FMN) binding site and a Fe-S cluster[43]. The hydrolysis of L-glutamine in the amidotransferase domain yields ammonium and L-glutamate. The ammonium is then combined with 2-oxoglutarate in the FMN binding domain to produce a second molecule of L-glutamate[44]. The G→A transition mutation found in the *orb1-1* mutant damages the first position of codon 579 of the At5g04140.1 coding sequence (exon 10), causing a lysine (K) for glutamate (E) substitution in the glutamate synthase central-N domain (Fig. 5c). A G→A transition mutation was found in the *orb1-2* mutant. This mutation alters the third position of codon 295, replacing a tryptophan (W) with a stop codon. A C→T transition mutation was found in the *orb1-3* mutant. This mutation damages the second position of codon 149 (exon 2) of the gene, causing a phenylalanine (F) for serine (S) substitution. The *orb1-2* and *orb1-3* mutations affect the glutamine amidotransferase type 2 domain. The T-DNA insertion of *orb1-4* is located in intron 28, according to a flanking sequence recovered from the left border of the T-DNA that is available from GenBank (accession number BH251122.1). Complementation crosses involving the EMS-induced *orb1-1* and *orb1-3* alleles as well as the insertional *orb1-4* allele showed that all these mutations damage the same gene (Fig. 6).

Expression pattern of the *ORB1* gene. The Arabidopsis eFP Browser (http://www.bar.utoronto.ca/)[45] and Transcriptome Variation Analyses (TraVA; http://travadb.org)[46] databases indicate that the expression of *ORB1* is most intense in above-ground tissues. To define the expression pattern of *ORB1* more precisely, we generated an $ORB1_{pro}$:*GUS* construct. By transforming *Ler* plants, we isolated two independent transformants expressing the $ORB1_{pro}$:*GUS* transgene. In young seedlings (7 das), we observed GUS signal in cotyledons, leaf primordia,

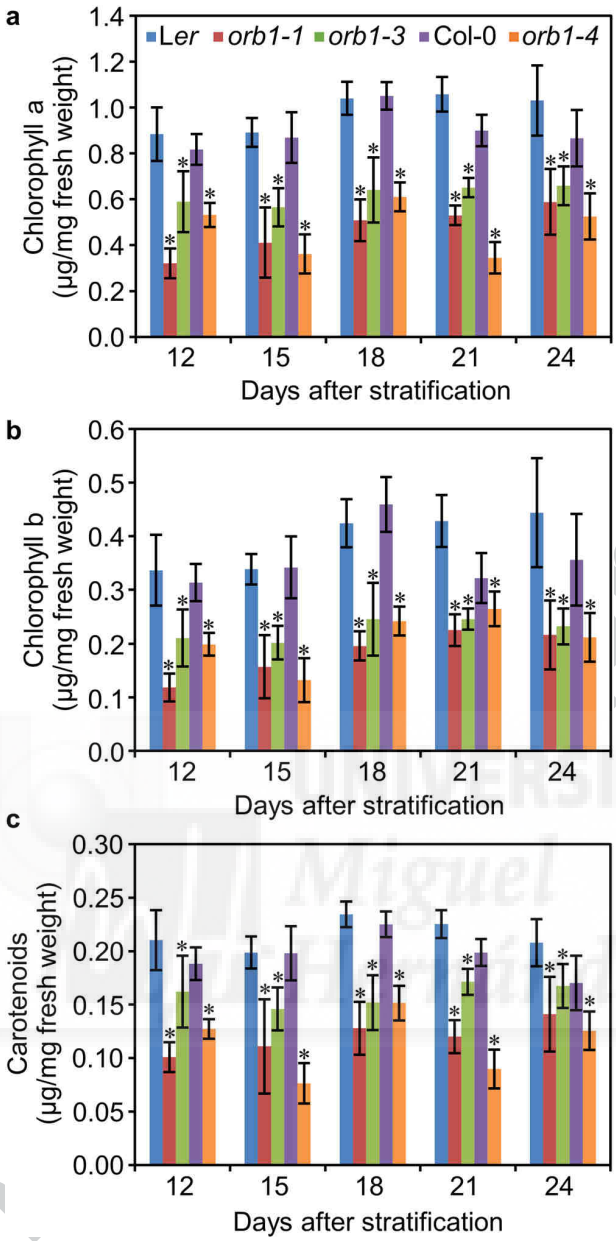


Figure 3. Pigment content in *orb1* mutants. Concentration of (**a**) chlorophyll a, (**b**) chlorophyll b, and (**c**) carotenoids in L*er*, *orb1-1*, *orb1-3*, Col-0 and *orb1-4* above-ground tissues. Plants were collected 12, 15, 18, 21 and 24 das. Error bars indicate standard deviations. Asterisks indicate values significantly different from the corresponding wild type in a Mann-Whitney U-test (*$p < 0.05$, $n = 5$).

shoot apical meristems and roots (Fig. 7a). In roots, the highest expression was detected at the vascular cylinder and root apices (Fig. 7c). In cotyledons and leaves, the GUS signal was most intense at the veins and the hydathodes, but mesophyll cells were also stained (Fig. 7b,d–f). This expression pattern matches the *GLU1* expression pattern previously described by other authors in tobacco and Arabidopsis[47,48]. In cauline leaves, no expression of the transgene was observed (Fig. 7g). In immature flowers, we observed GUS staining at the sepals and the style (Fig. 7h). In mature flowers, GUS signal was detected at the anther filaments, the style and the venation pattern of petals and sepals (Fig. 7i). In immature siliques, GUS expression was intense (Fig. 7j), but in mature siliques we only observed GUS signal at the receptacle (Fig. 7k).

RNA-seq analysis of the *orb1–3* mutant. To identify global changes in the transcriptome, we sequenced RNA samples isolated from *orb1-3* and Landsberg *erecta* (L*er*) vegetative rosettes, including three biological replicates of each genotype, using a strand-specific RNA-seq protocol. The reads were analysed using the Tophat2/Cufflinks pipeline[49–51], indicating the expression levels for each gene as FPKM values (fragments per kilobase

www.nature.com/scientificreports/

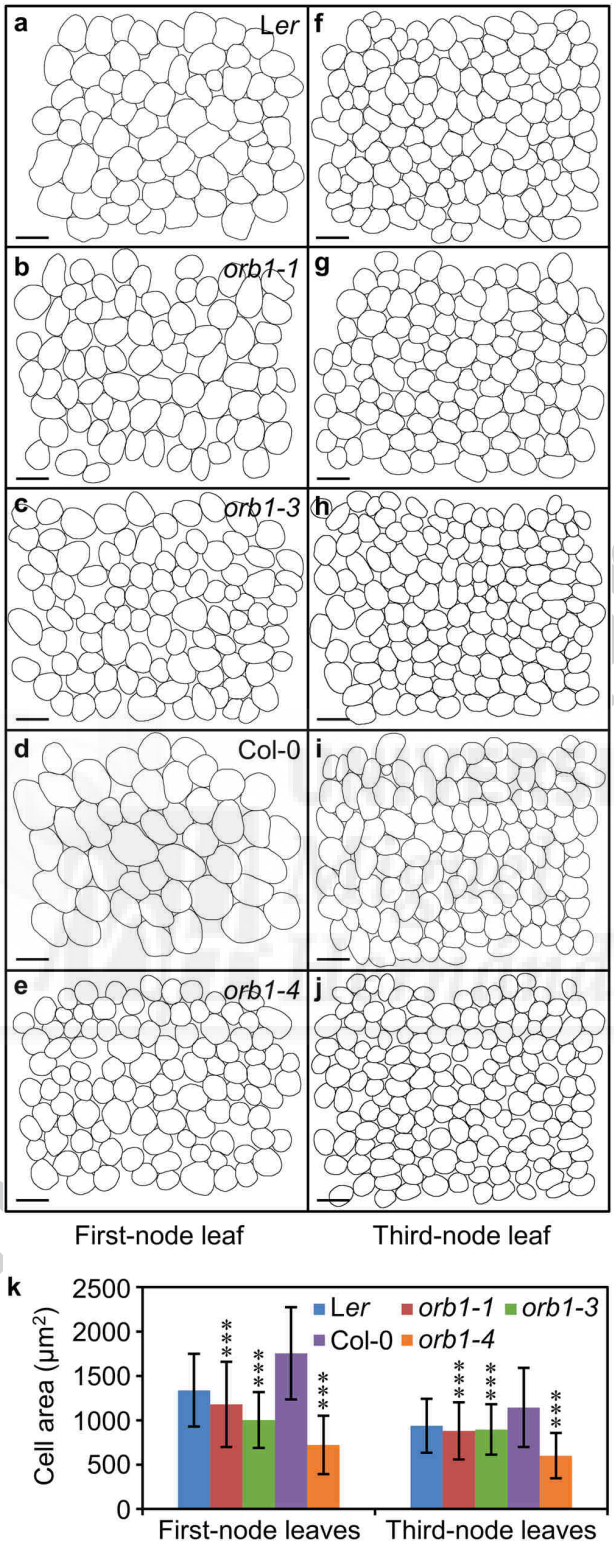


Figure 4. Morphometry of palisade mesophyll cells in *orb1* mutants. (**a–j**) Diagrams of the subepidermal layer of palisade mesophyll cells from (**a–e**) first- and (**f–j**) third-node leaves. (**k**) Palisade mesophyll cell area in first and third-node leaves. Plants were collected 21 das. Scale bars indicate 50 µm. Asterisks indicate values significantly different from the corresponding wild type in a Student's *t*-test (***$p < 0.001$, $n \geq 600$).

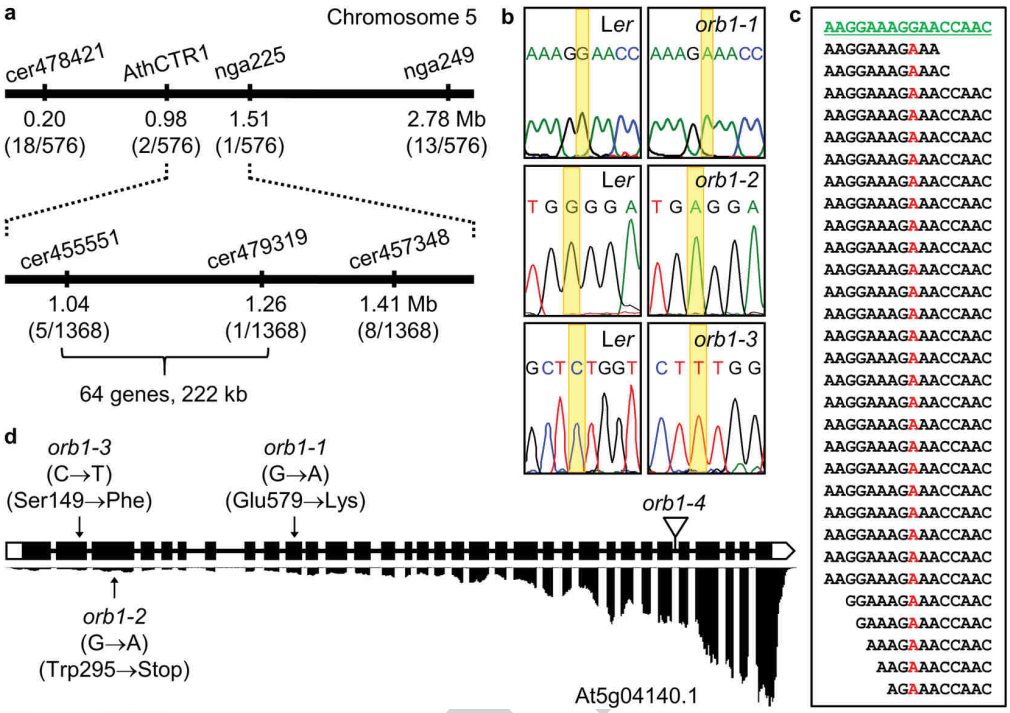


Figure 5. Positional cloning of the *ORB1* gene. (**a**) A mapping population of 684 F_2 plants derived from an *orb1-1* × Col-0 cross allowed us to define a candidate interval of 222 kb on chromosome 5. Names and physical map positions of the molecular markers used for linkage analysis are shown. The number of recombinant chromosomes found and the total of chromosomes analyzed are indicated in parentheses. (**b**) Electropherograms showing the point mutations in the *orb1-1*, *orb1-2* and *orb1-3* mutants. (**c**) Pileup of reads derived from the *orb1-1* genome. The reference sequence is shown in green, and the *orb1-1* mutation is highlighted in red. (**d**) Structure of the At5g04140.1 isoform of the *ORB1* gene and alignment of RNA-seq read pairs to the At5g04140.1 isoform, with indication of the nature and position of the *orb1* mutations. Boxes and lines between boxes indicate exons and introns, respectively. White boxes represent the 5′- and 3′-UTRs. A triangle indicates the T-DNA insertion in *orb1-4*.

of transcript per million fragments mapped; Supplementary Table S2). In line with the results of a microarray analysis previously performed using a T-DNA allele of the same gene (*glu1-2*, SALK_019917)[33], our results demonstrate that an extensive reprogramming of the transcriptome occurs in response to the *orb1-3* mutation. Out of 18856 different genes tested for differential expression levels, 6303 genes were found to be significantly differentially expressed between *orb1-3* and L*er* at a false discovery rate (FDR) threshold of 5%. Of these, 2833 genes were expressed at higher levels and 3470 genes at lower levels in *orb1-3* rosettes. When we used a stricter FDR threshold of 1%, 1862 genes and 2522 genes were respectively found to be expressed at higher and lower levels. We compared our results to the set of differentially expressed genes identified in the study of Kissen *et al.*[33]. Because the two studies differ in aspects such as ecotype (L*er* in this work versus Col-0 in that of Kissen *et al.*), choice of tissue (complete basal rosettes versus leaves), light regime (continuous lighting versus 16-hour photoperiod) and medium composition (half-strength versus full-strength MS salts), we expected to see differences in our results. A total of 1232 genes (28.1% of the 4384 differentially expressed genes selected at an FDR of 1%) were shared between the two studies, including 548 upregulated genes (29.4% of 1862) and 684 downregulated genes (25.6% of 2674). We expected that this set comprises genes whose expression reproducibly changes as a consequence of the *orb1* (*glu1*) mutations, regardless of other endogenous and environmental factors.

Different from the *glu1-2* mutant, which is a knock-down mutation[33], our *orb1-3* point mutation did not significantly affect the abundance of its own transcripts (Fig. 8a), suggesting that the expression of *ORB1*/*Fd-GOGAT1* is not induced by glutamine, one of the reaction substrates, which is known to accumulate at increased levels in loss-of-function *glu1* mutants[28, 33, 52]. In contrast to previous results[33], we found three genes encoding glutamine synthetases (At5g37600, At3g17820 and At5g16570) expressed at significantly reduced levels in *orb1-3* leaves, suggesting that the expression of these enzymes is subjected to product inhibition (Fig. 8b). We also found a drastic increase in the expression of *GLUTAMINE-DEPENDENT ASPARAGINE SYNTHASE 1* (*ASN1*, At3g47340), one of the three genes that encode glutamine-dependent asparagine synthetases in the Arabidopsis genome[53, 54]. Although *ASN1* was expressed at relatively low levels in the wild type, the *orb1-3* mutation caused a 60-fold increase in its expression (Fig. 8b). Asparagine synthetases, such as ASN1, transfer an amide group from glutamine to aspartate, yielding asparagine and glutamate. The expression of the genes encoding the small and large subunits of CARBAMOYL PHOSPHATE SYNTHETASE (CPS; *VEN3* and *VEN6*, respectively)[21] was also found to be increased in *orb1-3* rosettes. Because CPS catalyzes the production of glutamate and carbamoyl phosphate

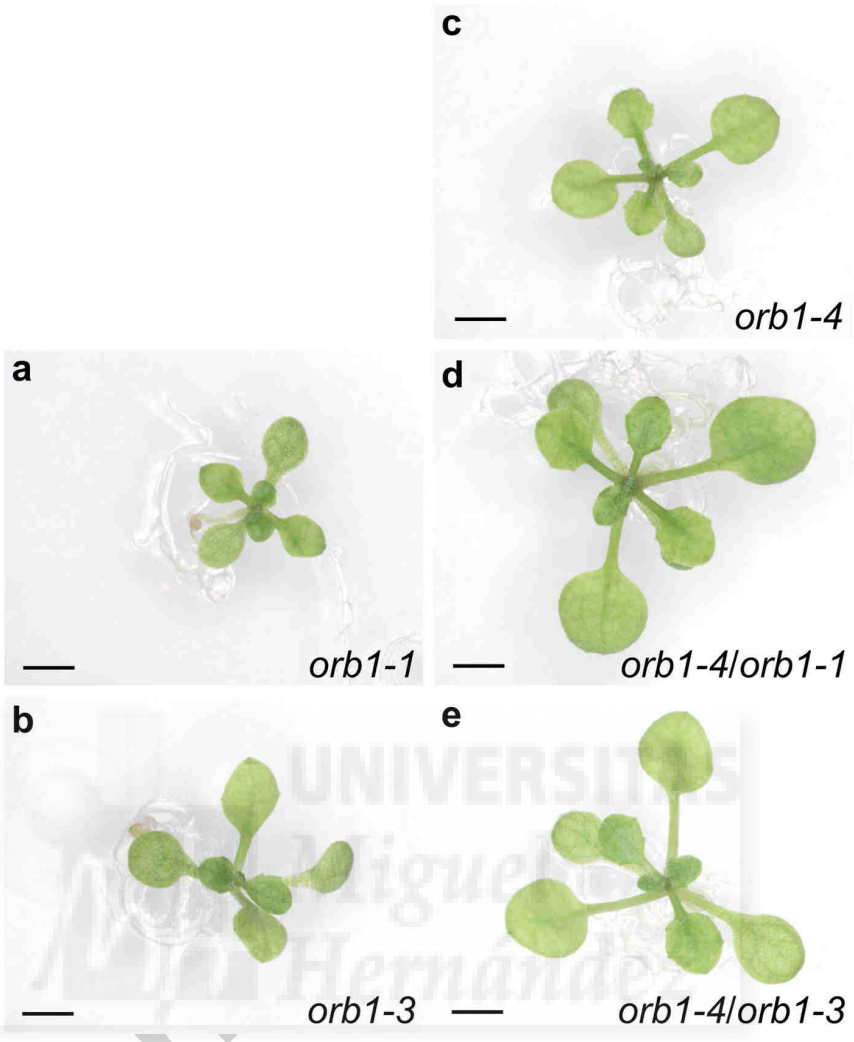


Figure 6. Allelism test of *orb1-1* and *orb1-3* point mutations with the insertional *orb1-4* mutation. Pictures were taken 14 das. Scale bars indicate 2 mm.

from glutamine, the enhanced levels of CPS and ASN1 might help to compensate the lack of Fd-GOGAT activity by producing glutamate and by reducing the elevated glutamine levels. Asparagine can in turn be converted into different amino acids by the activity of *ALANINE:GLYOXYLATE AMINOTRANSFERASE 1* (*AGT1*, At2g13360), which encodes an asparagine aminotransferase in the Arabidopsis genome[55], or into aspartic acid by the activity of two asparaginases, ASPARAGINASE A1 (ASPGA1, At5g08100) and ASPGB1 (At3g16150). Different from *ASN1*, the *AGT1* gene was expressed at high levels both in the wild type and the mutants (with FPKM values between 902 and 1235), suggesting that the conversion of Asn into Asp is not a limiting step. The expression of the two asparaginases was similarly not affected by the *orb1-3* mutation.

Two (At5g18170 and At5g07440) out of the three genes encoding glutamate dehydrogenases were expressed at elevated levels in the *orb1-3* mutant. In particular, the expression levels of *GLUTAMATE DEHYDROGENASE 2* (*GDH2*, At5g07440) shifted from 29 in Col-0 to 177 in the mutant (a fold change of ~6.1). Glutamate dehydrogenases convert glutamate into 2-oxoglutarate and hence, this increase in *GDH2* expression is at first sight unexpected because the increased GDH activity would further contribute to lowering the amount of glutamate. Counter-intuitively, the conversion of glutamate into 2-oxoglutarate might uncover a cellular strategy to duplicate the cellular pool of glutamate, as glutamate synthases such as Fd-GOGAT yield two glutamate molecules from each molecule of 2-oxoglutarate, in a reaction that also requires a molecule of glutamine.

To survey the metabolic pathways affected in the *orb1-3* mutant, we mapped the differentially expressed genes detected in our RNA-seq dataset using the KEGG PATHWAY online tool (http://www.genome.jp/kegg/pathway.html)[56,57]. In addition to the above-described genes involved in nitrogen metabolism, one noticeable characteristic of the map was the opposite regulation of enzymes in the biosynthesis and degradation pathways of fatty acids. While most genes encoding enzymes involved in their biosynthesis were overexpressed (for instance

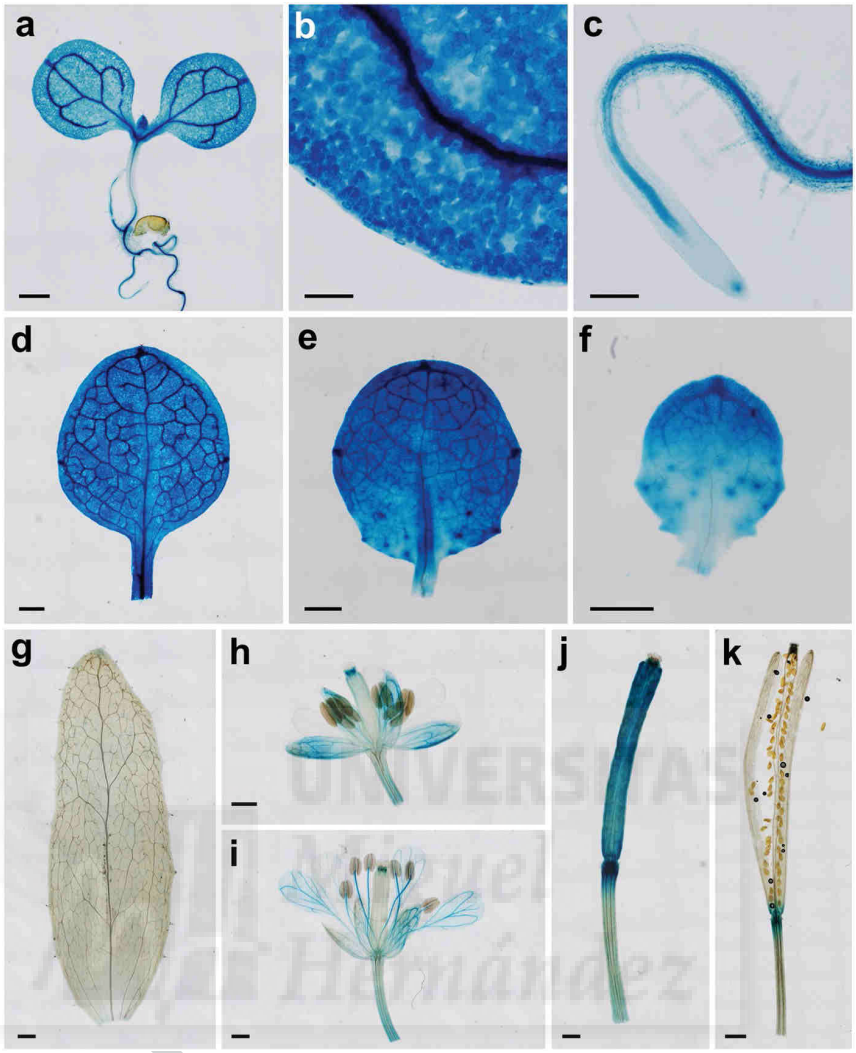


Figure 7. Visualization of *ORB1*$_{pro}$*:GUS* activity on a wild-type L*er* background. (**a**) Seedling, (**b**) cotyledon detail (vein and palisade mesophyll), (**c**) root, (**d**) first-node leaf, (**e**) third-node leaf, (**f**) fifth-node leaf, (**g**) cauline leaf, (**h**) immature flower, (**i**) mature flower, (**j**) immature silique, and (**k**) mature silique. Pictures were taken (**a–c**) 7, (**d–f**) 14, and (**g–k**) 40 das. Scale bars indicate (**k**) 1 mm, (**a,d–j**) 500 μm, and (**b,c**) 100 μm.

At3g04000, At5g10160, At2g05990 and At2g30200), most genes involved in their degradation, such as At4g29010, were repressed (Supplementary Fig. S1).

As also indicated by the Gene Ontology (GO) analysis below, we found many genes encoding subunits of the cytosolic ribosome among the set of upregulated genes: as many as 208 genes displayed a concerted increase in their expression levels (Fig. 8e and Supplementary Fig. S2). Although this result remained hidden in the supplemental tables of Kissen *et al.*, our results show that ribosomal proteins constitute the most abundant functional category among the genes shared by both studies. Protein biosynthesis is, undeniably, an important contributor to the nitrogen balance of the cell. Hence, this concerted upregulation possibly reflects a cellular response to altered amino acid levels when Fd-GOGAT1 function is impaired, and a concomitant altered translation.

GO analysis and singular enrichment analysis of differentially expressed genes. We assigned GO terms to the complete set of genes differentially expressed between L*er* and *orb1-3* using the GO annotation of the Arabidopsis genome that is available from TAIR. The 4384 differentially expressed genes (at an FDR of 1%) were assigned a total of 1218 GO terms from the 'molecular function' subontology, 1771 terms from the 'biological process' subontology and 337 terms from the 'cellular component' subontology (up to a total of 3326 GO terms). We next considered the distribution of GO terms in the overexpressed and the underexpressed genes separately. The overexpressed genes were assigned 1113 'biological process' terms, 269 'cellular component' terms and 644 'molecular function' terms. The set of underexpressed genes was assigned 1294 'biological process' terms, 192 'cellular component' terms and 124 'molecular function' terms. A total of 636 terms (35.91%) from the

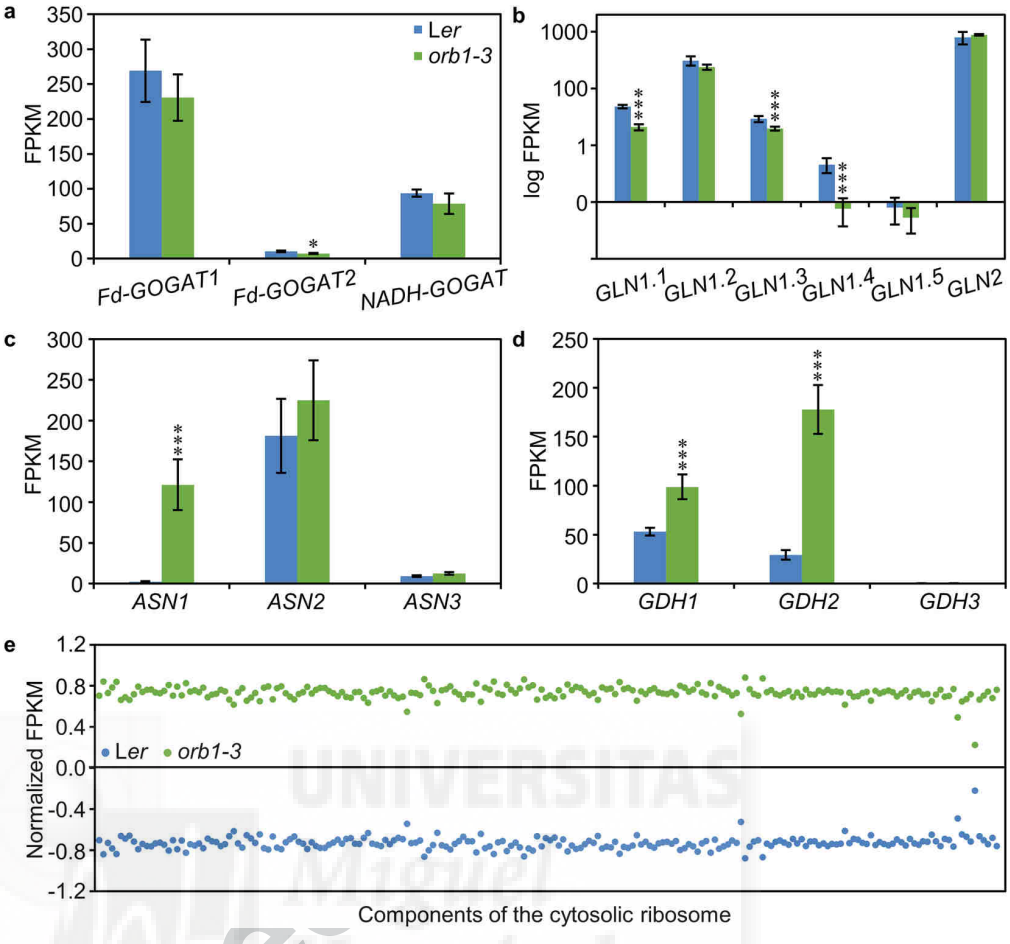


Figure 8. Differences in the expression of genes encoding glutamate-related enzymes and components of the cytosolic ribosome between L*er* and *orb1-3* plants. Expression levels of genes encoding (**a**) glutamate synthases (*Fd-GOGAT1*, *Fd-GOGAT2* and *NADH-GOGAT*), (**b**) glutamine synthetases (*GLN1.1*, *GLN1.2*, *GLN1.3*, *GLN1.4*, *GLN1.5* and *GLN2*), (**c**) asparagine synthetases (*ASN1*, *ASN2* and *ASN3*), (**d**) glutamate dehydrogenases (*GDH1*, *GDH2* and *GDH3*), and (**e**) several components of the cytosolic ribosome in L*er* and *orb1-3* plants, expressed as FPKM (fragments per kilobase of transcript per million fragments mapped). Error bars indicate standard deviations. Asterisks indicate values significantly different from L*er* as determined by Cuffdiff (*$q < 0.05$, ***$q < 0.001$, n = 3). The expression levels of the genes encoding components of the cytosolic ribosome were normalized based on the mean and standard deviation of the FPKM values obtained for each gene in all samples analyzed.

'biological process' subontology, 124 terms (36.8%) from the 'cellular component' subontology, and 350 terms (28.74%) from the 'molecular function' subontology were shared by the sets of underexpressed and overexpressed genes.

We performed singular enrichment analysis (SEA) for the GO terms assigned to the complete set of differentially expressed genes (Supplementary Table S3a). Thirty-three GO terms in the 'cellular component' subontology were significantly enriched. The most significantly enriched term was 'cytosolic ribosome' (GO:0022626), as 77.35% out of the 287 genes containing this term in the background set were differentially expressed. This term was followed by other highly significantly enriched terms related to ribosomes, such as 'ribosomal subunit' (GO:0033279), 'ribosome' (GO:0005840), 'cytosolic large ribosomal subunit' (GO:0022625), 'large ribosomal subunit' (GO:0015934), 'cytosolic small ribosome subunit' (GO:0022627), 'small ribosomal subunit' (GO:0015935), and 'ribonucleoprotein complex' GO:0030529). One hundred and eleven GO terms in the 'biological process' subontology were significantly enriched. In line with the most significantly enriched terms in the 'cellular component' subontology, the most significantly enriched term was 'translation' (GO:0006412), with 269 differentially expressed genes (45.98% out of the 585 genes containing this term in the background set). In line with previous results[33], other enriched GO terms from this subontology were related to the responses to biotic and abiotic stimuli, including 'defense response' (GO:0006952), 'response to biotic stimulus' (GO:0009607), 'response to stress' (GO:0006950), 'response to other organism' (GO:0051707), 'response to stimulus' (GO:0050896), 'response to wounding' (GO:0009611), 'response to external stimulus' (GO:0009605), 'response to chemical stimulus'

(GO:0042221), 'response to bacterium' (GO:0009617), 'response to organic substance' (GO:0010033), 'response to fungus' (GO:0009620), and 'response to salicylic acid stimulus' (GO:0009751). 30 GO terms were significantly enriched in the 'molecular function' subontology. The most significantly enriched term was 'structural constituent of ribosome' (GO:0003735), in agreement with the most significantly enriched terms in the 'biological process' and 'cellular component' subontologies. This GO term was ascribed to 227 different genes, out of the 327 genes that contained this term in the background set (69.42%). A variety of different catalytic and binding activities were also enriched in this set.

We also performed enrichment analysis on the sets of overexpressed and underexpressed genes separately (Supplementary Table S3b and S3c, respectively). For the overexpressed genes, we found 21 enriched GO terms in the 'biological process' subontology. Many of these terms matched the terms identified when the analysis was performed with the complete set of differentially expressed genes. At the top of the list were terms such as 'translation' (258 genes), 'ribosome biogenesis' (113 genes) and 'ribonucleoprotein complex biogenesis' (118 genes). These numbers closely matched the numbers obtained when all the differentially expressed genes were taken together, showing that the vast majority of the genes involved in translation and ribosome biogenesis have increased expression levels in the *orb1-3* mutant. This increase suggests that low amino acid levels compromise translation and trigger a cellular response aimed at compensating this defect. Such response would include a coordinated overexpression of most components of the large and small subunits of the cytosolic ribosome. Indeed, previous authors have found that genes encoding ribosomal proteins and other factors related to ribosome biogenesis are co-regulated by the $p33^{TCP20}$ protein, which binds to the GCCCR motif present in their promoters, providing a mechanism to ensure their appropriate stoichiometry. This motif was also found in the promoter of the cyclin *CYCB1;1* gene, linking ribosome availability with cell division[58]. In line with our results, mutations in the *ARABIDOPSIS PUMILIO 23* (*APUM23*) gene, which encodes an RNA-binding domain protein that functions in rRNA processing and ribosome assembly, are also known to cause a similar overexpression of genes encoding ribosomal proteins[59]. We found 43 enriched GO terms in the 'cellular component' subontology, the most significantly overrepresented of which being 'cytosolic ribosome' (GO:0022626; p = 2.71e-143), followed by other terms related to ribosomal function. Eight GO terms were found to be enriched in the 'molecular function' subontology. The most significantly overrepresented term was 'structural constituent of ribosome' (GO:0003735). For the underexpressed genes, we found 133 enriched GO terms in the 'biological process' subontology, 50 in the 'molecular function' subontology, and 10 in the 'cellular component' subontology. The most significantly enriched terms in 'biological process' were 'defense response' (GO:0006952) and other terms related to responses to biotic and abiotic stimuli. The 'molecular function' terms included different binding and catalytic activities, as well as an important number of transcriptional regulators (246). The most significantly enriched terms in the 'cellular component' subontology were related to membranes.

We also performed SEA using the set of 1232 differentially expressed genes shared by both studies (Supplementary Figure S3). In this smaller set, terms related to translation and the function of ribosomes appeared enriched to a greater extent than in the broader set of 4384 differentially expressed genes selected at an FDR of 1%, discussed above (Supplementary Table S3). Unlike the terms related to ribosomes, numerous terms turned out to be not enriched when we considered the smaller set of differentially expressed genes shared by both studies. Lost terms included many related to the response to various types of biotic and abiotic stimuli, suggesting that the stress responses largely depend on the genetic background (i.e. they are accession-specific) or other external factors. The set of shared upregulated genes was also enriched in ribosome-related terms, and included some new terms that are not enriched when the upregulated and downregulated genes were taken together. Examples of such additional terms are 'rRNA processing' (GO:0006364) and 'rRNA metabolic process' (GO:0016072). In link with the pale green phenotype of the mutants, the set of shared downregulated genes was enriched in terms related to plastid function and photosynthesis, which were not enriched in the set of all the differentially expressed genes taken together.

Concluding Remarks

Glutamic acid plays a central role in the assimilation of nitrogen and the biosynthesis of amino acids. Our results show that altered glutamate biosynthesis impairs plant growth and triggers a concerted transcriptional response that includes the upregulation of more than a hundred genes required for ribosome biogenesis. Understanding the relationship between protein synthesis and amino acid availability, and how they are regulated during plant growth and development, should help to engineer plants with increased biomass.

Methods

Plant material, growth conditions and crosses. Seeds of the Columbia-0 (Col-0) and Landsberg erecta (L*er*) wild-type accessions of *Arabidopsis thaliana* (L.) Heynh., as well as the T-DNA line SALK_011035 (N511035), were obtained from the Nottingham Arabidopsis Stock Centre (NASC). *orb1-1*, *orb1-2* and *orb1-3* mutants were isolated after a EMS-induced mutagenesis of L*er* seeds[26]. Plant culture and allelism tests were performed as reported in Berná *et al.* and Ponce *et al.*[26,60].

Plant morphometry. Pictures and drawings from Arabidopsis rosettes, leaves and palisade mesophyll cells were obtained as previously described[61]. Rosette area and mesophyll cell area measurements were performed using the NIS Elements AR 3.1 image analysis package (Nikon).

Fresh weight, dry weight and pigment concentration. Fresh and dry weight quantification, as well as pigment content determination of L*er*, *orb1-1*, *orb1-3*, Col-0 and *orb1-4* at different ages were performed as previously reported in Casanova-Sáez *et al.*[62].

Statistical analyses. We carried out statistical analyses to compare the phenotypic traits of *orb1-1*, *orb1-3* and *orb1-4* mutants with their corresponding wild-type plants. We used the Mann-Whitney U-test for 10 or less replicates and the Student's *t*-test for more than 10 replicates.

GUS staining. To induce GUS activity, plant tissues were incubated in 90% acetone for 10 min at −20 °C, and then transferred into GUS stain solution (2 mM 5-bromo-4-chloro-3-indolyl-β-glucuronic acid, 50 mM sodium phosphate, pH 7.2, 5 mM potassium ferrocyanide, 50 mM potassium ferricyanide, and 0.2% Triton X-100). Once in GUS stain solution, samples were infiltrated under vacuum for 20 min in the dark and then dark-incubated for 2 hours at 37 °C. GUS stain solution was removed and stained-tissues were cleared with 70% ethanol. Samples were mounted in an 8:2:1 (chloral hydrate:glycerol:water) solution and examined with a Nikon D-Eclipse C1 microscope.

Identification of the *orb1-1* mutation. We combined map-based cloning and next-generation sequencing approaches to identify the gene affected by the *orb1-1*, *orb1-2* and *orb1-3* mutations. We performed the low-resolution mapping of the *orb1-1* mutation as described in Ponce *et al.*[35]. We used the SSLP and insertion/deletion (indel) markers cer478421, nga225, nga249, cer455551, cer479319 and cer457348 to reduce the candidate interval to a 222-kb region, by genotyping F_2 plants derived from an *orb1-1* × Col-0 cross using PCR or Sanger sequencing. Primers used for fine-mapping of the At5g04140 gene are shown in Supplementary Table S1.

Genome resequencing and RNA-seq. To identify the *orb1-1* mutation within the candidate region, we resequenced the genome of the *orb1-1* mutant at the Beijing Genomics Institute (BGI). Genomic DNA was isolated as described in Mateo-Bonmatí *et al.*[34] and sequenced using the Illumina HiSeq2000 platform with 90 nt-long paired-end reads. The paired-end reads obtained from the resequencing of the *orb1-1* genome were aligned to the TAIR10 reference genome using Bowtie 2 (version 2.1.0)[38]. The resulting alignment was visualized using Tablet (version 1.15.09.01)[63].

Total RNA was extracted from ~100 mg of L*er* and *orb1-3* rosettes collected 16 das using TRI Reagent (Sigma-Aldrich). Library construction and RNA sequencing was performed by StabVida (Caparica, Portugal), using 3 biological replicates per genotype. Libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina). Paired-end sequencing was performed by multiplexing the libraries in an Illumina HiSeq2500 system. The sequence alignment and the quantification of gene expression levels were performed as previously described in Mandel *et al.*[64]. Reads were aligned to the TAIR10 version of the Arabidopsis genome using TopHat v.2.0.12 and Bowtie2 v.2.1.0 with the following parameters: -p 8 (number of threads), supplying the annotation for the nuclear genes in a general feature format (GFF) file, and using default values for all other parameters. The resulting alignments were quantified with Cuffdiff v.2.2.1 masking the rRNA, tRNA, snRNA and snoRNA genes for quantification purposes. Differentially expressed genes were mapped to metabolic pathways using the KEGG PATHWAY online tool (http://www.genome.jp/kegg/pathway.html)[56,57].

SEA was performed using agriGO (http://bioinfo.cau.edu.cn/agriGO/)[65] with the default options (statistical test: hypergeometric, multi-test adjustment method: Yekutieli, significance level: 0.01). The latest release of TAIR's GO annotation was downloaded from TAIR on 13/02/2017. For the SEA, we separately used the sets of underexpressed and overexpressed genes as well as the complete list of differentially expressed genes as queries, including customized annotation data taken from TAIR[66]. As the customized annotated reference, we used the set of genes that had been tested for differential gene expression (marked "OK" in the output of Cufflinks[50,67]).

Constructs and plant transformation. A 1943 bp fragment containing the intergenic region between At5g04130 and At5g04140 was amplified from Col-0 genomic DNA using the Phusion polymerase (Thermo Scientific) and the ORB1-pro-F and ORB1-pro-R primers (which contain *att*B1 and *att*B2 sites; Supplementary Table S1). This fragment was cloned into pGEM-T Easy221 vector (provided by Prof. B. Scheres) using BP clonase II (Life technologies). The insert was transferred to pMDC163[68] using LR clonase II (Life technologies) to generate the $ORB1_{pro}$*:GUS* construct. L*er* plants were transformed with the *Agrobacterium tumefaciens* strain C58C1 using the floral dip method[69]. T_1 transgenic plants were selected on Petri dishes supplemented with 15 $\mu g \cdot ml^{-1}$ hygromycin B (Invitrogen).

References

1. Gonzalez, N., Vanhaeren, H. & Inzé, D. Leaf size control: complex coordination of cell division and expansion. *Trends Plant Sci.* **17**, 332–340 (2012).
2. Hepworth, J. & Lenhard, M. Regulation of plant lateral-organ growth by modulating cell number and size. *Curr. Opin. Plant Biol.* **17**, 36–42 (2014).
3. Lewis, M. W. & Hake, S. Keep on growing: building and patterning leaves in the grasses. *Curr. Opin. Plant Biol.* **29**, 80–86 (2016).
4. Gallardo, F. *et al*. Expression of a conifer glutamine synthetase gene in transgenic poplar. *Planta* **210**, 19–26 (1999).
5. Gastal, F. & Nelson, C. J. Nitrogen use within the growing leaf blade of tall fescue. *Plant Physiol.* **105**, 191–197 (1994).
6. Granier, C. & Tardieu, F. Multi-scale phenotyping of leaf expansion in response to environmental changes: the whole is more than the sum of parts. *Plant Cell Environ.* **32**, 1175–1184 (2009).
7. Lewis, C. E., Noctor, G., Causton, D. & Foyer, C. H. Regulation of assimilate partitioning in leaves. *Funct. Plant Biol.* **27**, 507–519 (2000).
8. Sulpice, R. *et al*. Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions. *Plant Physiol.* **162**, 347–363 (2013).
9. Cho, H. T. & Cosgrove, D. J. Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **97**, 9783–9788 (2000).
10. Gonzalez, N. *et al*. Increased leaf size: different means to an end. *Plant Physiol.* **153**, 1261–1279 (2010).
11. Horiguchi, G., Kim, G. T. & Tsukaya, H. The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J.* **43**, 68–78 (2005).
12. Hu, Y., Poh, H. M. & Chua, N. H. The Arabidopsis *ARGOS-LIKE* gene regulates cell expansion during organ growth. *Plant J.* **47**, 1–9 (2006).

13. Mizukami, Y. & Fischer, R. L. Plant organ size control: *AINTEGUMENTA* regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. USA* **97**, 942–947 (2000).
14. White, D. W. *PEAPOD* regulates lamina size and curvature in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **103**, 13238–13243 (2006).
15. Disch, S. *et al*. The E3 ubiquitin ligase BIG BROTHER controls Arabidopsis organ size in a dosage-dependent manner. *Curr. Biol.* **16**, 272–279 (2006).
16. Xia, T. *et al*. The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in Arabidopsis. *Plant Cell* **25**, 3347–3359 (2013).
17. Foyer, C. H. & Ferrario, S. Modulation of carbon and nitrogen metabolism in transgenic plants with a view to improved biomass production. *Biochem. Soc. Trans.* **22**, 909–915 (1994).
18. Van Dingenen, J., Blomme, J., Gonzalez, N. & Inzé, D. Plants grow with a little help from their organelle friends. *J. Exp. Bot.* **67**, 6267–6281 (2016).
19. Casanova-Sáez, R., Candela, H. & Micol, J. L. Combined haploinsufficiency and purifying selection drive retention of RPL36a paralogs in Arabidopsis. *Sci. Rep.* **4**, 4122 (2014).
20. Horiguchi, G. *et al*. Differential contributions of ribosomal protein genes to *Arabidopsis thaliana* leaf development. *Plant J.* **65**, 724–736 (2011).
21. Mollá-Morales, A. *et al*. Analysis of *ven3* and *ven6* reticulate mutants reveals the importance of arginine biosynthesis in Arabidopsis leaf development. *Plant J.* **65**, 335–345 (2011).
22. Quesada, V. *et al*. *PORPHOBILINOGEN DEAMINASE* deficiency alters vegetative and reproductive development and causes lesions in Arabidopsis. *PLoS One* **8**, e53378 (2013).
23. Rubio-Díaz, S. *et al*. Cell expansion-mediated organ growth is affected by mutations in three *EXIGUA* genes. *PLoS One* **7**, e36500 (2012).
24. Stepanova, A. N., Hoyt, J. M., Hamilton, A. A. & Alonso, J. M. A link between ethylene and auxin uncovered by the characterization of two root-specific ethylene-insensitive mutants in Arabidopsis. *Plant Cell* **17**, 2230–2242 (2005).
25. Van Minnebruggen, A. *et al*. The *ang3* mutation identified the ribosomal protein gene *RPL5B* with a role in cell expansion during organ growth. *Physiol. Plant.* **138**, 91–101 (2010).
26. Berná, G., Robles, P. & Micol, J. L. A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. *Genetics* **152**, 729–742 (1999).
27. Lea, P. J. & Miflin, B. J. Alternative route for nitrogen assimilation in higher plants. *Nature* **251**, 614–616 (1974).
28. Ferrario-Méry, S. *et al*. Modulation of amino acid metabolism in transformed tobacco plants deficient in Fd-GOGAT. *Plant Soil* **221**, 67–79 (2000).
29. Miflin, B. J. & Lea, P. J. *Ammonia assimilation*. (Academic Press, 1980).
30. Ghaffari, M. R. *et al*. The metabolic signature of biomass formation in barley. *Plant Cell Physiol.* **57**, 1943–1960 (2016).
31. Stitt, M. *et al*. Steps towards an integrated view of nitrogen metabolism. *J. Exp. Bot.* **53**, 959–970 (2002).
32. Yamaya, T. *et al*. Genetic manipulation and quantitative-trait loci mapping for nitrogen recycling in rice. *J. Exp. Bot.* **53**, 917–925 (2002).
33. Kissen, R. *et al*. Transcriptional profiling of an *Fd-GOGAT1/GLU1* mutant in *Arabidopsis thaliana* reveals a multiple stress response and extensive reprogramming of the transcriptome. *BMC Genomics* **11**, 190 (2010).
34. Mateo-Bonmatí, E., Casanova-Sáez, R., Candela, H. & Micol, J. L. Rapid identification of *angulata* leaf mutations using next-generation sequencing. *Planta* **240**, 1113–1122 (2014).
35. Ponce, M. R., Robles, P., Lozano, F. M., Brotons, M. A. & Micol, J. L. Low-resolution mapping of untagged mutations. *Methods Mol. Biol.* **323**, 105–113 (2006).
36. Robles, P. & Micol, J. L. Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. *Mol. Genet. Genomics* **266**, 12–19 (2001).
37. Lamesch, P. *et al*. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–1210 (2012).
38. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
39. Coschigano, K. T., Melo-Oliveira, R., Lim, J. & Coruzzi, G. M. Arabidopsis *gls* mutants and distinct *Fd-GOGAT* genes. Implications for photorespiration and primary nitrogen assimilation. *Plant Cell* **10**, 741–752 (1998).
40. Somerville, C. R. & Ogren, W. L., *Isolation of photorespiration mutants in Arabidopsis thaliana*. (Elsevier Biomedical Press, 1982).
41. Somerville, C. R. & Ogren, W. L. Inhibition of photosynthesis in Arabidopsis mutants lacking leaf glutamate synthase activity. *Nature* **286**, 257–259 (1980).
42. Suzuki, A. & Knaff, D. B. Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism. *Photosynth. Res.* **83**, 191–217 (2005).
43. Finn, R. D. *et al*. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).
44. van den Heuvel, R. H., Curti, B., Vanoni, M. A. & Mattevi, A. Glutamate synthase: a fascinating pathway from L-glutamine to L-glutamate. *Cell. Mol. Life Sci.* **61**, 669–681 (2004).
45. Winter, D. *et al*. An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**, e718 (2007).
46. Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D. & Penin, A. A. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* **88**, 1058–1070 (2016).
47. Potel, F. *et al*. Assimilation of excess ammonium into amino acids and nitrogen translocation in *Arabidopsis thaliana*-roles of glutamate synthases and carbamoylphosphate synthetase in leaves. *FEBS J.* **276**, 4061–4076 (2009).
48. Ziegler, C. *et al*. Regulation of promoter activity of ferredoxin-dependent glutamate synthase. *Plant Physiol. Biochem.* **41**, 649–655 (2003).
49. Kim, D. *et al*. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
50. Trapnell, C. *et al*. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
51. Trapnell, C. *et al*. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
52. Dzuibany, C. *et al*. Regulation of nitrate reductase transcript levels by glutamine accumulating in the leaves of a ferredoxin-dependent glutamate synthase-deficient *gluS* mutant of *Arabidopsis thaliana*, and by glutamine provided via the roots. *Planta* **206**, 515–522 (1998).
53. Lam, H. M., Hsieh, M. H. & Coruzzi, G. Reciprocal regulation of distinct asparagine synthetase genes by light and metabolites in *Arabidopsis thaliana*. *Plant J.* **16**, 345–353 (1998).
54. Lam, H. M., Peng, S. S. & Coruzzi, G. M. Metabolic regulation of the gene encoding glutamine-dependent asparagine synthetase in *Arabidopsis thaliana*. *Plant Physiol.* **106**, 1347–1357 (1994).
55. Zhang, Q. *et al*. Characterization of Arabidopsis serine:glyoxylate aminotransferase, AGT1, as an asparagine aminotransferase. *Phytochemistry* **85**, 30–35 (2013).
56. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
57. Aoki-Kinoshita, K. F. & Kanehisa, M. Gene annotation and pathway mapping in KEGG. *Methods Mol. Biol.* **396**, 71–91 (2007).

58. Li, C., Potuschak, T., Colon-Carmona, A., Gutierrez, R. A. & Doerner, P. *Arabidopsis* TCP20 links regulation of growth and cell division control pathways. *Proc. Natl. Acad. Sci. USA* **102**, 12978–12983 (2005).
59. Abbasi, N. *et al*. APUM23, a nucleolar Puf domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in Arabidopsis. *Plant J.* **64**, 960–976 (2010).
60. Ponce, M. R., Quesada, V. & Micol, J. L. Rapid discrimination of sequences flanking and within T-DNA insertions in the *Arabidopsis* genome. *Plant J.* **14**, 497–501 (1998).
61. Muñoz-Nortes, T., Pérez-Pérez, J. M., Ponce, M. R., Candela, H. & Micol, J. L. The ANGULATA7 gene encodes a DnaJ-like zinc-finger-domain protein involved in chloroplast function and leaf development in Arabidopsis. *Plant J*. (2016).
62. Casanova-Sáez, R., Mateo-Bonmatí, E., Kangasjärvi, S., Candela, H. & Micol, J. L. Arabidopsis ANGULATA10 is required for thylakoid biogenesis and mesophyll development. *J. Exp. Bot.* **65**, 2391–2404 (2014).
63. Milne, I. *et al*. Tablet-next generation sequence assembly visualization. *Bioinformatics* **26**, 401–402 (2009).
64. Mandel, T. *et al*. Differential regulation of meristem size, morphology and organization by the ERECTA, CLAVATA and class III HD-ZIP pathways. *Development* **143**, 1612–1622 (2016).
65. Du, Z., Zhou, X., Ling, Y., Zhang, Z. & Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–70 (2010).
66. Berardini, T. Z. *et al*. Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiol.* **135**, 745–755 (2004).
67. Trapnell, C. *et al*. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
68. Curtis, M. D. & Grossniklaus, U. A Gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469 (2003).
69. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).

Acknowledgements

We thank J.M. Serrano, F.M. Lozano, A. Torregrosa and J.M. Sánchez-Larrosa for their technical assistance, and Prof. Ben Scheres for the pGEM-T Easy221 plasmid. Research in the laboratory of J.L.M. was supported by grants from the Ministerio de Economía y Competitividad of Spain (BIO2014-53063-P) and the Generalitat Valenciana (PROMETEOII/2014/006). H.C. was a recipient of a Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073). T.M.-N. held a predoctoral fellowship from the Generalitat Valenciana (ACIF/2013/273).

Author Contributions

Resources and Funding Acquisition, J.L.M.; Conceptualization, H.C., J.M.P.-P., and J.L.M.; Supervision, H.C., J.M.P.-P., and J.L.M.; Methodology, H.C., J.M.P.-P., and J.L.M.; Investigation, T.M.-N., R.S.-M., J.M.P.-P., and H.C.; Writing – Original Draft, T.M.-N., H.C., and J.L.M.; Writing, Review & Editing, all authors.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-06335-4

Competing Interests: The authors have no competing financial interests as defined by Nature Publishing Group, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

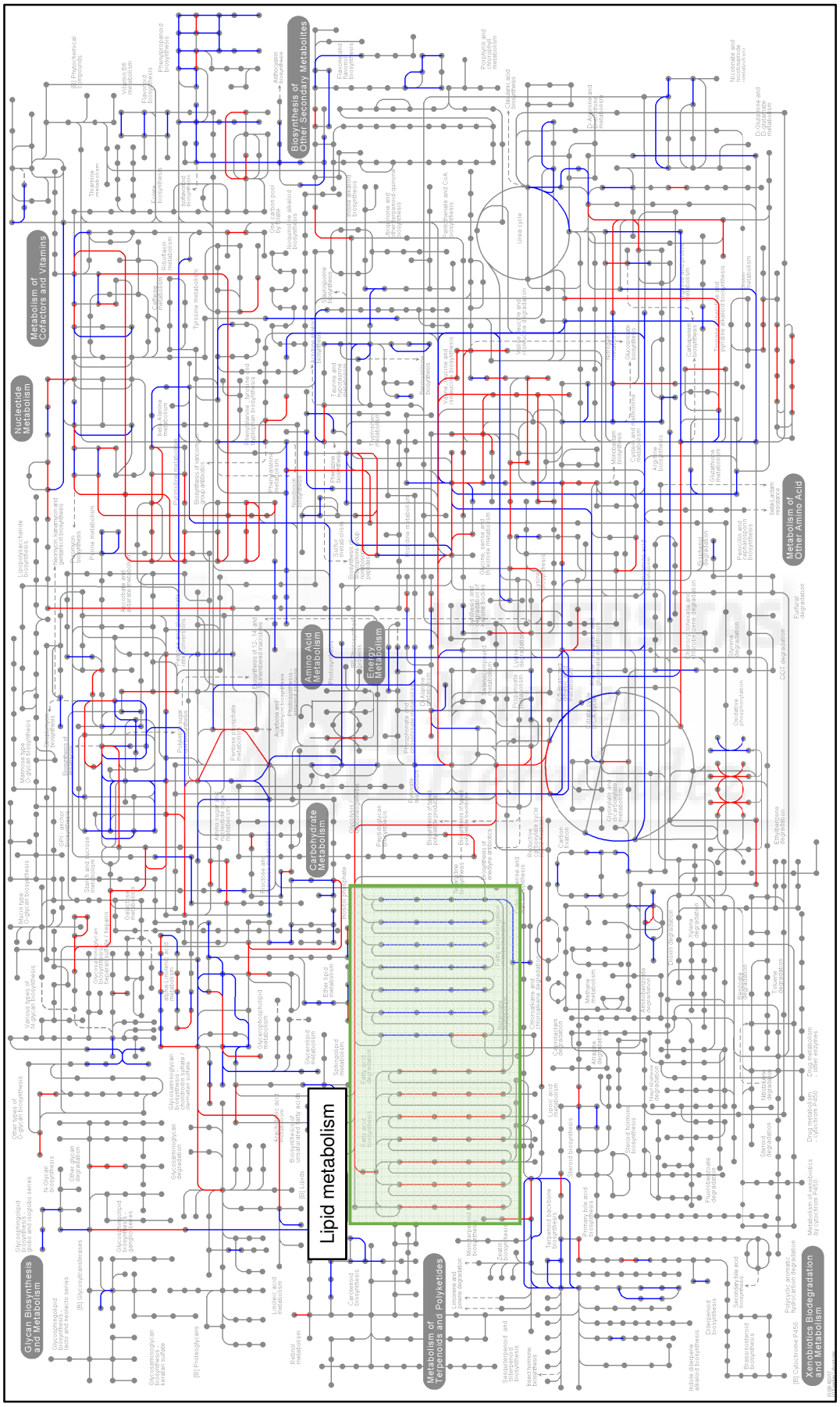
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Deficient glutamate biosynthesis triggers a concerted upregulation of ribosomal protein genes

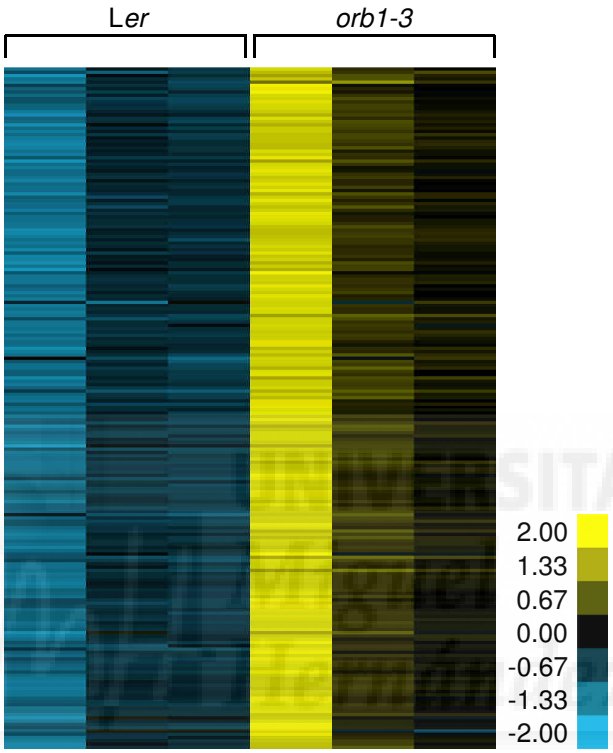
Tamara Muñoz-Nortes, José Manuel Pérez-Pérez, Raquel Sarmiento-Mañús, Héctor Candela, and José Luis Micol

Supplementary Information

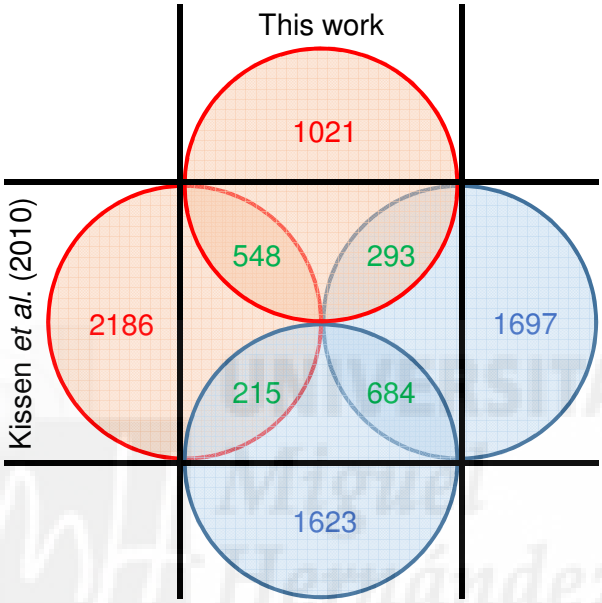
(Figures S1-S3, and Table S1)



Supplementary Figure S1. Schematic view of metabolic pathways affected in the *orb1-3* mutant. The map was generated using the KEGG PATHWAY online tool [http://www.genome.jp/kegg/pathway.html; Kanehisa, M., and Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27-30 (2000)]. Nodes represent metabolites and edges connecting adjacent nodes represent metabolic reactions. Grey nodes and edges represent reactions whose genes are not differentially expressed in our RNA-seq analysis. Upregulated and downregulated genes in *orb1-3* are shown in red and blue, respectively. The green rectangle highlights the fatty acid biosynthesis and degradation pathways.



Supplementary Figure S2. Heat map showing the expression levels of genes encoding components of the cytosolic ribosome in L*er* and *orb1-3* plants. Each column represents the values of fragments per kilobase of transcript per million fragments mapped (FPKM) obtained from each RNA sample in an RNA-seq experiment. The expression levels were normalized based on the mean and standard deviation of the FPKM values obtained for each gene in the six samples analyzed (three L*er* and three *orb1-3*). The heat map was generated performing a hierarchical clustering of arrays with Cluster 3.0, and visualized using Java Treeview 3.0.



Supplementary Figure S3. Venn diagram showing similarities and differences between the microarray analysis performed by Kissen *et al.* (2010), and our RNA-seq analysis. Upregulated and downregulated genes are shown in red and blue, respectively. Common genes between both studies are shown in green.

Supplementary Table S1. Primers used in this work

Purpose	Oligonucleotide name(s)	Oligonucleotide sequences (5'→3')	
		Forward primer (F)	Reverse primer (R)
Linkage analysis	cer478421_F/R	CTCCAATTGCAGCCGAAACC	TGTAGTCCGAGAATCTAAAAATG
	AthCTR1_F/R	TATCAACAGAAACGCACCGAG	CCACTTGTTTCTCTCTCTAG
	cer455551_F/R	CATGCAAATTAGTTTACTGATTGCT	AAGCACACCTACAAGTCCGATTGA
	cer479319_F/R	TCAGAAGTCATTTGCAACTTGTAGA	CACAAATGATATATGGCTGTCGT
	cer457348_F/R	CTCACCACCCATCAGTTCATCT	CGGACGCAAATAAGTTTAATCGAGT
	nga225_F/R	GAAATCCAAATCCCAGAGAGG	TCTCCCCACTAGTTTTGTGTCC
	nga249_F/R	TACCGTCAATTTCATCGCC	GGATCCCTAACTGTAAAATCCC
Sequencing of candidate mutations	At5g04140_F1/R1	AGAAGCTCTCATCACTCATCTG	CTCAAATATGTTTTCAATAACTACA
	At5g04140_F2/R2	CTTGCTCCTTTTGATAAGTTG	CAGATGTTGTTACCTCAGGATAT
	At5g04140_F3/R3	TCATGATTAGAAGTGGAAGAAC	CTCATCAACAAACCTGGCCAT
	At5g04140_F4/R4	TCGGAGTTGTACCAGTTGATG	GACATTTTACCAGTCTATCGG
	At5g04140_F5/R5	CAAAGGTTCTGTCCACATATTTC	CCCACCTCATCAAAGGTGAGT
	At5g04140_F6/R6	TTATTGTGGTGCTCAGATATTTG	GATTGATCACCTGATGTAGATC
	At5g04140_F7/R7	AAGTTGCACAAGGTGCCAAGC	GCTGACCTCTTCTGCTACGTA
	At5g04140_F8/R8	CAGTGGGTGTAGCAAGTCAG	GCTAGTTTTACCTTTCCAACGT
	At5g04140_F9/R9	TGCAGGACAAGTGAACCTAAC	GCCACAACCTGCTCTTGAATG
	At5g04140_R10		GAATGTTGTTAGTTTATGTTAGAC
Gateway cloning	ORB1pro_F/R	GGGGACAAGTTTGTACAAAAAAGCAGGCT TCACCTTCACTGTGTAAAAGA	GGGGACCACTTTGTACAAGAAAGCTGGGT ATTGCATCGCCATTGAAATGA
Genotyping of bacterial clones and transgenic plants	pGEMT221_F/R	GTTGTAAAACGACGGCCAGTG	GGAAACAGCTATGACCATGATT
	pMDC163_F/R	CTCTAGCATTCGCCATTCAGG	ATTGCCCGGCTTTCTTGTAAC

UNIVERSITAS
Miguel
Hernández

Suitability of two distinct approaches for the high-throughput study of the post-embryonic effects of embryo-lethal mutations in Arabidopsis

Tamara Muñoz-Nortes, Héctor Candela, and José Luis Micol

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Spain

Corresponding author: J.L. Micol (telephone: 34 96 665 85 04; fax: 34 96 665 85 11; E-mail: jlmicol@umh.es)

Running head: Clonal analysis of embryo-lethal mutants.

Keywords: embryonic-lethal mutants, clonal analysis, *Arabidopsis thaliana*

Character count: 24865. Word count: 4293.

Figures: 7 Tables: 2 Supplemental Figures: 0 Supplemental Tables: 0

Abstract

Several hundred genes are required for embryonic and gametophytic development in Arabidopsis, as inferred from the lethality of their mutations. Despite many of these genes are normally expressed throughout the plant life cycle, the corresponding mutants usually die at early stages, preventing the study of their post-embryonic functions using conventional methods. Clonal analysis techniques can provide an effective solution to this problem by uncovering the effects of embryo-lethal mutations in sectors of mutant cells within an otherwise normal adult plant. We selected 24 embryo-lethal mutations (*emb*) for clonal analysis, with a focus on their effects on leaf development. For the induction of mutant sectors in adult plants, we used two approaches: one based on the X-ray irradiation of so-called 'cell autonomy' lines (CAUT), and another based on the site-specific excision of transgenes mediated by Cre recombinase. This study will advance our knowledge on the functions of selected *EMB* genes, some of which encode transcription factors and components of the epigenetic machinery.

Introduction

Mutational approaches have greatly advanced our understanding of developmental processes in plants and animals. The isolation and characterization of viable mutants with defective growth and pattern formation has been crucial to identify both housekeeping and regulatory genes that are required for the organism to attain its normal size and shape. By focusing on viable mutations, however, these screenings are likely to have missed many genes that play important post-embryonic roles, because they are essential in early developmental stages and there are not viable alleles to study. This is particularly important in plants, whose development takes place mostly post-embryonically, after the basic body plan is laid out during the embryogenesis. Post-embryonic development includes the development of important plant organs, such as the leaves. Indeed, numerous viable mutants identified in such screenings turned out to be hypomorphic (partial loss-of-function) alleles of genes otherwise known only by their embryonic lethal effects. Some examples are the *angulata1-1* (*anu1-1*), *anu7-1*, *anu9-1* and *scabra1-1* (*sca1-1*) mutants of *Arabidopsis thaliana* (hereafter, Arabidopsis), identified in a large-scale screen for viable mutants with abnormal leaf shape, size and pigmentation, which were later found to be hypomorphic alleles of the *SECA2*, *EMBRYO DEFECTIVE 2737* (*EMB2737*), *NON-INTRINSIC ABC PROTEIN 14* (*NAP14*) and *EMB3113* genes[1-3]. Another example is the *incurvata2-1* (*icu2-1*) mutant, identified in the same screen and found to be the first viable allele of the *ICU2* gene, which encodes the catalytic subunit of DNA polymerase α[4]. Because a significant fraction of the genes in the Arabidopsis genome is known to correspond to essential functions, and many such genes are expressed beyond the embryogenesis in wild-type plants, we hypothesized that many of them might also perform important roles in adult plants, after the embryogenesis has been completed.

Clonal analysis has been used to study embryo-lethal mutations by inducing genetic mosaics in many organisms, such as *Drosophila melanogaster*, maize and Arabidopsis[5-12]. Clonal analysis experiments typically combine a lethal gene and a cell-autonomous reporter gene or mutation with an easy-to-score phenotype, in order to identify induced mutant sectors that exhibit a post-embryonic mutant phenotype. By inducing mutant sectors in phenotypically wild-type plants, clonal analysis has helped researchers to answer questions regarding the phenotypic effects caused by the complete inactivation of embryo-lethal (*EMB*) genes in the tissues of an adult plant, the site of action of gene products, the cell autonomy and the cell lethality of lethal mutations. Several experimental approaches are available to perform clonal analysis in plants, including methods based on the loss of a chromosome arm after irradiation[6], the mobilization of transposons[5], or the use of transgenic approaches (e.g. based on the

induction of site-specific recombinases to induce heritable changes in a given linage of cells[7,8,10,12,13]. In this work, we performed a pilot experiment aimed at determining which strategy is best suited for the high-throughput identification of the post-embryonic effects of a set of embryonic-lethal mutations. We tested two different strategies, one involving the use of X-rays and CAUT (cell autonomy) lines, and another based on the site-specific excision of transgenes mediated by Cre recombinase.

Results and Discussion

In an attempt to select an efficient strategy that is suitable for the systematic identification of essential genes that also function post-embryonically, we have carried out pilot experiments using two different approaches aimed at inducing somatic sectors that express the mutant phenotype, one based on the use of CAUT lines[14], and another based on the use of the Cre-*loxP* site-specific recombination system[7]. We focused on a subset of 24 *EMBRYO DEFECTIVE* (*EMB*) genes selected from the SeedGenes database (http://www.seedgenes.org/), which includes comprehensive information on the embryonic lethal genes of Arabidopsis[15]. *EMB* genes were selected based on the availability of embryo-lethal mutant alleles and on their expression patterns beyond the embryogenesis (Table 1), particularly focusing on genes that are normally expressed during the vegetative phase (in leaves and basal rosettes) according to publicly available data from the electronic Fluorescent Pictograph (eFP) browser database[16,17]. The genes selected encode proteins as diverse as transcription factors, proteasome subunits or epigenetic factors, which were considered good candidates to control leaf development at the transcriptional or post-transcriptional levels. We also selected some genes encoding proteins containing conserved domains whose functions remain unknown.

Sector induction using CAUT lines and X-rays

For the induction of marked somatic sectors in Arabidopsis, we initially took advantage of the availability of CAUT lines with insertions located on every chromosome arm[14]. For 13 different *EMB* genes (Table 1), we selected CAUT lines carrying an insertion of the *CHLORATA-42* (*CH-42*) gene located between the *EMB* gene and the centromere of the corresponding chromosome; this gene encodes the CHLI subunit of magnesium chelatase, which is required for chlorophyll biosynthesis. By choosing this configuration, we expect that all marked (yellow) sectors found after X-ray irradiation have also lost the wild-type allele of the *EMB* gene. To implement this strategy (Figure 1), we systematically crossed heterozygous *EMB*/*emb* plants to the homozygous *ch-42*/*ch-42* mutant and isolated F_2 plants displaying the recessive yellow phenotype caused by *ch-42*. The presence of aborted embryos or collapsed seeds in the siliques of these plants allowed us to select *ch-42*/*ch-42* plants segregating the corresponding *emb* mutation in the F_3 progeny (Figure 2a, b). Plants with the *EMB*/*emb*; *ch-42/ch-42* genotype were subsequently crossed to appropriate CAUT lines. Ten different CAUT lines were used for this purpose (Table 1). Whenever possible, we selected CAUT lines carrying the *CH-42* insertion that maps closest to the *EMB* gene, because a higher frequency of chromosomal breaks is expected to occur as the distance between the insertion and the centromere increases. This crossing scheme allowed us to select phenotypically wild-

type (green) plants that carry an insertion of the *CH-42* transgene in the F_2 generation. F_3 families segregating individual *emb* mutations were then established from F_2 plants that had aborted embryos in their siliques. Sibling families not segregating the *emb* mutations were also established from each cross as a control. We tested the Mendelian segregation of the yellow *ch-42* phenotype in these F_3 families. Unexpectedly, we found a high number of plants exhibiting a yellow phenotype in seven (out of the thirteen) families segregating aborted seeds, suggesting that the *CH-42* transgene fails to complement the *ch-42* allele (possibly due to silencing) or that it is located at a higher-than-expected chromosomal distance from the corresponding *EMB* gene.

In phenotypically wild-type *ch-42*/*ch-42*; *EMB CH-42*/*emb* – plants, X-rays can cause chromosomal breaks between the centromere and the T-DNA insertion, and are expected to generate hemizygous yellow sectors when the acentric fragment carrying the extra copy of *CH-42* and the *EMB* wild-type allele is lost. A drawback of irradiating F_3 families, which comprise seeds with a mixture of genotypes, is that recombination events between the loci of the T-DNA insertions and the linked *EMB* genes might lead to yellow sectors that still keep a functional copy of the *EMB* gene. Any developmental or other visible phenotypes occurring specifically in the yellow sectors can be attributed to the post-embryonic effects of the corresponding *emb* mutation only if they are not observed in the irradiated control families.

Two different X-ray dosages were used to induce sectors. On the one hand, water-imbibed seeds were subjected to a dosage of 1000 rad (10 Gy) based on previous reports from the Arabidopsis and maize literature[6]. On the other, dry seeds received a dosage of 16000 rad (160 Gy), as previously described[14]. The irradiation of dry seeds allowed us to stagger the sowing of the irradiated families. Plants were periodically examined under the stereomicroscope to identify yellow sectors. The temperature sensitivity of the *ch-42* mutation, which determined a paler pigmentation at 26°C than at 20°C, made the yellow sectors easier to spot and helped us to select plants with the correct genotype (Figure 2c, d). Sectors occurred at a very low frequency in the families irradiated at 1000 rad (Figure 3a-c). In these families, we only found 6 sectors, one half of which appeared in control families lacking an *emb* mutation (Figure 3c). Three of these sectors, including two in the control families, were completely albino, rather than yellow, suggesting that rearrangements caused by X-rays lead to visible phenotypes even when *emb* mutations are not involved. By contrast, we found sectors in every family in about 1% of the plants irradiated at 16000 rad (Figure 3d-f), a frequency that is roughly similar to the frequency reported by Furner *et al.* (2008). In the six families that exhibited a clear distortion of the Mendelian segregation of the yellow phenotype caused by *ch-42*, we found somatic sectors in both types of irradiated families (segregating and not

segregating the *emb* mutation; Figure 3f), making it difficult to draw conclusions on the post-embryonic roles of the corresponding genes.

Incidentally, this approach occasionally allowed us to find escapers for some *emb* mutations, i.e. plants that completed the embryogenesis and reached the seedling stage or beyond, potentially providing information on the post-embryonic function of the genes. Escapers were found for mutant alleles of three *EMB* genes (Figure 4), in all cases at a very low frequency in the F_2 generation (0.72% for *emb1135*, 1.92% for *emb1706-1*, and 0.48% for *emb2410-1*). The majority of escapers were pale green, as expected from our crossing scheme, and exhibited additional developmental phenotypes. The *emb1135* escapers were small, with fused cotyledons, wrinkled surface and irregular margins (Figure 4a). The *emb2410* escapers expanded their cotyledons and then died (Figure 4e, f). The *emb1706* escapers formed small rosettes, which included leaves with long petioles and adaxially curved leaf laminae (Figure 4b). When transferred to soil, these plants produced numerous secondary shoots (Figure 4c) with abnormally patterned flowers (Figure 4d).

Sector induction using Cre recombinase

We also tested a strategy based on the site-specific excision of transgenes driven by a heat-inducible Cre recombinase (Figure 5). To this end, we prepared two Gateway-compatible versions of the pCB1 vector (see Material and Methods), which is intended for the induction of clonal sectors by means of the Cre-mediated excision of a cassette containing a wild-type copy of the gene of interest (Figure 6). We used the Gateway cloning technology to systematically create 20 entry clones, each containing a different genomic region able to complement the embryonic lethality of a selected *emb* mutation (Table 1). These entry clones were transferred to the Gateway-compatible version of pCB1 using LR reactions in order to obtain constructs for plant transformation. Because the Gateway cassette is flanked by two *loxP* sites, the genomic inserts of these constructs can be excised by expressing Cre to produce GFP-marked, *emb*/*emb* mutant sectors.

In order to obtain transgenic lines for 20 non-allelic *emb* mutations, we first transformed homozygous HS_{pro}*:Cre* plants with the pCB1-Gateway constructs, each carrying a wild-type copy of a different *EMB* gene. The resulting T_1 transformants are expected to carry insertions of two T-DNAs, one from the pCB1-Gateway vector and another to allow the inducible expression of Cre driven by a heat shock promoter. These transgenic plants were subsequently crossed to *EMB*/*emb* heterozygotes to isolate plants carrying the *emb* mutation and both constructs. The F_1 and F_2 progenies of these crosses were genotyped by PCR to verify the presence of both constructs before sector

induction. Plates containing 6-days-after-sowing seedlings were sealed with Parafilm and heat-shocked for 30 min at 37°C in a water bath. We reproducibly found leaf sectors for four different *emb* mutations: *emb1408-1*, *emb1586-1*, *emb1637-1* and *emb2001-1* (Figure 7). However, sectors similar to those for *emb1637-1* and *emb2001-1* occurred in the corresponding control lines (i.e. with the pCB1-Gateway and HS_{pro}*:Cre* constructs but without an *emb* mutation; Figure 7g, h), showing that sectors with a mutant phenotype can arise from Cre-induced chromosomal rearrangements even in the absence of an embryonic-lethal mutation. In addition to these, we found some heat-shocked families segregating plants with impaired growth and a pale green phenotype (Figure 7e). These plants exhibited intense GFP fluorescence (Figure 7f), suggesting that the phenotype might arise from the occurrence of Cre-mediated recombination generally in the affected tissues.

Concluding remarks

In this report, we have tested two different strategies for the induction of somatic sectors in adult plants. The first approach, based on the use of CAUT lines, did not scale up well for high-throughput studies. In addition to being labour-intensive and time-consuming, this strategy required a complex crossing scheme with several generations before plant materials were ready for irradiation. According to Furner *et al*. (2008), the timing required for preparing a single line is about 40 weeks. This approach is further complicated when the *emb* mutations reside in the same chromosome arm as the *ch-42* marker (on chromosome 4) or when they map very close to a centromere. The latter problem might make it difficult to identify an appropriate CAUT line for a given *emb* mutation, and a short distance between the *CH-42* transgene and the centromere is expected to result in a low frequency of sectors. Furthermore, scoring the boundaries of yellow sectors is a problematic task, particularly when the sectors are small or hard to distinguish from other pale-green necrotic sectors that occur non-specifically (i.e. which might also be present in control families) as a secondary effect of the X-ray treatment.

Implementing the second strategy, based on the use of the site-specific Cre recombinase and transgenes, was more straightforward. To establish an efficient cloning pipeline, we first prepared a Gateway destination vector based on the pCB1 vector, which has previously been used effectively to characterize the effects of individual embryonic-lethal mutations in somatic sectors in *Arabidopsis thaliana*[7]. We found that a skilled operator can efficiently streamline the making of entry clones containing large genomic inserts by using high-fidelity DNA polymerases (e.g. Phusion High-Fidelity DNA Polymerase) and primers containing *att*B1 and *att*B2 sites for subsequent recombination into the Gateway-compatible version of pCB1.

Alternatively, additional information might be obtained from the characterization of hypomorphic (non-null) alleles of *EMB* genes, which might be difficult to isolate. As an example, mutations in *EMB2107* and *EMB1611*[18,19] have recently been found to cause post-embryonic phenotypes in leaves. These new alleles will be ideal controls in future clonal analysis experiments with a larger number of plants or aimed at defining an optimal set of experimental conditions. Fine-tuning the X-ray dosages or the duration of the heat-shock treatment should help to minimize the secondary effects of both treatments while optimizing the frequency of somatic sectors specifically being due to loss of *EMB* functions.

Methods

Plant materials, growth conditions and crosses

Seeds of the *Arabidopsis thaliana* L. Heynh. wild-type accessions Landsberg *erecta* (L*er*) and Columbia-0 (Col-0), as well as heterozygous *EMB*/*emb* lines and CAUT lines (Table 1) were obtained from the Nottingham Arabidopsis Stock Centre (NASC; http://arabidopsis.info/). Transgenic seeds carrying the HS_{pro}*:CRE* construct were kindly supplied by Dr. Guy Wachsmann. Seed sterilization, sowing, plant culture and crosses were performed as previously described[20,21]. Briefly, seeds were sown on plates containing Murashige and Skoog (MS) agar medium [half-strength MS salts, 0.7% plant agar (Duchefa), pH 5.7, and 1% sucrose], stratified at 4°C in the dark for 24 h and then transferred to TC16 or TC30 growth chambers (Conviron) set to our standard conditions (continuous light at approximately 75 $\mu mol \cdot m^{-2} \cdot s^{-1}$, 20°C, 60-70% relative humidity). When required, plants were transferred to pots containing a 2:2:1 mixture of perlite:vermiculite:sphagnum moss and grown in walk-in growth chambers set to our standard conditions. For selection of transgenic plants, T_1 seeds were sown in flat pots containing perlite and river sand and were sub-irrigated with ATM supplemented with 15 mg/l glufosinate ammonium (Finale).

Irradiation and sector screening

Irradiation of Arabidopsis seeds was performed using a Philips MG102 X-ray cabin. Seeds were irradiated at doses of 10 Gy for sterilized seeds and 160 Gy for dry seeds, as previously described[6,14]. At least two control wild-type lines and two heterozygous *EMB*/*emb* lines of each of the 13 genotypes were irradiated. After irradiation, seeds were sown in Petri dishes and the resulting plants were checked periodically, looking for mutant sectors. Pictures of the different sectors were taken, and the leaves that contained them were collected and stored. Plants containing sectors were moved to soil pots in order to verify if they spread to other plant organs like secondary shoots, cauline leaves or flowers, as previously described[14].

Modification of pCB1 vector

We modified the pCB1 vector[7] for use with the Gateway cloning technology (Figure 6). For this, pCB1 was linearized with *Not*I, and the resulting cohesive ends were filled in with Klenow to generate blunt ends. A PCR product corresponding to a Gateway cassette (Frame A) was amplified with Phusion DNA polymerase (Finnzymes) and ligated to pCB1 using T4 ligase (Fermentas). The ligation products were transformed into the *Escherichia coli* DB3.1 strain, and colonies resistant to both kanamycin and chloramphenicol were selected. This modified plasmid was called pCB1-Gateway. After

purifying the plasmids that carried the insert of interest, its orientation was checked with a *Sma*I and *Sal*I double digestion. We obtained two different versions of the pCB1-Gateway vector, with the Gateway cassette oriented in both possible directions, (+) and (-).

Generation of pCB1-Gateway constructs

In order to introduce a wild-type copy of the *EMB* genes of interest into the pCB1-Gateway empty vector, we amplified genomic regions containing each *EMB* gene spanning from the end of the previous gene coding region to the beginning of the following gene coding region, to make sure that the regulatory sequences were also included. We designed primer pairs containing *attB1* and *attB2* sites (Table 2), in order to amplify the regions that contain each *EMB* gene of interest from its corresponding bacterial clone. These regions were PCR amplified using the Phusion polymerase (Finnzymes). The amplification products were purified and used in different BP reactions (Invitrogen), in which the pGEM-T Easy 221 plasmid was used as entry vector. Chemocompetent DH5*α Escherichia coli* cells were transformed by heat shock with the products of BP reactions. Colonies carrying the pGEM-T Easy 221 plasmid were selected in Petri dishes with LB medium supplemented with ampicillin. Insert presence was checked by rapid size screen with lysis buffer[22], digestion with the restriction enzyme *Not*I and PCR with plasmid and insert primers (Table 2). Positive colonies were used to perform LR reactions (Invitrogen) with the appropriate pCB1-Gateway destination vector. Each LR reaction was performed twice, using the pCB1-Gateway plasmids with the Gateway cassette in both orientations. Chemocompetent DH5*α Escherichia coli* cells were transformed by heat shock with the LR products and colonies carrying the pCB1-Gateway vector were selected in LB medium supplemented with kanamycin. The presence of each insert was checked by double digestion with *Xba*I and *Sma*I restriction enzymes. Positive clones were mobilized into *Agrobacterium tumefaciens* C58C1 pSOUP cells by electroporation. Every pCB1-Gateway construct was transferred to plants carrying the HS_{pro}*:CRE* construct by the floral dip method[23].

Heat shock sector induction

Plants carrying HS_{pro}*:CRE* and pCB1-Gateway constructs combined with *emb* mutations were sowed in Petri dishes. After growing for 6 days, plates were sealed with Parafilm and submerged in water at 37°C during 4 hours. They were put inside the plant growth chamber again and, after 5-6 days, the different lines were observed using fluorescence microscopy, in order to detect sectors with GFP signal.

Acknowledgements

We thank J.M. Serrano, J.M. Sánchez-Larrosa and A. Torregrosa for their excellent technical assistance. We thank Dr. Renzé Heidstra for sending the pCB1 vector, Dr. Guy Wachsmann for *HS$_{pro}$:CRE* seeds, and Dr. Ginés Morata for the use of facilities for irradiation with X-rays. Research in the laboratory of J.L.M. was supported by grants from the Ministerio de Economía y Competitividad of Spain (BIO2014-53063-P) and the Generalitat Valenciana (PROMETEO/2014/006). H.C. was a recipient of a Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073). T.M.-N. held a predoctoral fellowship from the Generalitat Valenciana (ACIF/2013/273).

References

1. Mateo-Bonmatí, E., Casanova-Sáez, R., Candela, H. & Micol, J. L. Rapid identification of *angulata* leaf mutations using next-generation sequencing. *Planta* **240**, 1113-1122 (2014).
2. Mateo-Bonmatí, E.*, et al.* Plastid control of abaxial-adaxial patterning. *Sci. Rep.* **5**, 15975 (2015).
3. Muñoz-Nortes, T., Pérez-Pérez, J. M., Ponce, M. R., Candela, H. & Micol, J. L. The *ANGULATA7* gene encodes a DnaJ-like zinc-finger-domain protein involved in chloroplast function and leaf development in Arabidopsis. *Plant J.* (2016).
4. Barrero, J. M., González-Bayon, R., del Pozo, J. C., Ponce, M. R. & Micol, J. L. *INCURVATA2* encodes the catalytic subunit of DNA polymerase alpha and interacts with genes involved in chromatin-mediated cellular memory in *Arabidopsis thaliana*. *Plant Cell* **19**, 2822-2838 (2007).
5. Becraft, P. W., Li, K., Dey, N. & Asuncion-Crabb, Y. The maize *dek1* gene functions in embryonic pattern formation and cell fate specification. *Development* **129**, 5217-5225 (2002).
6. Fu, S. & Scanlon, M. J. Clonal mosaic analysis of EMPTY PERICARP2 reveals nonredundant functions of the duplicated HEAT SHOCK FACTOR BINDING PROTEINs during maize shoot development. *Genetics* **167**, 1381-1394 (2004).
7. Heidstra, R., Welch, D. & Scheres, B. Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis* SCARECROW action in asymmetric cell division. *Genes Dev.* **18**, 1964-1969 (2004).
8. Latvala-Kilby, S. M. & Kilby, N. J. Uncovering the post-embryonic role of embryo essential genes in *Arabidopsis* using the controlled induction of visibly marked genetic mosaics: *EMB506*, an illustration. *Plant Mol. Biol.* **61**, 179-194 (2006).
9. Resino, J., Salama-Cohen, P. & García-Bellido, A. Determining the role of patterned cell proliferation in the shape and size of the *Drosophila* wing. *Proc. Natl. Acad. Sci. U.S.A* **99**, 7502-7507 (2002).
10. Serralbo, O., Pérez-Pérez, J. M., Heidstra, R. & Scheres, B. Non-cell-autonomous rescue of anaphase-promoting complex function revealed by mosaic analysis of *HOBBIT*, an *Arabidopsis CDC27* homolog. *Proc. Natl. Acad. Sci. U.S.A* **103**, 13250-13255 (2006).
11. Stern, C. Somatic crossing over and segregation in *Drosophila melanogaster*. *Genetics* **21**, 625-730 (1936).
12. Wildwater, M.*, et al.* The *RETINOBLASTOMA-RELATED* gene regulates stem cell maintenance in *Arabidopsis* roots. *Cell* **123**, 1337-1349 (2005).

13. Candela, H., Pérez-Pérez, J. M. & Micol, J. L. Uncovering the post-embryonic functions of gametophytic- and embryonic-lethal genes. *Trends Plant Sci.* **16**, 336-345 (2011).
14. Furner, I., Ellis, L., Bakht, S., Mirza, B. & Sheikh, M. CAUT lines: a novel resource for studies of cell autonomy in Arabidopsis. *Plant J.* **53**, 645-660 (2008).
15. Tzafrir, I.*, et al.* The Arabidopsis SeedGenes Project. *Nucleic Acids Res.* **31**, 90-93 (2003).
16. Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D. & Penin, A. A. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* **88**, 1058-1070 (2016).
17. Winter, D.*, et al.* An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLOS ONE* **2**, e718 (2007).
18. Book, A. J.*, et al.* The RPN5 subunit of the 26s proteasome is essential for gametogenesis, sporophyte development, and complex assembly in Arabidopsis. *Plant Cell* **21**, 460-478 (2009).
19. Leasure, C. D., Fiume, E. & Fletcher, J. C. The essential gene *EMB1611* maintains shoot apical meristem function during Arabidopsis development. *Plant J.* **57**, 579-592 (2009).
20. Berná, G., Robles, P. & Micol, J. L. A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. *Genetics* **152**, 729-742 (1999).
21. Ponce, M. R., Quesada, V. & Micol, J. L. Rapid discrimination of sequences flanking and within T-DNA insertions in the *Arabidopsis* genome. *Plant J.* **14**, 497-501 (1998).
22. Law, D. & Crickmore, N. Use of a simplified rapid size screen protocol for the detection of recombinant plasmids. *Tech. Tips Online* **2**, 136-137 (1997).
23. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735-743 (1998).

Muñoz-Nortes *et al.*, Figure 1

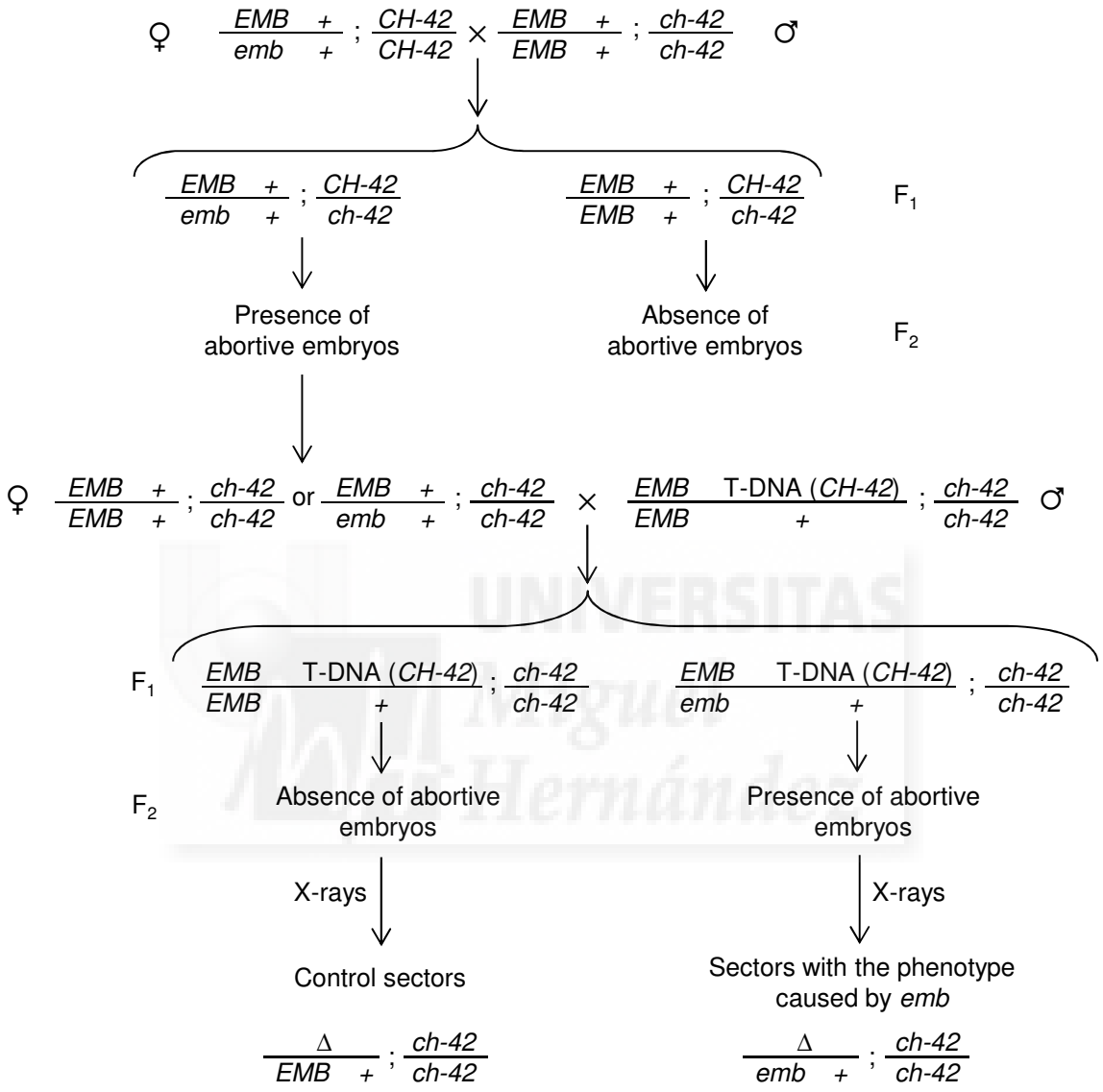


Figure 1. Detailed strategy to obtain hemizygous sectors for an embryo-lethal (*emb*) mutation by means of X-rays. Only the relevant genotype of each member from a pair of homolog chromosomes is indicated. The generation derived from a cross is indicated as F_1, and the progeny of its self-fertilization is indicated as F_2. The uppercase Greek letter delta (Δ) represents the loss of a chromosome fragment. In cells with the appropriate genotype, the loss of a chromosome fragment containing the *CHLORATA-42* (*CH-42*) transgene and the wild-type copy of the *EMB* gene gives rise to a cell with pale-green genotype which might be accompanied by a mutant phenotype caused by the *emb* mutation.

Muñoz-Nortes *et al.*, Figure 2

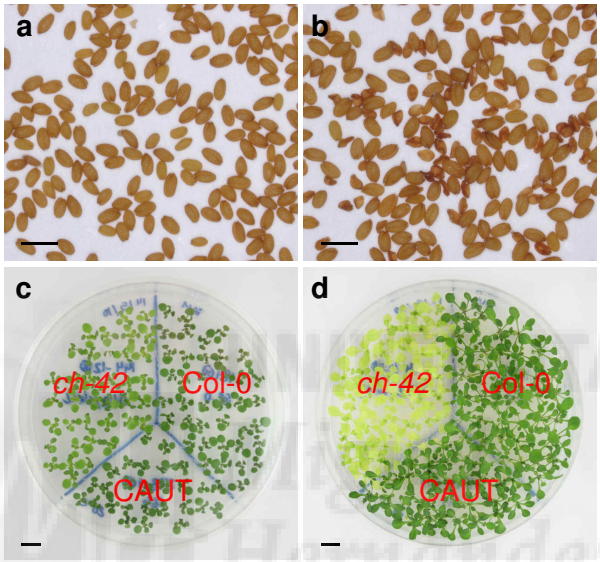


Figure 2. Selection of *EMB*/*emb* lines and effects of temperature on *ch-42* plants. (a, b) F_2 mature seeds derived from a cross involving *EMB*/*emb*;*CH-42*/*CH-42* and *EMB*/*EMB*;*ch-42*/*ch-42* plants. (a) Absence of abortive seeds indicates that the F_2 line does not carry the *emb* mutation, and (b) presence of abortive seeds indicates that the F_2 line carries the *emb* mutation. (c, d) Plants from different genotypes growing at (c) 20ºC, and (d) 26ºC. Scale bars represent (a, b) 1 mm, and (c, d) 1 cm.

Muñoz-Nortes *et al.*, Figure 3

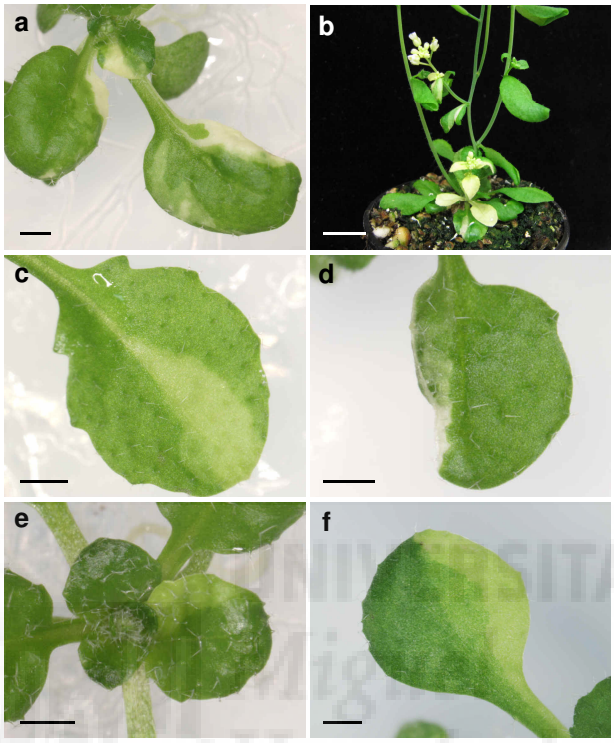


Figure 3. Sectors identified after X-rays irradiation. Plants from irradiated families segregating (a, b) *emb1441*, (d) *emb2001*, and (e) *emb1706* mutations. (c, f) Plants from irradiated families that are not segregating *emb* mutations. Plants were irradiated at dosages of (a-c) 1000 and (d-f) 16000 rad. Plants were collected (a, c-f) 14 and (b) 40 days after stratification. Scale bars represent (a, c-f) 1 mm and (b) 1 cm.

Muñoz-Nortes *et al.*, Figure 4

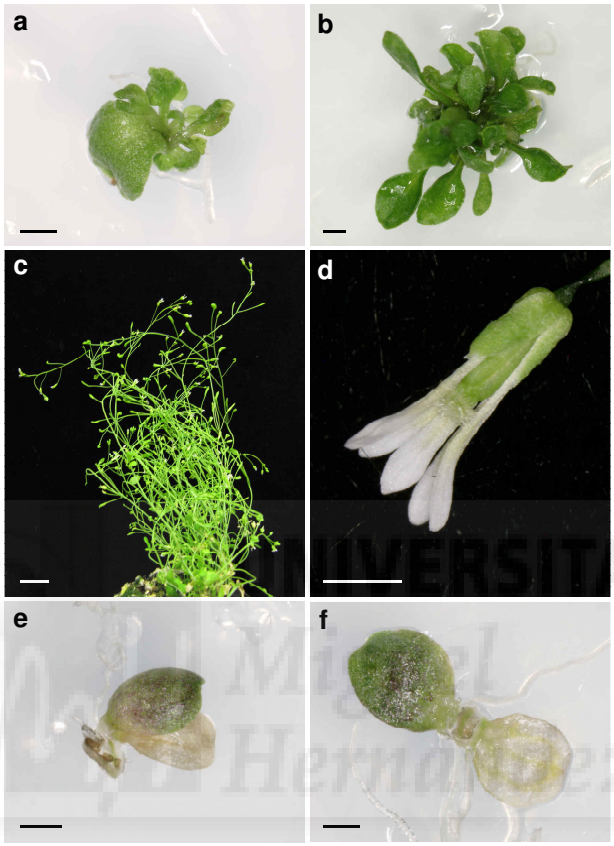


Figure 4. Putative scapers for (a) *emb1135*, (b-d) *emb1706*, and (e, f) *emb2410* mutations. Plants were collected (a, e, f) 21, (b) 40 and (c, d) 50 days after stratification. Scale bars represent (a, b, d-f) 1 mm, and (c) 1 cm.

Muñoz-Nortes *et al.*, Figure 5

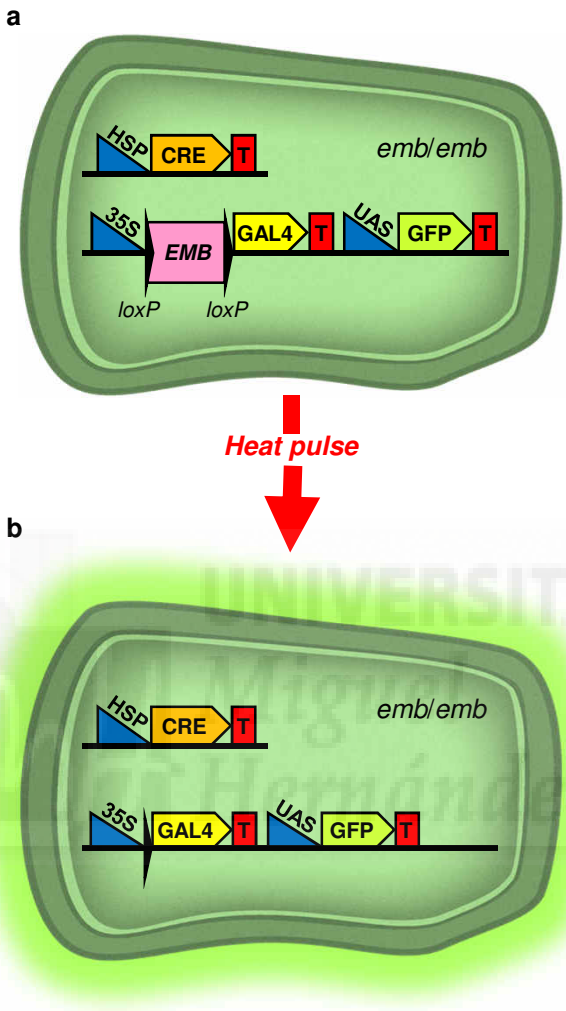


Figure 5. Transgene-mediated approach to generate hemizygous marked sectors for embryo-lethal mutations. (a) Cell with the appropriate genotype for induction of fluorescent sectors by heat shock. This cell is homozygous for the embryo-lethal mutation (*emb*/*emb*) and carries two different constructs, one of them providing a wild-type copy of an *EMB* gene that allows its normal development, and the other with a heat-shock promoter driving the inducible expression of Cre recombinase. (b) A heat pulse causes the activation of Cre and a concomitant loss of the wild-type copy of the *EMB* gene through the excision of the Gateway cassette mediated by the action of Cre recombinase on the *loxP* sites. The subsequent action of GAL4 on the UAS drives the expression of GFP and marks the cell, which is fluorescent and might exhibit any mutant phenotype associated with the loss of function of the *EMB* gene in adult tissues.

Muñoz-Nortes *et al.*, Figure 6

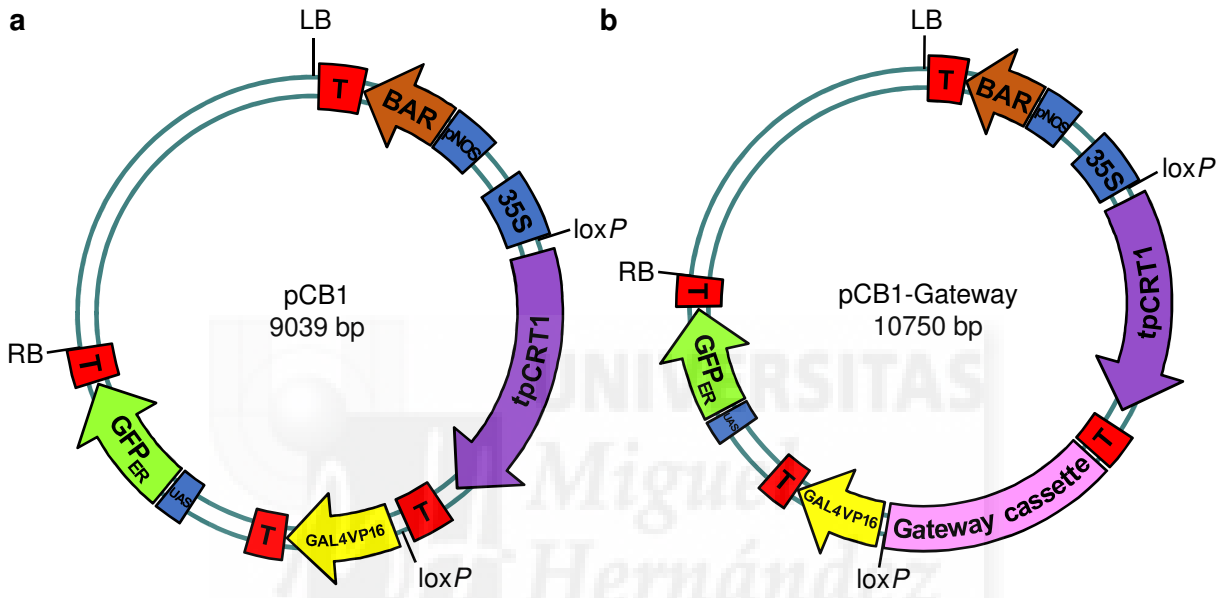


Figure 6. Maps of pCB1 and pCB1-Gateway vectors. (A) The pCB1 binary vector, and (B) the modified pCB1-Gateway vector. LB: T-DNA left border; T: transcriptional terminator; BAR: bialaphos resistance gene; pNOS: nopaline synthase promoter; 35S: constitutive promoter; lox*P*: Cre recombination site; tpCRT1: resistance gene; GAL4VP16: transcriptional activator; UAS: upstream activating sequence; GFP_{ER}: endoplasmic reticulum-localized green fluorescent protein; RB: T-DNA right border.

Muñoz-Nortes *et al.*, Figure 7

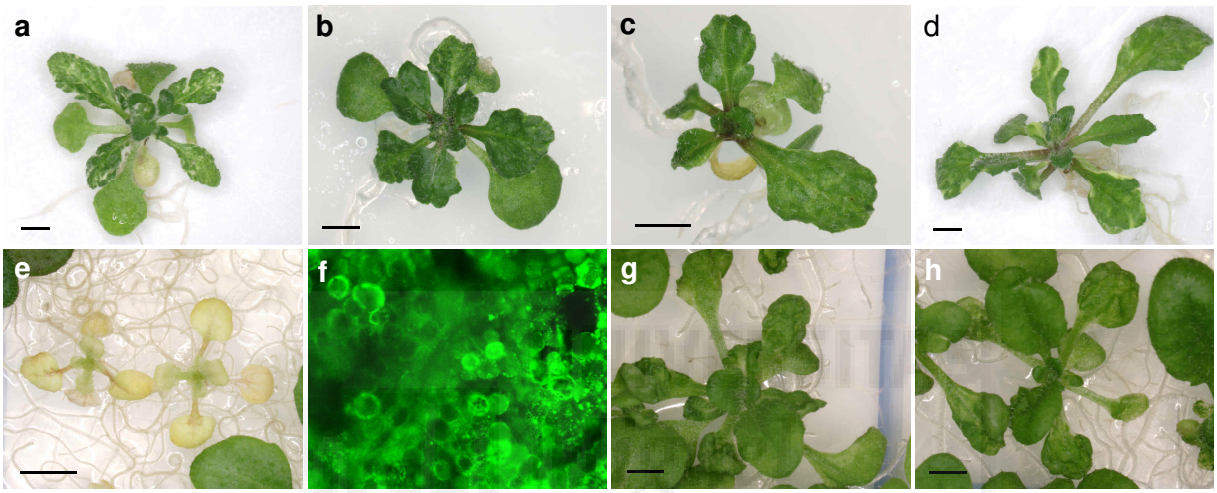


Figure 7. Observed phenotypes after inducing sectors by heat shock. (a-f) Plants carrying (a) *emb1408*, (b) *emb1586*, (c) *emb1637* and (d, e) *emb2001* mutations. (g, h) Control plants carrying the HS_{pro}:Cre and pCB1-Gateway constructs, but not an *emb* mutation. (e) Pale-green plants with impaired growth. (f) Intense GFP fluorescence in one of the plants shown in (e). Scale bars represent (a-e, g,h) 2 mm.

Table 1. *EMB* genes, CAUT lines and pCB1 constructs used in this work

Gene name	AGI code	Chromosome	Coordinates	Mutant allele	CAUT line	pCB1
ATSWI3A	AT2G47620	2	19531947 - 19534401	*atswi3a-1*	7F	Yes
EMB1135	AT1G79350	1	29844633 - 29853414	*emb1135*	C381	-
EMB1381	AT2G31340	2	13361506 - 13365200	*emb1381-1*	-	Yes
EMB1408	AT5G67570	5	26952352 - 26955543	*emb1408*	-	Yes
EMB1441	AT5G49930	5	20308033 - 20312808	*emb1441-1*	L82	Yes
EMB1513	AT2G37920	2	15868580 - 15870071	*emb1513-1*	-	Yes
EMB1586	AT1G12770	1	4351064 - 4353685	*emb1586-1*	-	Yes
EMB1611	AT2G34780	2	14668653 - 14673904	*emb1611*	L40	Yes
EMB1637	AT3G57870	3	21428496 - 21430200	*emb1637*	25_12	Yes
EMB1674	AT1G58210	1	21553621 - 21558056	*emb1674-1*	-	Yes
EMB1688	AT1G67440	1	25263804 - 25265719	*emb1688-1*	-	Yes
EMB1691	AT4G09980	4	6247735 - 6252288	*emb1691-1*	L104	Yes
EMB1706	AT4G10760	4	6619817 - 6623351	*emb1706-1*	L104	Yes
EMB1745	AT1G13120	1	4469181 - 4473213	*emb1745*	-	Yes
EMB1895	AT4G20060	4	10854790 - 10859330	*emb1895-1*	-	Yes
EMB1923	AT4G28210	4	13990617 - 13992078	*emb1923-1*	L4	-
EMB1990	AT3G07430	3	2379193 - 2380198	*emb1990-1*	C413	Yes
EMB2001	AT2G22870	2	9739457 - 9741104	*emb2001-1*	30B4	Yes
EMB2036	AT5G66055	5	26417156 - 26419264	*emb2036-1*	-	Yes
EMB2107	AT5G09900	5	3089278 - 3092595	*emb2107*	-	Yes
EMB2301	AT2G46770	2	19220727 - 19222916	*emb2301*	7F	-
EMB2410	AT2G25660	2	10916203 - 10927390	*emb2410-1*	30B4	-
EMB2736	AT3G19980	3	6961736 - 6965108	*emb2736*	-	Yes
EMB3008	AT5G39750	5	15906875 - 15907942	*emb3008*	B111	Yes

Table 2. Primers used in this work

Gene name	Amplified region (bp)	Primers
ATSWI3A	4001	F: ggggacaagtttgtacaaaaaagcaggctACTTTCAGGTTGTTCACCAGA
		R: ggggaccactttgtacaagaaagctgggtTCTCACGTATTCCTGTCACCA
EMB1381	5694	F: ggggacaagtttgtacaaaaaagcaggctTTGGACCGTAATAACATCCCG
		R: ggggaccactttgtacaagaaagctgggtCAAAAGAGAATCCATTTCCAC
EMB1408	5191	F: ggggacaagtttgtacaaaaaagcaggctCGATCAAGCTTTGGGATCTCG
		R: ggggaccactttgtacaagaaagctgggtCCGAATATGAAAAGGCATGTC
EMB1441	8456	F: ggggacaagtttgtacaaaaaagcaggctGCTCAATTGGTAGTTGTTCTG
		R: ggggaccactttgtacaagaaagctgggtTACAAGGCCCACCCAAAGTTT
EMB1513	4593	F: ggggacaagtttgtacaaaaaagcaggctAGGCGTAAGCTCACTGTGTTG
		R: ggggaccactttgtacaagaaagctgggtTTCGAAAGAAAAATCCGACAA
EMB1586	3901	F: ggggacaagtttgtacaaaaaagcaggctGTGTTCATGACCCACGACATT
		R: ggggaccactttgtacaagaaagctgggtTTTGGCAATGGCACTAAACAA
EMB1611	7464	F: ggggacaagtttgtacaaaaaagcaggctCCTGGAAACATGACTTCGGTC
		R: ggggaccactttgtacaagaaagctgggtGGCCAGTAAAACCACCAAACC
EMB1637	4001	F: ggggacaagtttgtacaaaaaagcaggctGGTGGTGGTTTGTTGCCTTCT
		R: ggggaccactttgtacaagaaagctgggtGGGTTGGTTGCTGTTGAGATT
EMB1674	6435	F: ggggacaagtttgtacaaaaaagcaggctCACGCATGCAACAGAGATGAC
		R: ggggaccactttgtacaagaaagctgggtATGGCTCCTCTCTCCAAAGGA
EMB1688	3384	F: ggggacaagtttgtacaaaaaagcaggctGTGACTTGTTGTTTTGGTTAG
		R: ggggaccactttgtacaagaaagctgggtTTGAACTATCACGTCTTTTCC
EMB1691	7693	F: ggggacaagtttgtacaaaaaagcaggctGCCGGGTAGAGAAATACACTG
		R: ggggaccactttgtacaagaaagctgggtACCAATTTGTGGTGCGGTTGC
EMB1706	8220	F: ggggacaagtttgtacaaaaaagcaggctATCTCCTTCAAAGTTCAGCTC
		R: ggggaccactttgtacaagaaagctgggtATCTTGCTTGTGAGAAAGGCA
EMB1745	6539	F: ggggacaagtttgtacaaaaaagcaggctTGCAGGAGTAAACACAAGCGC
		R: ggggaccactttgtacaagaaagctgggtATAGAGAGAGGGTTGAGGAG
EMB1895	8109	F: ggggacaagtttgtacaaaaaagcaggctGTCTAGAGTCATGTTAGGTGG
		R: ggggaccactttgtacaagaaagctgggtTGACGTGGTGATTCTCAGTGG
EMB1990	2533	F: ggggacaagtttgtacaaaaaagcaggctTGTGTCATGGATTACTAATTT
		R: ggggaccactttgtacaagaaagctgggtCGATTTCTGGATTTGAGGTTG
EMB2001	3822	F: ggggacaagtttgtacaaaaaagcaggctCATATATGTGTTGAAAACTCA
		R: ggggaccactttgtacaagaaagctgggtGTTTGCTTGTTATATTGTGTA
EMB2036	3501	F: ggggacaagtttgtacaaaaaagcaggctTCGTCGCTGGTTCTATGGTTT
		R: ggggaccactttgtacaagaaagctgggtCTCTCAAGGAAACGTGCAAGA
EMB2107	4995	F: ggggacaagtttgtacaaaaaagcaggctCAGAGATTACAAGATATCCTG
		R: ggggaccactttgtacaagaaagctgggtACTGACTCCAGCAAAATCGGC
EMB2736	5372	F: ggggacaagtttgtacaaaaaagcaggctACAGGTATGGGCATCAGGTTT
		R: ggggaccactttgtacaagaaagctgggtACGAGCTCACAATCAGAGTAC
EMB3008	5953	F: ggggacaagtttgtacaaaaaagcaggctCTTCTGATCGGGTGCTTGATA
		R: ggggaccactttgtacaagaaagctgggtTGACTATGACGACTGTTGCTG

F: forward primer. R: reverse primer. *att*B1 and *att*B2 sites are represented in lower case.

UNIVERSITAS
Miguel
Hernández

VI.- ANEXO:
COMUNICACIONES A CONGRESOS

UNIVERSITAS
Miguel
Hernández

Genetic characterization of some leaf mutants with abnormal mesophyll growth

Candela, H., Galiana-Briones, Y., Muñoz-Nortes, T., Sánchez-García, A.B., Sempere-Ferrández, A., Casanova-Sáez, R., and Micol, J.L.

División de Genética and Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain

Despite their apparent structural simplicity, the leaves of *Arabidopsis thaliana* comprise several cell layers, each of them with a diversity of differentiated cell types. A mutational approach is ideal to identify genes that are important for the differentiation, patterning and/or proliferation of such cell types.

We will present our advances toward the map-based cloning of the *APICULATA1* (*API1*), *ANGULATA8* (*ANU8*), *ANU11*, *ANU12*, and *EROSA3* (*ERO3*) genes of *Arabidopsis thaliana*. Preliminary observations suggest that mutant alleles of these genes (which were induced by ethyl methanesulfonate) cause defective growth or differentiation of leaf mesophyll cells. The identification of the corresponding genes will help us to gain insight into the mechanisms that regulate the proliferation of mesophyll cell layers, coordinating their growth with that of other leaf tissues.

2009
Plant Growth Biology and Modeling Workshop
Elche
Póster

Genetic characterization of some leaf mutants with abnormal mesophyll growth

H. Candela, Y. Galiana-Briones, T. Muñoz-Nortes, A.B. Sánchez-García, A. Sempere-Ferrández, R. Casanova-Sáez, and J.L. Micol

División de Genética and Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
hcandela@umh.es jlmicol@umh.es http://genetica.umh.es

Despite their apparent structural simplicity, the leaves of *Arabidopsis thaliana* comprise several cell layers, each of them with a diversity of differentiated cell types. A mutational approach is ideal to identify genes that are important for the differentiation, patterning and/or proliferation of such cell types.

We present our advances toward the map-based cloning of the *APICULATA1* (*API1*), *ANGULATA8* (*ANU8*), *ANU11*, *ANU12*, and *EROSA3* (*ERO3*) genes of *Arabidopsis thaliana*[1]. Preliminary observations suggest that mutant alleles of these genes (which were induced by ethyl methanesulfonate) cause defective growth or differentiation of leaf mesophyll cells. The identification of the corresponding genes will help us to gain insight into the mechanisms that regulate the proliferation of mesophyll cell layers, coordinating their growth with that of other leaf tissues.

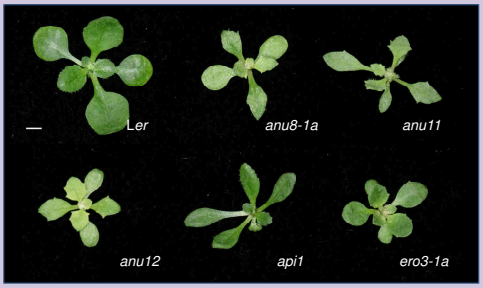


Figure 1.- Basal rosettes of the mutants described in this poster. The mutants were assigned to the Angulata (Anu), Apiculata (Api), and Erosa (Ero) phenotypic classes. Plants were grown at 20±1°C and 60-70% relative humidity, under constant fluorescent light (5,000 lx)[2]. Pictures were taken 19 days after stratification. Scale bar: 2 mm.

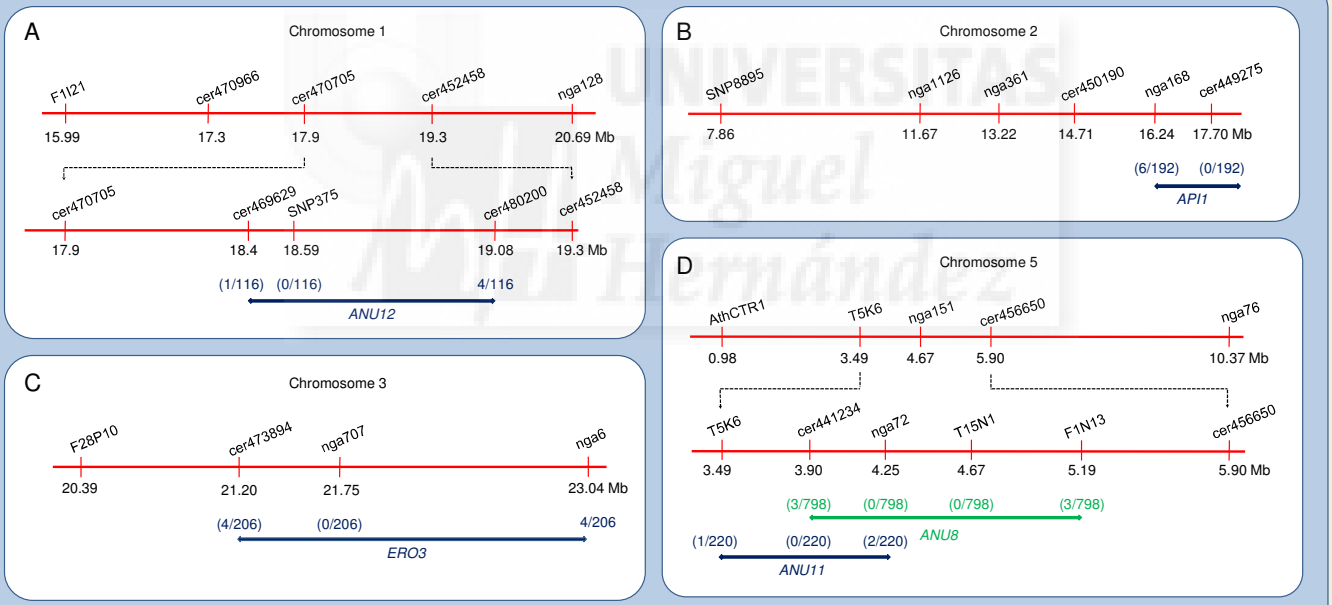


Figure 2.- Progress toward the positional cloning of the (A) *ANU12*, (B) *API1*, (C) *ERO3*, and (D) *ANU8* and *ANU11* genes of *Arabidopsis thaliana*. Values in parentheses indicates the number of informative recombinants found for each marker. Candidate intervals for each gene are given based on linkage analyses of the molecular markers shown.

REFERENCES

1.- Berná, G., Robles, P., and Micol, J.L. (1999). A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. *Genetics* **152,** 729-742.

2.- Ponce, M.R., Quesada, V., and Micol, J.L. (1998). Rapid discrimination of sequences flanking and within T-DNA insertions in the Arabidopsis genome. *Plant Journal* **14,** 497-501.

ACKNOWLEDGEMENTS

This work is supported by grants BIO2008-04075 from the Ministerio de Ciencia e Innovación of Spain, and PROMETEO/2009/112 from the Generalitat Valenciana to J.L.M.

Uncovering novel gene functions involved in leaf development

Candela, H., Casanova-Sáez, R., Muñoz-Nortes, T., Martínez-Asperilla, A., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante, Spain

The leaf of *Arabidopsis thaliana* comprises a limited number of differentiated cell types and hence is an ideal model for dissecting plant developmental processes, such as vasculature patterning or the establishment and maintenance of abaxial-adaxial polarity. We are following a two-pronged strategy to molecularly identify genes important for leaf development.

On the one hand, we are assigning a set of 28 viable leaf mutations to genomic intervals between 100 and 200 kb, which we plan to sequence using next-generation technologies to identify the causal mutations. On the other hand, we are systematically using the so-called cell autonomy (CAUT) lines to induce leaf sectors for embryo-lethal mutations, in an attempt to identify developmental functions that have eluded previous screens, which have mainly focused on viable leaf mutants.

2010
XVII Congress of the Federation of European Societies of Plant Biology
Valencia
Póster

Uncovering novel gene functions involved in leaf development

H. Candela, R. Casanova-Sáez, T. Muñoz-Nortes, A. Martínez-Asperilla, and J.L. Micol

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
hcandela@umh.es jlmicol@umh.es genetica.umh.es

The leaf of *Arabidopsis thaliana* comprises a limited number of differentiated cell types and hence is an ideal model for dissecting plant developmental processes, such as vascular patterning or the establishment and maintenance of abaxial-adaxial polarity. We are following a two-pronged strategy to molecularly identify genes important for leaf development.

On the one hand, we are assigning a set of 28 viable leaf mutations to genomic intervals between 100 and 200 kb. These mutants belong to the Angusta (Ang), Angulata (Anu), Apiculata (Api), Asymmetric leaves (As), Erosa (Ero), Incurvata (Icu) and Extrahydathodes (Ehy) phenotypic classes[1,2] (Figure 1). Preliminary low-resolution map positions were available for 24 genes[2]. We are refining those positions and have determined the positions of three genes that had not been mapped previously (*API2*, *EHY*, and *ICU7*) (Figure 2). In the first year, our strategy has already allowed us to clone three genes (*ANG1*, *API2* and *ANU10*) and to establish candidate intervals shorter than 200 kb for four more genes (*ANU1*, *ANU3*, *ANU4* and *ANU9*). For such short intervals, we plan to use next-generation sequencing technologies to identify the causal mutations.

On the other hand, in an attempt to identify developmental functions that have eluded previous screens, we are using the so-called cell autonomy (CAUT) lines to induce leaf sectors for embryo-lethal mutations (Figure 3). These mutations damage 35 genes that we have selected based on several criteria: (1) their mutations are lethal, (2) they are also expressed at post-embryonic stages, and (3) they encode proteins with potentially regulatory functions.

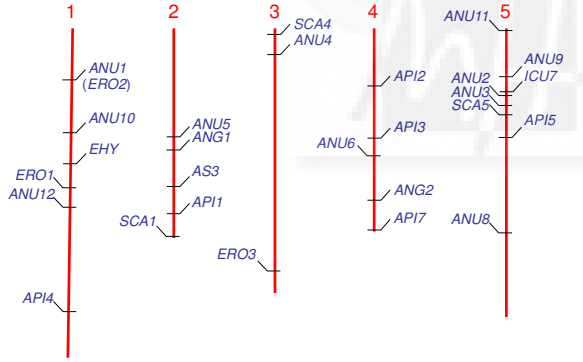


Figure 2.- Genetic map of *Arabidopsis thaliana*. Low-resolution map positions for most of the genes under study were already available[2] and have subsequently been refined by us. Using such positions as a starting point, we are fine mapping all the genes in a positional cloning approach.

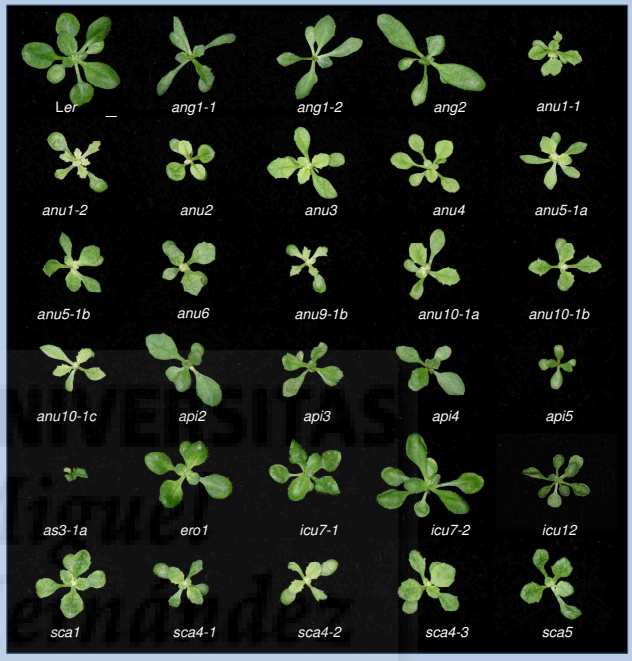
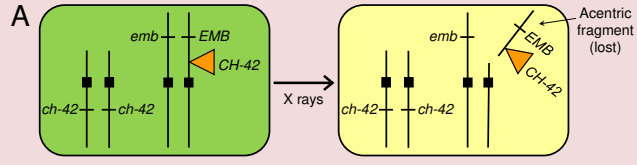


Figure 1.- Rosettes of plants carrying mutant alleles of selected *ANGUSTA* (*ANG*), *ANGULATA* (*ANU*), *APICULATA* (*API*), *ASYMMETRIC LEAVES* (*AS*), *EROSA* (*ERO*), *INCURVATA* (*ICU*) and *SCABRA* (*SCA*) genes. *ang* mutants have narrow leaves. *anu* mutant leaves have prominent teeth and light pigmentation. *api* mutant leaves are pointed, with prominent teeth but normal pigmentation. *ero* mutant leaves have uneven surfaces. *as* and *icu* mutant leaves curl abnormally. *sca* mutant leaves are variegated. Plants were grown for 19 days at 20±1°C and 60-70% relative humidity under constant light (5,000 lx)[3]. Scale bar: 2 mm.



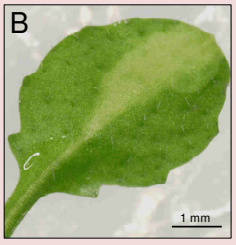


Figure 3.- CAUT lines are homozygous for a recessive yellow (non-photosynthetic) marker, *ch-42*, but they exhibit a wild-type (green) phenotype due to genetic complementation with a wild-type allele provided by a transgene (*CH-42*). (A) Sectors of yellow cells appear occasionally when the chromosome arm that carries the transgene is lost after a treatment with a high dose of X-rays. In a strain with the appropriate genotype, loss of the chromosome arm can also be used to create a sector that is hemizygous for an *embryo-lethal* mutation (*emb*), which will be marked by the yellow phenotype. Sectors are screened for mutant phenotypes. (B) Leaf with a sector that is hemizygous for the *emb1990* mutation, induced by X-rays (16,000 rad).

REFERENCES

1.- Berná, G., Robles, P., and Micol, J.L. (1999). A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. *Genetics* **152,** 729-742.

2.- Robles, P., and Micol, J.L. (2001). Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. *Molecular Genetics and Genomics* **266,** 12-19.

3.- Ponce, M.R., Quesada, V., and Micol, J.L. (1998). Rapid discrimination of sequences flanking and within T-DNA insertions in the Arabidopsis genome. *Plant Journal* **14,** 497-501.

4.- Furner, I., Ellis, L., Bakht, S., Mirza, B., and Sheikh, M. (2008). CAUT lines: a novel resource for studies of cell autonomy in Arabidopsis. *Plant Journal* **53,** 645-660.

ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M. is supported by grants from the Ministerio de Ciencia e Innovación of Spain [BIO2008-04075 and CSD2007-00057 (TRANSPLANTA)], the Generalitat Valenciana (PROMETEO/2009/112) and the European Commission [LSHG-CT-2006-037704 (AGRON-OMICS)]. R.C.-S. holds a fellowship from the Ministerio de Ciencia e Innovación. H.C. is the recipient of a Marie Curie International Reintegration Grant.

Búsqueda de nuevas funciones génicas implicadas en el desarrollo foliar

Candela, H., Casanova-Sáez, R., Muñoz-Nortes, T., Martínez-Asperilla, A., y Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante

La estructura de la hoja de *Arabidopsis thaliana* es muy simple, ya que incluye pocos histotipos. Es por eso que constituye un buen modelo para la disección genética de determinados procesos del desarrollo vegetal, como la formación del patrón vascular o el establecimiento y mantenimiento de la polaridad abaxial-adaxial.

Hemos iniciado dos abordajes paralelos para la identificación molecular de nuevos genes implicados en el desarrollo foliar. Estamos sometiendo a análisis de ligamiento a 28 mutantes foliares viables, con el objetivo de definir intervalos candidatos de 100-200 kb, que secuenciaremos utilizando tecnologías de alto rendimiento. Pretendemos también identificar aspectos del desarrollo foliar que hayan pasado inadvertidos en escrutinios anteriores, cuyo objetivo fue el aislamiento de mutantes viables. Estamos empleando para ello la colección de líneas CAUT ("cell autonomy") a fin de inducir sectores foliares que manifiesten los efectos locales de la insuficiencia de función de genes letales embrionarios.

2010
X Reunión de Biología Molecular de Plantas
Valencia
Póster

Búsqueda de nuevas funciones génicas implicadas en el desarrollo foliar

H. Candela, R. Casanova-Sáez, T. Muñoz-Nortes, A. Martínez-Asperilla y J.L. Micol

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante.
hcandela@umh.es jlmicol@umh.es genetica.umh.es

La estructura de la hoja de Arabidopsis thaliana es muy simple, ya que incluye pocos histotipos. Es por eso que constituye un buen modelo para la disección genética de determinados procesos del desarrollo vegetal, como la formación del patrón vascular o el establecimiento y mantenimiento de la polaridad abaxial-adaxial. Hemos iniciado dos abordajes paralelos para la identificación molecular de nuevos genes implicados en el desarrollo foliar.

Por una parte, estamos sometiendo a análisis de ligamiento a 28 mutantes foliares viables[1,2], con el objetivo de definir intervalos candidatos de 100-200 kb, que secuenciaremos utilizando tecnologías de alto rendimiento. Estos mutantes pertenecen a las clases fenotípicas Angusta (Ang), Angulata (Anu), Apiculata (Api), Asymmetric leaves (As), Erosa (Ero), Incurvata (Icu) y Extrahydathodes (Ehy) (Figura 1). Hemos iniciado la cartografía de alta resolución a partir de las posiciones de mapa de baja resolución disponibles para 24 de los genes (Figura 2). También hemos cartografiado tres genes cuya posición se desconocía (*API2*, *EHY* e *ICU7*). En el primer año, esta estrategia nos ha permitido ya clonar tres genes (*ANG1*, *API2* y *ANU10*) y delimitar intervalos candidatos de tamaño inferior a 200 kb para otros cuatro genes (*ANU1*, *ANU3*, *ANU4* y *ANU9*). Para estos últimos, planeamos identificar las mutaciones responsables de los fenotipos observados utilizando las nuevas tecnologías de secuenciación masiva.

Por otra parte, pretendemos también identificar aspectos del desarrollo foliar que hayan pasado inadvertidos en escrutinios anteriores, cuyo objetivo fue el aislamiento de mutantes viables. Estamos empleando para ello la colección de líneas CAUT ("cell autonomy")[4], a fin de inducir sectores foliares que manifiesten los efectos locales de la insuficiencia de función de genes letales embrionarios (Figura 3). Estamos caracterizando 35 genes, que fueron seleccionados en base a: (1) la letalidad de sus mutaciones, (2) su expresión en etapas posteriores a la embriogénesis, y (3) la posible función reguladora de sus productos.

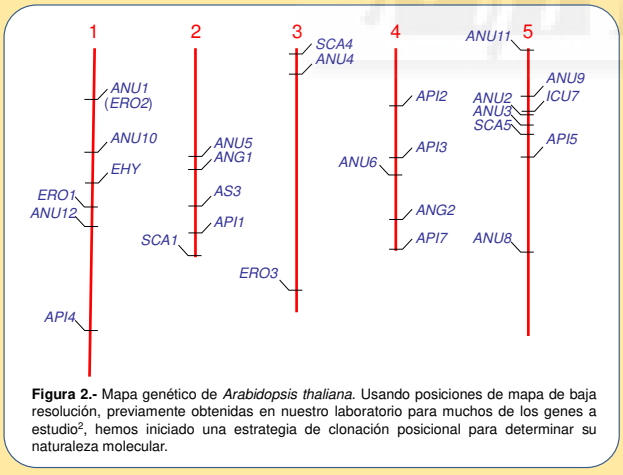


Figura 2.- Mapa genético de *Arabidopsis thaliana*. Usando posiciones de mapa de baja resolución, previamente obtenidas en nuestro laboratorio para muchos de los genes a estudio[2], hemos iniciado una estrategia de clonación posicional para determinar su naturaleza molecular.

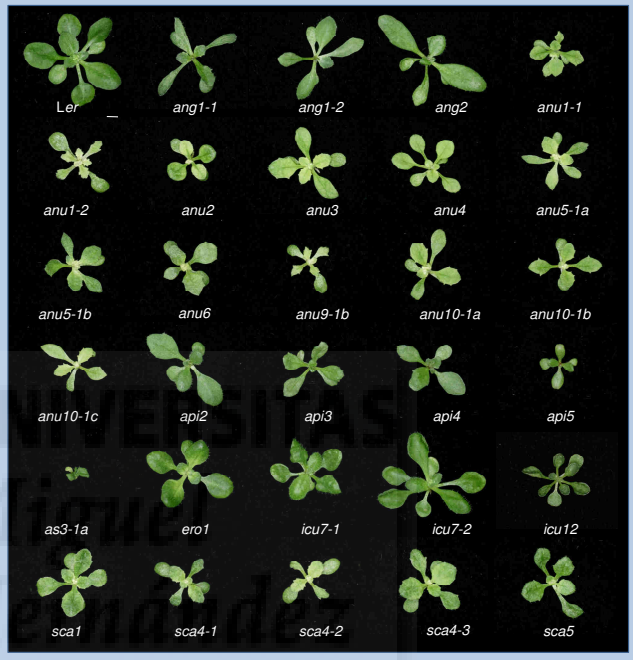


Figura 1.- Rosetas de plantas portadoras de alelos mutantes de algunos genes *ANGUSTA* (*ANG*), *ANGULATA* (*ANU*), *APICULATA* (*API*), *ASYMMETRIC LEAVES* (*AS*), *EROSA* (*ERO*), *INCURVATA* (*ICU*) y *SCABRA* (*SCA*). Los mutantes *ang* tienen hojas estrechas. Las hojas de los mutantes *anu* tienen pigmentación amarillenta y un margen foliar con indentaciones prominentes. Las de los mutantes *api* son apuntadas, con indentaciones prominentes y pigmentación normal. Las de los mutantes *ero* presentan una superficie irregular. Las hojas de los mutantes *as* e *icu* se curvan de manera anormal. Las de los mutantes *sca* son variegadas. Las plantas se cultivaron durante 19 días a 20±1°C y humedad relativa del 60-70% con iluminación constante (5.000 lx)[3]. Barra de escala: 2 mm.

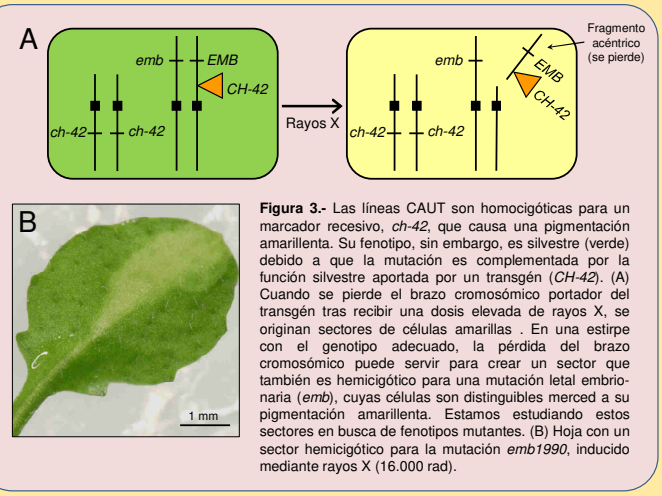


Figura 3.- Las líneas CAUT son homocigóticas para un marcador recesivo, *ch-42*, que causa una pigmentación amarillenta. Su fenotipo, sin embargo, es silvestre (verde) debido a que la mutación es complementada por la función silvestre aportada por un transgén (*CH-42*). (A) Cuando se pierde el brazo cromosómico portador del transgén tras recibir una dosis elevada de rayos X, se originan sectores de células amarillas . En una estirpe con el genotipo adecuado, la pérdida del brazo cromosómico puede servir para crear un sector que también es hemicigótico para una mutación letal embrionaria (*emb*), cuyas células son distinguibles merced a su pigmentación amarillenta. Estamos estudiando estos sectores en busca de fenotipos mutantes. (B) Hoja con un sector hemicigótico para la mutación *emb1990*, inducido mediante rayos X (16.000 rad).

BIBLIOGRAFÍA

1.- Berná, G., Robles, P., y Micol, J.L. (1999). A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. *Genetics* **152,** 729-742.

2.- Robles, P., y Micol, J.L. (2001). Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. *Molecular Genetics and Genomics* **266,** 12-19.

3.- Ponce, M.R., Quesada, V., y Micol, J.L. (1998). Rapid discrimination of sequences flanking and within T-DNA insertions in the Arabidopsis genome. *Plant Journal* **14,** 497-501.

4.- Furner, I., Ellis, L., Bakht, S., Mirza, B., y Sheikh, M. (2008). CAUT lines: a novel resource for studies of cell autonomy in Arabidopsis. *Plant Journal* **53,** 645-660.

AGRADECIMIENTOS

La investigación en el laboratorio de J.L.M. está financiada por proyectos del Ministerio de Ciencia e Innovación [BIO2008-04075 y CSD2007-00057 (TRANSPLANTA)], la Generalitat Valenciana (PROMETEO/2009/112) y la Comisión Europea [LSHG-CT-2006-037704 (AGRONOMICS)]. R.C.-S. es becario del Ministerio de Ciencia e Innovación. H.C. es beneficiario de una Marie Curie International Reintegration Grant.

Clonal analysis of the post-embryonic function of embryo-lethal genes in *Arabidopsis thaliana*

Muñoz-Nortes,T., Tremiño, L., Madrigal, P., Candela, H., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Spain

Several hundred genes in the *Arabidopsis thaliana* genome are necessary for embryonic and gametophytic development as inferred from the lethality of their loss-of-function mutations. Despite many embryo-lethal genes are normally expressed throughout all stages of plant development, the corresponding mutants die at early stages, preventing the study of their post-embryonic functions using conventional methods. Clonal analysis techniques provide an effective solution to this problem by uncovering the effects of embryo-lethal mutations in sectors of mutant cells within an otherwise normal adult plant.

We have selected 35 embryo-lethal mutants of *Arabidopsis thaliana* for clonal analysis experiments, with a focus on their effects on leaf development. For the induction of mutant sectors in adult plants, we are using two different approaches: one based on the X-ray irradiation of so-called 'cell autonomy' lines (CAUT; Furner *et al.*, 2008), and another based on the site-specific excision of transgenes mediated by Cre recombinase (Heidstra *et al.*, 2004).

References

1. Furner, I., Ellis, L., Bakht, S., Mirza, B., and Sheikh, M. (2008). CAUT lines: a novel resource for studies of cell autonomy in Arabidopsis. *Plant J.* **53,** 645-660.
2. Heidstra, R., Welch, D., and Scheres, B. (2004). Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis* SCARECROW action in asymmetric cell division. *Genes Dev.* **18**, 1964-1969.

2011
II Plant Growth Biology and Modeling Workshop
Elche
Póster

Clonal analysis of the post-embryonic function of embryo-lethal genes in *Arabidopsis thaliana*

Muñoz-Nortes, T., Tremiño, L., Madrigal, M.P., Candela, H., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
tmunoz@umh.es hcandela@umh.es jlmicol@umh.es http://genetica.umh.es

As inferred from the lethality of their loss-of-function mutations, several hundred genes in the *Arabidopsis thaliana* genome are necessary for embryonic and gametophytic development. Despite many embryo-lethal genes are normally expressed throughout all stages of plant development, the corresponding mutants die at early stages, preventing the study of their post-embryonic functions using conventional methods. Clonal analysis techniques provide an effective solution to this problem by uncovering the effects of embryo-lethal mutations in sectors of mutant cells within an otherwise normal adult plant[1].

We have selected 35 embryo-lethal mutants of *Arabidopsis thaliana* for clonal analysis experiments, with a focus on their effects on leaf development. Some of the corresponding genes encode transcription factors or participate in the chromatin-mediated regulation of gene expression. Other group of genes have putative roles in proteasome-mediated protein degradation.

For the induction of mutant sectors in adult plants, we are using two different approaches: one based on the X-ray irradiation of so-called 'cell autonomy' (CAUT) lines[2], and another based on the site-specific excision of transgenes mediated by Cre recombinase.

For the first approach, we generated lines that are homozygous for the pale-green *chlorata-42* (*ch-42*) mutation and heterozygous for selected *embryo-defective* (*emb*) mutations[3]. Individual lines were then crossed to appropriate CAUT lines carrying a transgene with a wild-type copy of *CHLORATA-42* located between the centromere and the corresponding *EMB* gene (Figure 1). To induce hemizygous sectors for each *emb* mutation, we have already irradiated (with X-rays) seeds from lines carrying 13 different *emb* mutations. The examination of irradiated lines showed pale-green and albino sectors (Figure 2). In some families, we occasionally found plants that were homozygous for an *emb* mutation but were able to grow despite their embryo-lethality (escapers, Figure 3).

For the second approach, we are generating constructs (carrying a wild-type copy of a given *EMB* gene) to complement the embryonic lethality of the corresponding *emb* mutation. We have adapted the pCB1 vector[4] for use with the Gateway technology. In this vector, the wild-type copy of an *EMB* gene is flanked by two *loxP* sites, making possible to create GFP-marked, mutant sectors when the Cre recombinase is expressed under the control of a heat shock promoter (Figure 4).

MATERIALS AND METHODS

Plants were grown at 20±1°C or 26±1°C and 60-70% relative humidity under continuous fluorescent light (5,500 lx). Irradiation of seeds with X-rays was performed using a Philips MG102 cabin and the doses of irradiation were 10 Gy for wet seeds and 160 Gy for dry seeds. We cloned a Gateway cassette in the *Not*I site of pCB1, in both orientations, to generate the pCB1-Gateway(+) and pCB1-Gateway(-) vectors. Large genomic inserts were amplified using *Phusion* polymerase and primers containing *attB1* and *attB2* sites, subsequently cloned into a donor vector using a BP reaction, and finally mobilized into pCB1-Gateway(+) or pCB1-Gateway(-) using an LR reaction (Invitrogen).

ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M is supported by grants BIO2008-04075 and CSD2007-00057 (TRANSPLANTA) from the Ministerio de Ciencia e Innovación of Spain, PROMETEO/2009/112 from the Generalitat Valenciana and LSHG-CT-2006-037704 (AGRON-OMICS) from the European Commission. We want to thank Ginés Morata, Renze Heidstra and Guy Wachsman for sharing their materials. H.C. is the recipient of a Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073).

REFERENCES

1.- Candela, H., Pérez-Pérez, J.M., and Micol, J.L. (2011). *Trends Plant Sci.* **16**, 336-345.
2.- Furner, I., Ellis, L., Bakht, S., Mirza, B., and Sheikh, M. (2008). *Plant J.* **53,** 645-660.
3.- Meinke, D.W., and Sussex, I.M. (1979b). *Dev. Biol.* **72,** 62-72.
4.- Heidstra, R., Welch, D., and Scheres, B. (2004). *Genes Dev.* **18**, 1964-1969.

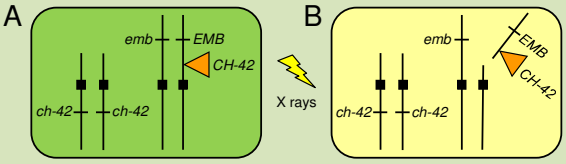


Figure 1.- Approach based on CAUT lines. (A) CAUT lines are homozygous for a recessive yellow (non-photosynthetic) marker, *ch-42*, but they exhibit a wild-type (green) phenotype due to genetic complementation with a wild-type allele supplied by a transgene (*CH-42*). (B) Sectors of yellow cells appear occasionally when the chromosome arm that carries the transgene is lost after a treatment with a high dose of X-rays. In a strain with the appropriate genotype, loss of the chromosome arm can also be used to create a sector that is also hemizygous for an *embryo-defective* mutation (*emb*), which will be marked by the yellow phenotype. Then, sectors have to be screened in order to find mutant phenotypes related to the *emb* mutation.

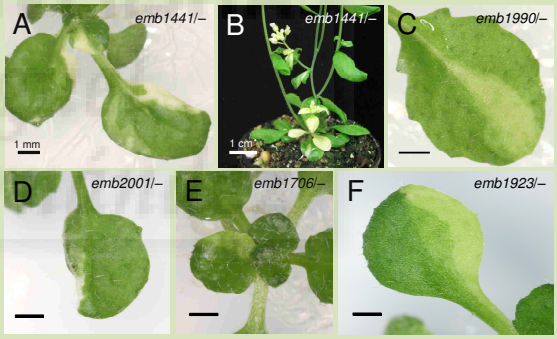


Figure 2.- Irradiation with X-rays induce pale-green or completely albino sectors in several leaves, and these sectors are assumed to be hemizygous for an *emb* mutation. (A-F) Leaves with sectors that are hemizygous for (A, B) *emb1441*, (C) *emb1990*, (D) *emb2001*, (E) *emb1706*, and (F) *emb1923* mutations.

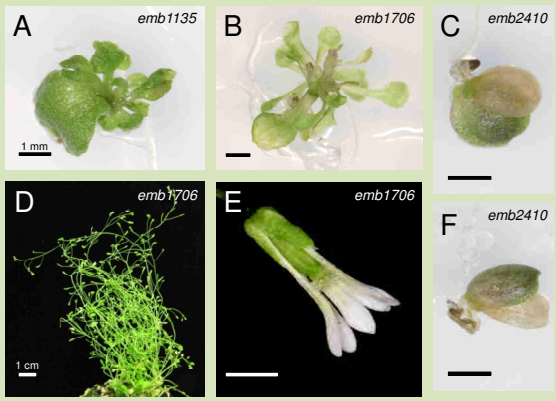


Figure 3.- Escapers are embryo-lethal mutants that survive beyond the embryogenesis and exhibit a postembryonic phenotype putatively associated with the *emb* mutation. (A-F) Escapers found in the screening of irradiated lines that were segregating for the (A) *emb1135* (B, D, E) *emb1706*, and (C, F) *emb2410* mutations.

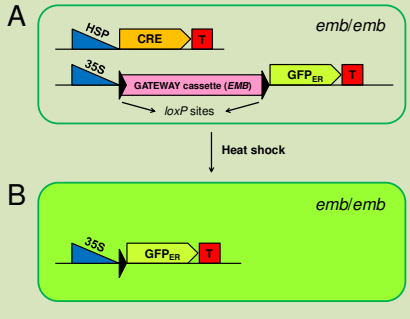


Figure 4.- Transgene-based approach. (A) An homozygous *emb*/*emb* plant carries two different constructs, one of them providing a wild-type copy of an *EMB* gene that allows its normal development, and the other with an heat-shock promoter driving the inducible expression of Cre recombinase. (B) A heat pulse causes the activation of Cre and a concomitant loss of the wild-type copy of the *EMB* gene in some cells. Then, the expression of GFP under the 35S promoter marks the lineage of those cells, which might exhibit any mutant phenotype that is associated with the loss of function of the *EMB* gene in adult tissues.

Forward and reverse approaches to the genetic dissection of leaf development

Mollá-Morales, A., Sarmiento-Mañús, R., Ferrández-Ayela, A., Rubio-Díaz, S, Muñoz-Viana, R., Esteve-Bruna, D., Casanova-Sáez, R., Muñoz-Nortes, T., Wilson-Sánchez, D., González-Bayón, R., Jover-Gil, S., Candela, H., Pérez-Pérez, J.M., Ponce, M.R., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández, 03202 Elche, Alicante, Spain.

Plant leaves are the best solar panels ever built, and they also perform well as air purifiers and food factories. Leaves efficiently trap sunlight, remove carbon dioxide from the air, and are the ultimate source of most of the oxygen that we breathe and of the food that we eat. Understanding how a leaf is made is important for several reasons, which include gaining knowledge of the biology and evolution of a multicellular organ with no equivalents in the animal kingdom, as well as identifying —and eventually manipulating, to increase crop yield— the genetic, environmental, and hormonal cues that determine its final architecture and function.

To shed light on the making of plant leaves, we took a forward genetics approach to the saturation of the Arabidopsis genome with viable mutations causing abnormal leaf morphology. The identified mutations fell into 147 complementation groups. Using a high throughput gene mapping method that we developed, we have already cloned 47 of these genes identified by mutation. The products of these genes participate in various developmental processes, such as polar cell expansion, transduction of hormonal signals, gene regulation, plastid biogenesis, and chromatin remodeling, among others. The broad spectrum of leaf morphological alterations that we identified is helping to dissect specific leaf developmental processes.

We are now combining traditional linkage analysis and next-generation sequencing techniques in order to positionally clone 40 non-allelic mutations already isolated in our laboratory, which affect leaf morphology. In addition, we have started to use clonal analysis to study essential, embryonic-lethal genes that are expressed in wild-type leaves.

We also aim at identifying genes involved in the development of an organ —the leaf— at a scale with no precedent in plants, and perhaps animals. The Ecker laboratory is producing 50,000 sequence-indexed homozygous T-DNA lines representing all the protein coding genes in the Arabidopsis genome. We are searching this collection for leaf morphological aberrations, and have already identified 340 genes required for normal leaf

development. These leaf mutants provide an opportunity to propose models and test hypotheses about how genes control plant development at the organ level.

2011

II Plant Growth Biology and Modeling Workshop

Elche

Conferencia pronunciada por J.L. Micol, por invitación.

Análisis clonal de las funciones postembrionarias de genes letales embrionarios en *Arabidopsis thaliana*

Muñoz-Nortes T, Tremiño L, Candela H, y Micol JL

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante, Spain

El genoma de *Arabidopsis thaliana* contiene centenares de genes que son necesarios para el desarrollo embrionario y/o gametofítico, cuyas mutaciones de pérdida de función causan letalidad. Aunque muchos de los genes letales embrionarios se expresan también en otras etapas del desarrollo de la planta, sus funciones postembrionarias no pueden ser estudiadas mediante métodos convencionales, dado que los mutantes mueren en estadios tempranos. Las técnicas del análisis clonal proporcionan una solución a este problema, al posibilitar el estudio del fenotipo de las mutaciones letales embrionarias en grupos de células de plantas adultas.

Hemos elegido 35 genes letales embrionarios para su análisis clonal en las hojas de *Arabidopsis thaliana*. Para la inducción de sectores mutantes en plantas adultas estamos utilizando dos abordajes distintos: uno basado en la irradiación de líneas CAUT (de “cell autonomy”; Furner *et al*., 2008) con rayos X, y otro basado en la recombinación específica de sitio mediada por la recombinasa Cre en plantas transgénicas.

2011
XXXVIII Congreso de la Sociedad Española de Genética
Murcia
Póster

Análisis clonal de las funciones postembrionarias de genes letales embrionarios en *Arabidopsis thaliana*

Muñoz-Nortes, T., Tremiño, L., Candela, H., y Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
tmunoz@umh.es hcandela@umh.es jlmicol@umh.es http://genetica.umh.es

El genoma de *Arabidopsis thaliana* contiene centenares de genes necesarios para el desarrollo embrionario y/o gametofítico, cuyas mutaciones de pérdida de función causan letalidad. Aunque muchos de los genes letales embrionarios se expresan también en otras etapas del desarrollo de la planta, sus funciones postembrionarias no pueden ser estudiadas mediante métodos convencionales, dado que los mutantes mueren en estadios tempranos. Las técnicas del análisis clonal proporcionan una solución a este problema, al posibilitar el estudio del fenotipo de las mutaciones letales embrionarias en grupos de células de plantas adultas[1].

Hemos elegido 35 mutantes letales embrionarios para su análisis clonal en las hojas de *Arabidopsis thaliana*. Algunos de los genes correspondientes a estas mutaciones codifican factores de transcripción o participan en la regulación de la expresión génica mediada por la cromatina. Otros podrían estar implicados en la degradación de proteínas por el proteasoma.

Para la inducción de sectores mutantes en plantas adultas estamos utilizando dos abordajes distintos: uno basado en la irradiación de líneas CAUT (de *cell autonomy*)[2] con rayos X, y otro basado en la recombinación específica de sitio de transgenes mediada por la recombinasa Cre.

Para el primer abordaje, hemos generado líneas homocigóticas para la mutación *chlorata-42* (*ch-42*) y heterocigóticas para las mutaciones letales embrionarias (*emb*)[3] de interés. Cada una de estas líneas fue cruzada con la línea CAUT apropiada, portadora de un transgén con una copia silvestre de *CHLORATA-42* situada entre el centrómero y el gen *EMB* correspondiente (Figura 1). Con el fin de inducir sectores hemicigóticos para cada mutación, hemos irradiado con rayos X semillas procedentes de líneas portadoras de 13 mutaciones *emb* distintas. En el escrutinio de estas líneas encontramos sectores de color verde pálido y sectores albinos (Figura 2). Ocasionalmente encontramos también plantas que, siendo homocigóticas para una de las mutaciones *emb* a estudio, fueron capaces de crecer a pesar de su letalidad embrionaria (*escapers*, Figura 3).

Para el segundo abordaje, estamos generando construcciones portadoras de una copia silvestre de un gen *EMB* dado, con el fin de complementar la letalidad embrionaria causada por las mutaciones *emb*. Hemos adaptado el vector pCB1[4] a la tecnología Gateway. En este vector, la copia silvestre del gen *EMB* de interés está flanqueada por dos sitios *loxP*, haciendo posible la generación de sectores mutantes marcados con GFP cuando la recombinasa Cre se expresa bajo el control de un promotor de choque térmico, como se muestra en la Figura 4.

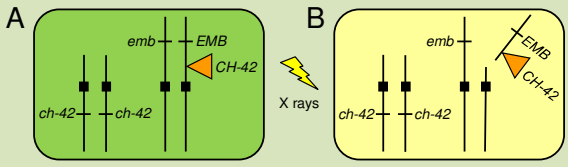


Figura 1.- Estrategia basada en las líneas CAUT. (A) Las líneas CAUT son homocigóticas para el marcador amarillento *ch-42*, pero muestran un fenotipo silvestre (verde) debido a que un transgén con una copia silvestre *CH-42* complementa el fenotipo mutante. (B) Aparecen ocasionalmente sectores de células amarillentas cuando el brazo cromosómico portador del transgén se pierde tras irradiación con rayos X. En una estirpe con el genotipo apropiado, la pérdida del brazo cromosómico puede aprovecharse para generar un sector hemicigótico para una mutación letal embrionaria (*emb*), que quedará marcado por su color amarillo. Deben ser examinados a continuación los sectores para encontrar fenotipos mutantes relacionados con la mutación *emb*.

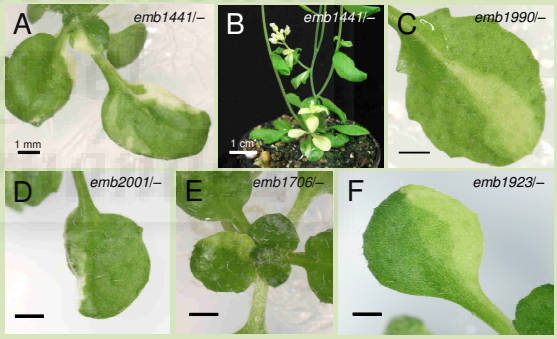


Figura 2.- La irradiación con rayos X induce sectores de color verde pálido o completamente albinos en algunas hojas, y estos sectores son supuestamente hemicigotos para una mutación *emb*. (A-F) Hojas con sectores hemicigóticos para las mutaciones (A, B) *emb1441*, (C) *emb1990*, (D) *emb2001*, (E) *emb1706* y (F) *emb1923*.

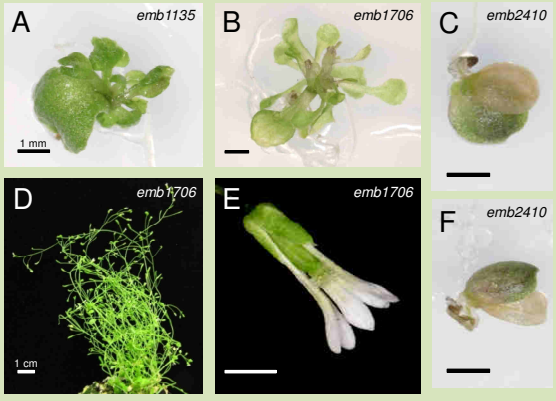


Figura 3.- Los *escapers* son mutantes letales embrionarios que consiguen superar la fase de embriogénesis y muestran un fenotipo postembrionario supuestamente causado por la mutación *emb*. (A-F) *Escapers* encontrados en el escrutinio de familias irradiadas en las que segregaban mutantes (A) *emb1135* (B, D, E) *emb1706* y (C, F) *emb2410*.

MATERIALES Y MÉTODOS

Las plantas se cultivaron a 20±1°C o 26±1°C con una humedad relativa del 60-70% y bajo luz continua (5.500 lx). La irradiación de semillas con rayos X se llevó a cabo usando una cabina Philips MG102, y las dosis de irradiación fueron 10 Gy para semillas en agua y 160 Gy para semillas secas. Clonamos un cassette Gateway en el sitio *Not*I del vector pCB1 en las dos orientaciones posibles, para generar los vectores pCB1-Gateway(+) y pCB1-Gateway(-). Amplificamos insertos grandes de ADN genómico utilizando la polimerasa *Phusion* y cebadores con sitios *attB1* y *attB2*, y los clonamos en un vector de entrada mediante una reacción BP. A continuación movilizamos dichos insertos hacia los vectores de destino pCB1-Gateway(+) y pCB1-Gateway(-) mediante una reacción LR (Invitrogen).

AGRADECIMIENTOS

La investigación en el laboratorio de J.L.M está financiada por el Ministerio de Ciencia e Innovación [BIO2008-04075 y CSD2007-00057 (TRANSPLANTA)], la Generalitat Valenciana (PROMETEO/2009/112) y la Comisión Europea [LSHG-CT-2006-037704 (AGRON-OMICS)]. H.C. es beneficiario de una Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073). Agradecemos a Ginés Morata, Renze Heidstra y Guy Wachsman los materiales cedidos.

BIBLIOGRAFÍA

1.- Candela, H., Pérez-Pérez, J.M., y Micol, J.L. (2011). *Trends Plant Sci.* **16**, 336-345.
2.- Furner, I., Ellis, L., Bakht, S., Mirza, B., y Sheikh, M. (2008). *Plant J.* **53,** 645-660.
3.- Meinke, D.W., y Sussex, I.M. (1979b). *Dev. Biol.* **72,** 62-72.
4.- Heidstra, R., Welch, D., y Scheres, B. (2004). *Genes Dev.* **18**, 1964-1969.

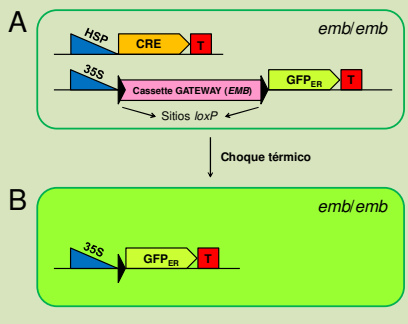


Figura 4.- Estrategia basada en transgenes. (A) Una planta de genotipo mutante *emb*/*emb* es portadora de dos construcciones distintas, una que aporta una copia silvestre de un gen *EMB* y hace posible su desarrollo normal y la otra con un promotor de choque térmico que controla la expresión de la recombinasa Cre. (B) Un choque térmico provoca la activación de Cre y la pérdida concomitante del alelo silvestre *EMB* en algunas células. La expresión de la proteína fluorescente verde (GFP) bajo el control del promotor 35S marca el linaje de esas células, que pueden mostrar un fenotipo mutante asociado con la disfunción local del gen *EMB* en la planta adulta.

Análisis clonal de las funciones postembrionarias de genes letales embrionarios en *Arabidopsis thaliana*

Tamara Muñoz-Nortes, Héctor Candela y José Luis Micol

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante

El genoma de *Arabidopsis thaliana* contiene centenares de genes que son necesarios para el desarrollo embrionario y/o gametofítico, cuyas mutaciones de pérdida de función causan letalidad recesiva. Aunque muchos de los genes letales embrionarios se expresan también en otras etapas del desarrollo de la planta, sus funciones postembrionarias no pueden ser estudiadas mediante métodos convencionales, dado que los mutantes mueren en estadios tempranos. Las técnicas de análisis clonal proporcionan una solución a este problema, al posibilitar el estudio del fenotipo de las mutaciones letales embrionarias en grupos de células de plantas adultas.

Hemos elegido 35 genes letales embrionarios para su análisis clonal en las hojas de *Arabidopsis thaliana*. Para la inducción de sectores mutantes en plantas adultas estamos utilizando dos abordajes distintos, uno basado en la irradiación de líneas CAUT (de “cell autonomy”; Furner *et al*., 2008) con rayos X, y otro basado en la recombinación específica de sitio mediada por la recombinasa Cre en plantas transgénicas. Hemos irradiado semillas procedentes de líneas portadoras de 13 mutaciones *emb* distintas, con el fin de inducir sectores hemicigóticos para cada mutación. Hemos realizado a continuación un escrutinio de las plantas obtenidas a partir de las semillas irradiadas. También hemos generado construcciones portadoras de una copia silvestre de 17 genes *EMB* distintos, con el fin de complementar la letalidad embrionaria. Combinando estas construcciones con otras en las que la recombinasa Cre se expresa bajo el control de un promotor de choque térmico generaremos sectores mutantes *emb*/*emb* marcados con una proteína verde fluorescente.

2012
XI Reunión de Biología Molecular de Plantas
Segovia
Póster

Análisis clonal de las funciones postembrionarias de genes letales embrionarios en *Arabidopsis thaliana*

Muñoz-Nortes, T., Candela, H., y Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante.
tmunoz@umh.es hcandela@umh.es jlmicol@umh.es http://genetica.umh.es

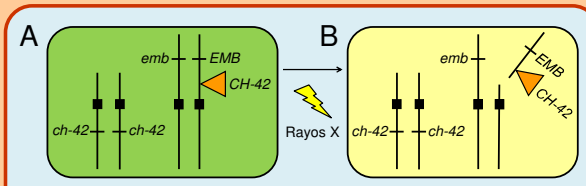


Figura 1.- Estrategia basada en las líneas CAUT. (A) Las líneas CAUT son líneas transgénicas homocigóticas para una mutación recesiva, *chlorata-42* (*ch-42*), que confiere a la planta una pigmentación amarillenta. Las plantas CAUT, sin embargo, son verdes, ya que son portadoras de un transgén que contiene una copia del alelo silvestre del gen *CH-42*. (B) Aparecerán sectores marcados (amarillentos) cuando el brazo cromosómico portador del transgén *CH-42* se pierda a consecuencia de una ruptura cromosómica inducida por los rayos X. En una línea con el genotipo apropiado, la pérdida de un brazo cromosómico puede generar un sector hemicigótico para una mutación letal embrionaria (*emb*), y a la vez marcarlo con el fenotipo amarillento.

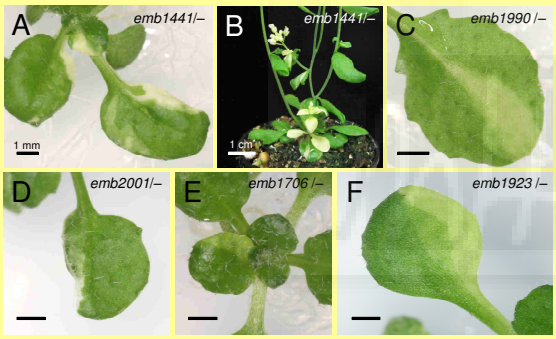


Figura 2.- La irradiación con rayos X induce sectores de color verde pálido o completamente albinos en algunas hojas, presuntamente hemicigóticos para una mutación *emb*. (A-F) Hojas con sectores hemicigóticos para las mutaciones (A, B) *emb1441*, (C) *emb1990*, (D) *emb2001*, (E) *emb1706* y (F) *emb1923*.

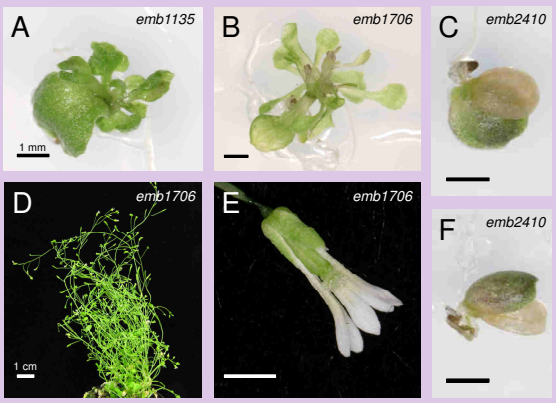


Figura 3.- Los *escapers* son mutantes letales embrionarios que consiguen superar la embriogénesis y muestran un fenotipo postembrionario supuestamente causado por la mutación *emb*. (A-F) *Escapers* encontrados en el escrutinio de familias irradiadas en las que segregaban mutantes (A) *emb1135*, (B, D, E) *emb1706* y (C, F) *emb2410*.

El genoma de *Arabidopsis thaliana* contiene centenares de genes necesarios para el desarrollo embrionario y/o el gametofítico, cuyas mutaciones de pérdida de función causan letalidad. Aunque muchos de los genes letales embrionarios se expresan también en otras etapas del desarrollo de la planta, sus funciones postembrionarias no pueden ser estudiadas mediante métodos convencionales, dado que los mutantes mueren en estadios tempranos. Las técnicas del análisis clonal proporcionan una solución a este problema, al posibilitar el estudio del fenotipo de las mutaciones letales embrionarias en grupos de células de plantas adultas[1].

Hemos elegido 35 mutaciones letales embrionarias para su análisis clonal en las hojas de *Arabidopsis thaliana*. Algunos de los correspondientes genes codifican factores de transcripción o participan en la regulación de la expresión génica mediada por la cromatina. Otros podrían estar implicados en la degradación de proteínas por el proteasoma.

Para la inducción de sectores mutantes en plantas adultas estamos utilizando dos abordajes distintos: uno basado en la irradiación de líneas CAUT (de *cell autonomy*)[2] con rayos X, y otro basado en la recombinación específica de sitio de transgenes mediada por la recombinasa Cre.

Para el primer abordaje, hemos obtenido líneas homocigóticas para la mutación *chlorata-42* (*ch-42*) y heterocigóticas para las mutaciones letales embrionarias (*emb*)[3] de interés. Cada una de estas líneas fue cruzada por la línea CAUT apropiada, portadora de un transgén con una copia silvestre de *CHLORATA-42,* situada entre el centrómero y el gen *EMB* correspondiente (Figura 1). Con el fin de inducir sectores hemicigóticos, hemos irradiado con rayos X semillas procedentes de líneas portadoras de 13 mutaciones *emb* distintas. En el escrutinio de estas líneas hemos encontrado sectores de color verde pálido y albinos (Figura 2). Ocasionalmente encontramos también plantas que, siendo homocigóticas para una de las mutaciones *emb* a estudio, fueron capaces de crecer (*escapers*; Figura 3).

Para el segundo abordaje, hemos generado construcciones portadoras de una copia silvestre de 17 genes *EMB* distintos, con el fin de complementar la letalidad embrionaria causada por las mutaciones *emb*, adaptando para ello el vector pCB1[4] a la tecnología Gateway. La copia silvestre de cada gen *EMB* de interés está flanqueada en pCB1 por dos sitios *loxP*, haciendo posible la generación de sectores mutantes marcados con GFP cuando se combina en una misma planta la construcción portadora del alelo silvestre de un gen *EMB* dado con otra en la que la recombinasa Cre se expresa bajo el control de un promotor de choque térmico (Figura 4).

MATERIALES Y MÉTODOS

Las plantas se cultivaron a 20±1°C o 26±1°C con una humedad relativa del 60-70% e iluminación continua (5.500 lx). La irradiación de semillas con rayos X se llevó a cabo usando una cabina Philips MG102 a 10 Gy (semillas en agua) y 160 Gy (semillas secas). Se clonó un casete Gateway en el sitio *Not*I del vector pCB1 en las dos orientaciones posibles, a fin de generar los vectores pCB1-Gateway(+) y pCB1-Gateway(-). Se amplificaron insertos grandes de ADN genómico utilizando la polimerasa *Phusion* y cebadores con sitios *attB1* y *attB2* en sus extremos 5'. Los productos así obtenidos se clonaron en un vector de entrada mediante una reacción BP y se subclonaron en los vectores de destino pCB1-Gateway(+) y pCB1-Gateway(-) mediante una reacción LR. Estas construcciones fueron transferidas de *Escherichia coli* DH5α a *Agrobacterium tumefaciens* C58C1, y se utilizaron para infectar plantas de Arabidopsis portadoras de la construcción HS:Cre. Las plantas transformantes se seleccionaron en arena de río suplementada con Basta (15 mg/l).

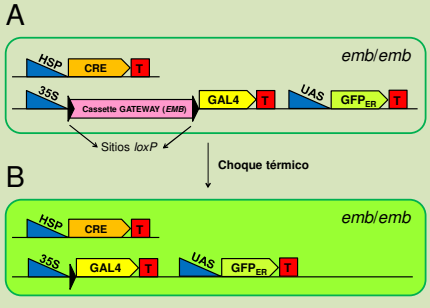


Figura 4.- Estrategia basada en transgenes. (A) Una planta de genotipo mutante *emb/emb* es portadora de dos construcciones distintas, una que aporta una copia silvestre de un gen *EMB* y hace posible su desarrollo normal, y otra con un promotor de choque térmico que controla la expresión de la recombinasa Cre. (B) Un choque térmico provoca la activación de Cre y la pérdida concomitante del alelo silvestre *EMB* en algunas células. La expresión de la GFP marca el linaje de estas últimas, que pueden manifestar el fenotipo mutante causado por la disfunción local del gen *EMB* en la planta adulta.

BIBLIOGRAFÍA

1.- Candela, H., Pérez-Pérez, J.M., y Micol, J.L. (2011). *Trends Plant Sci.* **16**, 336-345.
2.- Furner, I., Ellis, L., Bakht, S., Mirza, B., y Sheikh, M. (2008). *Plant J.* **53,** 645-660.
3.- Meinke, D.W., y Sussex, I.M. (1979). *Dev. Biol.* **72,** 62-72.
4.- Heidstra, R., Welch, D., y Scheres, B. (2004). *Genes Dev.* **18**, 1964-1969.

AGRADECIMIENTOS

La investigación en el laboratorio de J.L.M. está financiada por el Ministerio de Economía y Competitividad [BFU2011-22825 y CSD2007-00057 (TRANSPLANTA)], la Generalitat Valenciana (PROMETEO/2009/112) y la Comisión Europea [LSHG-CT-2006-037704 (AGRON-OMICS)]. H.C. es beneficiario de una Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073). Agradecemos a Renze Heidstra y Guy Wachsman los materiales cedidos y a Ginés Morata el uso de sus instalaciones.

Clonal analysis strategies to uncover the post-embryonic function of embryo-lethal genes in Arabidopsis

Muñoz-Nortes,T., Pérez-Howell, O., Candela, H., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante, Spain

Several hundred genes are required for embryonic and gametophytic development in Arabidopsis, as inferred from the lethality of their mutations. Despite many embryo-lethal genes are normally expressed throughout plant development, the corresponding mutants usually die at early stages, preventing the study of their post-embryonic functions using conventional methods. Clonal analysis techniques provide an effective solution to this problem by uncovering the effects of embryo-lethal mutations in sectors of mutant cells within an otherwise normal adult plant. We selected 35 embryo-lethal mutations for clonal analysis, with a focus on their effects on leaf development. For the induction of mutant sectors in adult plants, we are using two approaches: one based on the X-ray irradiation of so-called 'cell autonomy' lines (CAUT), and another based on the site-specific excision of transgenes mediated by Cre recombinase. To induce hemizygous clonal sectors, we irradiated and screened plants carrying 13 different *emb* mutations. We have also prepared transgenes to complement the embryo-lethality of 17 different *emb* mutants. Each of these transgenes carries a wild-type *EMB* gene flanked by *lox*P sites. The inducible expression of Cre creates mutant sectors that also express the green fluorescent protein. This study will advance our knowledge on the functions of selected *EMB* genes, some of which encode transcription factors and components of the chromatin remodeling machinery.

2013
SEB Annual Meeting
Valencia
Póster

Clonal analysis strategies to uncover the post-embryonic function of embryo-lethal genes in *Arabidopsis thaliana*

Muñoz-Nortes, T., Pérez-Howell, O., Candela, H., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
tmunoz@umh.es hcandela@umh.es jlmicol@umh.es http://genetica.umh.es

As inferred from the lethality of their loss-of-function mutations, several hundred genes in the *Arabidopsis thaliana* genome are necessary for embryonic and gametophytic development. Despite many embryo-lethal genes are normally expressed throughout all stages of plant development, the corresponding mutants die at early stages, preventing the study of their post-embryonic functions using conventional methods. Clonal analysis techniques provide an effective solution to this problem by uncovering the effects of embryo-lethal mutations in sectors of mutant cells within an otherwise normal adult plant[1].

We have selected 35 embryo-lethal mutants of *Arabidopsis thaliana* for clonal analysis experiments, with a focus on their effects on leaf development. Some of the corresponding genes encode transcription factors or participate in the chromatin-mediated regulation of gene expression. Other group of genes have putative roles in proteasome-mediated protein degradation.

For the induction of mutant sectors in adult plants, we are using two different approaches: one based on the X-ray irradiation of so-called 'cell autonomy' (CAUT) lines[2], and another based on the site-specific excision of transgenes mediated by Cre recombinase.

For the first approach, we generated lines that are homozygous for the pale-green *chlorata-42* (*ch-42*) mutation and heterozygous for selected *embryo-defective* (*emb*) mutations[3]. Individual lines were then crossed to appropriate CAUT lines carrying a transgene with a wild-type copy of *CHLORATA-42* located between the centromere and the corresponding *EMB* gene (Figure 1). To induce hemizygous sectors for each *emb* mutation, we have already irradiated (with X-rays) seeds from lines carrying 13 different *emb* mutations. The examination of irradiated lines showed pale-green and albino sectors (Figure 2). In some families, we occasionally found plants that were homozygous for an *emb* mutation but were able to grow despite their embryo-lethality (escapers, Figure 3).

For the second approach, we are generating constructs (carrying a wild-type copy of a given *EMB* gene) to complement the embryonic lethality of the corresponding *emb* mutation. We have adapted the pCB1 vector[4] for use with the Gateway technology. In this vector, the wild-type copy of an *EMB* gene is flanked by two *loxP* sites, making possible to create GFP-marked, mutant sectors when the Cre recombinase is expressed under the control of a heat shock promoter (Figure 4).

We have generated double transgenic lines for 17 different genotypes. Each line contains: (1) a pCB1-Gateway vector carrying a wild-type copy of the corresponding *EMB* gene, (2) the construct HS:Cre, where Cre recombinase expression is controlled by a heat shock promoter.

We have crossed the double transgenics with heterozygous *EMB*/*emb* plants, and we are about to induce the Cre recombinase expression by means of a heat shock in the F_3 generation.

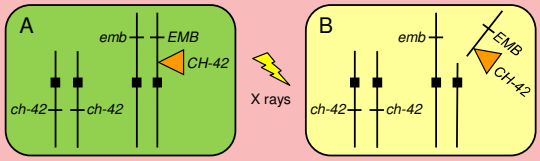


Figure 1.- Approach based on CAUT lines. (A) CAUT lines are homozygous for a recessive yellow (non-photosynthetic) marker, *ch-42*, but they exhibit a wild-type (green) phenotype due to genetic complementation with a wild-type allele supplied by a transgene (*CH-42*). (B) Sectors of yellow cells appear occasionally when the chromosome arm that carries the transgene is lost after a treatment with a high dose of X-rays. In a strain with the appropriate genotype, loss of the chromosome arm can also be used to create a sector that is also hemizygous for an *embryo-defective* mutation (*emb*), which will be marked by the yellow phenotype. Then, sectors have to be screened in order to find mutant phenotypes related to the *emb* mutation.

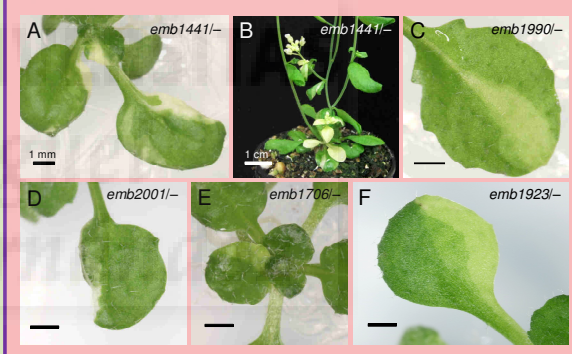


Figure 2.- Irradiation with X-rays induce pale-green or completely albino sectors in several leaves, and these sectors are assumed to be hemizygous for an *emb* mutation. (A-F) Leaves with sectors that are hemizygous for (A, B) *emb1441*, (C) *emb1990*, (D) *emb2001*, (E) *emb1706*, and (F) *emb1923* mutations.

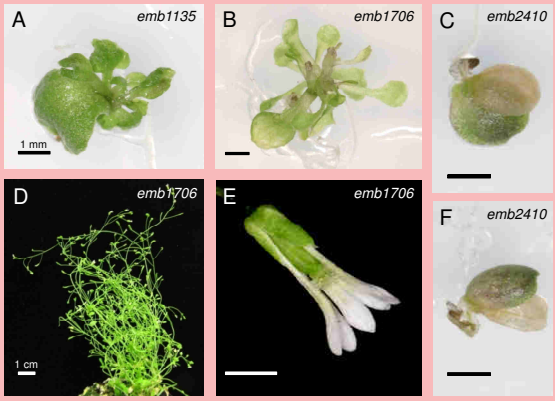


Figure 3.- Escapers are embryo-lethal mutants that survive beyond the embryogenesis and exhibit a postembryonic phenotype putatively associated with the *emb* mutation. (A-F) Escapers found in the screening of irradiated lines that were segregating for the (A) *emb1135* (B, D, E) *emb1706*, and (C, F) *emb2410* mutations.

MATERIALS AND METHODS

Plants were grown at 20±1°C or 26±1°C and 60-70% relative humidity under continuous fluorescent light (5,500 lx). Irradiation of seeds with X-rays was performed using a Philips MG102 cabin and the doses of irradiation were 10 Gy for wet seeds and 160 Gy for dry seeds. We cloned a Gateway cassette in the *Not*I site of pCB1, in both orientations, to generate the pCB1-Gateway(+) and pCB1-Gateway(-) vectors. Large genomic inserts were amplified using *Phusion* polymerase and primers containing *attB1* and *attB2* sites, subsequently cloned into a donor vector using a BP reaction, and finally mobilized into pCB1-Gateway(+) or pCB1-Gateway(-) using an LR reaction (Invitrogen).

REFERENCES

1.- Candela, H., Pérez-Pérez, J.M., and Micol, J.L. (2011). *Trends Plant Sci.* **16**, 336-345.
2.- Furner, I., Ellis, L., Bakht, S., Mirza, B., and Sheikh, M. (2008). *Plant J.* **53,** 645-660.
3.- Meinke, D.W., and Sussex, I.M. (1979b). *Dev. Biol.* **72,** 62-72.
4.- Heidstra, R., Welch, D., and Scheres, B. (2004). *Genes Dev.* **18**, 1964-1969.

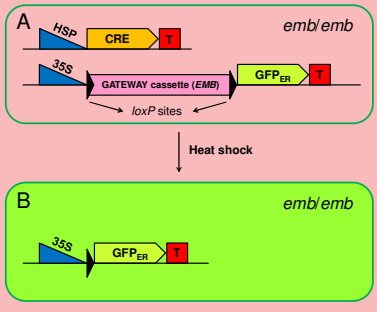


Figure 4.- Transgene-based approach. (A) An homozygous *emb*/*emb* plant carries two different cons-tructs, one of them providing a wild-type copy of an *EMB* gene that allows its normal development, and the other with an heat-shock promoter driving the inducible expression of Cre recombinase. (B) A heat pulse causes the activation of Cre and a concomitant loss of the wild-type copy of the *EMB* gene in some cells. Then, the expression of GFP under the 35S promoter marks the lineage of those cells, which might exhibit any mutant phenotype that is associated with the loss of function of the *EMB* gene in adult tissues.

ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M is supported by grants BFU2011-22825 and CSD2007-00057 (TRANSPLANTA) from the Ministerio de Ciencia e Innovación of Spain, PROMETEO/2009/112 from the Generalitat Valenciana and LSHG-CT-2006-037704 (AGRON-OMICS) from the European Commission. We want to thank Ginés Morata, Renze Heidstra and Guy Wachsman for sharing their materials and facilities. H.C. is the recipient of a Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073). T.M. is the recipient of a VALi+d fellowship from the Generalitat Valenciana (ACIF/2013/273).

Using clonal analysis approaches to study Arabidopsis embryo-lethal genes in post-embryonic stages

Muñoz-Nortes, T., Lázaro-Frías, A., Candela, H., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante, Spain

Several hundred genes are required for embryonic and gametophytic development in Arabidopsis, as inferred from the lethality of their mutations. Despite many embryo-lethal genes are normally expressed throughout plant development, the corresponding mutants usually die at early stages, preventing the study of their post-embryonic functions using conventional methods. Clonal analysis techniques provide an effective solution to this problem by uncovering the effects of embryo-lethal mutations in sectors of mutant cells within an otherwise normal adult plant. We selected 35 embryo-lethal mutations for clonal analysis, with a focus on their effects on leaf development. For the induction of mutant sectors in adult plants, we are using two approaches: one based on the X-ray irradiation of so-called 'cell autonomy' lines (CAUT), and another based on the site-specific excision of transgenes mediated by Cre recombinase. To induce hemizygous clonal sectors, we irradiated and screened plants carrying 13 different *emb* mutations. We have also prepared transgenes to complement the embryo-lethality of 17 different *emb* mutants. Each of these transgenes carries a wild-type *EMB* gene flanked by *lox*P sites. The inducible expression of Cre creates mutant sectors that also express the green fluorescent protein. This study will advance our knowledge on the functions of selected *EMB* genes, some of which encode transcription factors and components of the chromatin remodeling machinery.

2014
XII Reunión Biología Molecular de Plantas
Cartagena
Póster

Using clonal analysis approaches to study Arabidopsis embryo-lethal genes in post-embryonic stages

Muñoz-Nortes, T., Lázaro-Frías, A., Candela, H., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
tmunoz@umh.es hcandela@umh.es jlmicol@umh.es http://genetics.umh.es

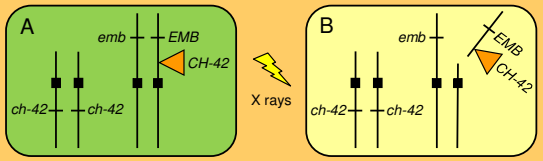


Figure 1.- Approach based on CAUT lines. (A) CAUT lines are homozygous for a recessive yellow (non-photosynthetic) marker, *ch-42*, but they exhibit a wild-type (green) phenotype due to genetic complementation with a wild-type allele supplied by a transgene (*CH-42*). (B) Sectors of yellow cells appear occasionally when the chromosome arm that carries the transgene is lost after a treatment with a high dose of X-rays. In a strain with the appropriate genotype, loss of the chromosome arm can also be used to create a sector that is also hemizygous for an *embryo-defective* mutation (*emb*), which will be marked by the yellow phenotype. Then, sectors have to be screened in order to find mutant phenotypes related to the *emb* mutation.

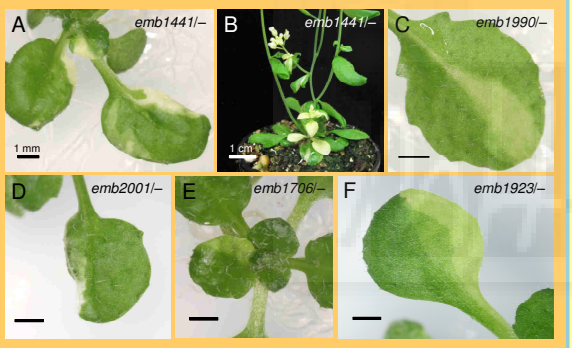


Figure 2.- Irradiation with X-rays induce pale-green or completely albino sectors in several leaves, and these sectors are assumed to be hemizygous for an *emb* mutation. (A-F) Leaves with sectors that are hemizygous for (A, B) *emb1441*, (C) *emb1990*, (D) *emb2001*, (E) *emb1706*, and (F) *emb1923* mutations.

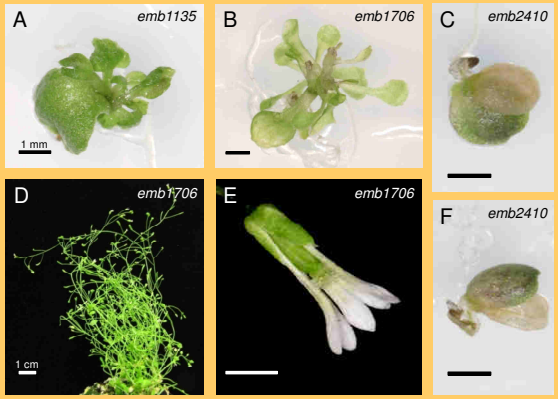


Figure 3.- Escapers are embryo-lethal mutants that survive beyond the embryogenesis and exhibit a postembryonic phenotype putatively associated with the *emb* mutation. (A-F) Escapers found in the screening of irradiated lines that were segregating for the (A) *emb1135* (B, D, E) *emb1706*, and (C, F) *emb2410* mutations.

As inferred from the lethality of their loss-of-function mutations, several hundred genes in the *Arabidopsis thaliana* genome are necessary for embryonic and gametophytic development. Despite many embryo-lethal genes are normally expressed throughout all stages of plant development, the corresponding mutants die at early stages, preventing the study of their post-embryonic functions using conventional methods. Clonal analysis techniques provide an effective solution to this problem by uncovering the effects of embryo-lethal mutations in sectors of mutant cells within an otherwise normal adult plant[1].

We have selected 35 embryo-lethal mutants of *Arabidopsis thaliana* for clonal analysis experiments, with a focus on their effects on leaf development. Some of the corresponding genes encode transcription factors or participate in the chromatin-mediated regulation of gene expression. Other group of genes have putative roles in proteasome-mediated protein degradation.

For the induction of mutant sectors in adult plants, we are using two different approaches: one based on the X-ray irradiation of so-called 'cell autonomy' (CAUT) lines[2], and another based on the site-specific excision of transgenes mediated by Cre recombinase.

For the first approach, we generated lines that are homozygous for the pale-green *chlorata-42* (*ch-42*) mutation and heterozygous for selected *embryo-defective* (*emb*) mutations[3]. Individual lines were then crossed to appropriate CAUT lines carrying a transgene with a wild-type copy of *CHLORATA-42* located between the centromere and the corresponding *EMB* gene (Figure 1). To induce hemizygous sectors for each *emb* mutation, we have already irradiated (with X-rays) seeds from lines carrying 13 different *emb* mutations. The examination of irradiated lines showed pale-green and albino sectors (Figure 2). In some families, we occasionally found plants that were homozygous for an *emb* mutation but were able to grow despite their embryo-lethality (escapers, Figure 3).

For the second approach, we are generating constructs (carrying a wild-type copy of a given *EMB* gene) to complement the embryonic lethality of the corresponding *emb* mutation. We have adapted the pCB1 vector[4] for use with the Gateway technology. In this vector, the wild-type copy of an *EMB* gene is flanked by two *loxP* sites, making possible to create GFP-marked, mutant sectors when the Cre recombinase is expressed under the control of a heat shock promoter (Figure 4).

We have generated double transgenic lines for 17 different genotypes. Each line contains: (1) a pCB1-Gateway vector carrying a wild-type copy of the corresponding *EMB* gene, (2) the construct HS:Cre, where Cre recombinase expression is controlled by a heat shock promoter.

We have crossed the double transgenics with heterozygous *EMB*/*emb* plants, and we are about to induce the Cre recombinase expression by means of a heat shock in the F_3 generation.

MATERIALS AND METHODS

Plants were grown at 20±1°C or 26±1°C and 60-70% relative humidity under continuous fluorescent light (5,500 lx). Irradiation of seeds with X-rays was performed using a Philips MG102 cabin and the doses of irradiation were 10 Gy for wet seeds and 160 Gy for dry seeds. We cloned a Gateway cassette in the *Not*I site of pCB1, in both orientations, to generate the pCB1-Gateway(+) and pCB1-Gateway(-) vectors. Large genomic inserts were amplified using *Phusion* polymerase and primers containing *attB1* and *attB2* sites, subsequently cloned into a donor vector using a BP reaction, and finally mobilized into pCB1-Gateway(+) or pCB1-Gateway(-) using an LR reaction (Invitrogen).

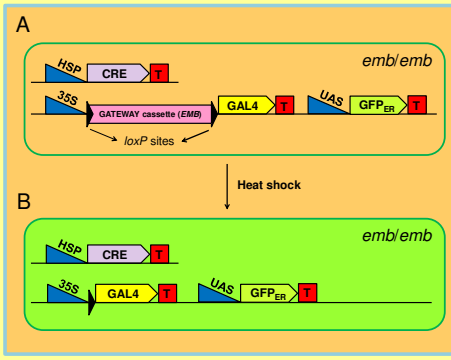


Figure 4.- Transgene-based approach. (A) A homozygous *emb*/*emb* plant carries two different constructs, one of them providing a wild-type copy of an *EMB* gene that allows its normal development, and the other with a heat-shock promoter driving the inducible expression of Cre recombinase. (B) A heat pulse causes the activation of Cre and a concomitant loss of the wild-type copy of the *EMB* gene in some cells. Then, the expression of GFP under the 35S promoter marks the lineage of those cells, which might exhibit any mutant phenotype that is associated with the loss of function of the *EMB* gene in adult tissues.

REFERENCES

1.- Candela, H., Pérez-Pérez, J.M., and Micol, J.L. (2011). *Trends Plant Sci.* **16**, 336-345.
2.- Furner, I., Ellis, L., Bakht, S., Mirza, B., and Sheikh, M. (2008). *Plant J.* **53,** 645-660.
3.- Meinke, D.W., and Sussex, I.M. (1979). *Dev. Biol.* **72,** 62-72.
4.- Heidstra, R., Welch, D., and Scheres, B. (2004). *Genes Dev.* **18**, 1964-1969.

ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M is supported by grants BFU2011-22825 and CSD2007-00057 (TRANSPLANTA) from the Ministerio de Ciencia e Innovación of Spain, PROMETEO/2009/112 from the Generalitat Valenciana and LSHG-CT-2006-037704 (AGRON-OMICS) from the European Commission. We want to thank Ginés Morata, Renze Heidstra and Guy Wachsman for sharing their materials and facilities. T.M. is the recipient of a VALi+d fellowship from the Generalitat Valenciana (ACIF/2013/273).

The *ANGULATA7* essential gene is required for leaf development

Tamara Muñoz-Nortes, Alejandro Ruiz-Bayón, José Manuel Pérez-Pérez, María Rosa Ponce, Héctor Candela and José Luis Micol

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante, Spain

In a large-scale screen for EMS-induced mutants with abnormal leaf shape, we isolated the *angulata7-1* (*anu7-1*) recessive mutant, which exhibits pale-green leaves with prominent marginal teeth. *anu7-1* vegetative leaves are small and exhibit reduced chlorophyll and carotenoid levels compared to wild type. Map-based cloning allowed us to identify the causal mutation of the phenotype of *anu7-1*: a G-to-A transition that is predicted to cause a G-to-E amino acid substitution in ANU7, a chloroplast-localized protein of unknown function. A 35_{pro}*:ANU7* transgene fully complemented the mutant phenotype of *anu7-1*. We also identified two insertional, embryo-lethal alleles: *anu7-2* and *anu7-3*. A number of nuclear genes encoding transcription factors were found deregulated in a microarray analysis of *anu7-1* RNA, suggesting a role for ANU7 in retrograde signalling.

2015
26th International Conference on Arabidopsis Research
París
Póster

The *ANGULATA7* essential gene is required for leaf development

Tamara Muñoz-Nortes, Alejandro Ruiz-Bayón, José Manuel Pérez-Pérez, María Rosa Ponce, Héctor Candela and José Luis Micol

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
tmunoz@umh.es jlmicol@umh.es http://genetics.umh.es

In a large-scale screen for EMS-induced mutants with abnormal leaf shape, twenty mutant lines of the Angulata (Anu) phenotypic class were isolated and assigned to twelve different complementation groups[1]. We already identified five *ANU* genes: *ANU1*, *ANU4*, *ANU9*, *ANU10* and *ANU12*; all of them were found to encode plastid-localized proteins involved in diverse processes[2,3].

Plants homozygous for the *angulata7-1* (*anu7-1*) mutation (Fig. 1A) exhibit pale-green leaves with prominent marginal teeth, the characteristic traits of the Anu class (Fig. 1B, C). *anu7-1* plants are small and have reduced chlorophyll and carotenoid levels compared to L*er* (Fig. 1B, F, G). Palisade cells in *anu7-1* leaves are larger and less densely packed than in the wild type, resulting in bigger intercellular spaces (Fig. 2).

Following a map-based cloning approach, we identified the causal mutation of the phenotype of the *anu7-1* mutant: a G→A transition that is predicted to cause a glycine→glutamate amino acid substitution in a putative plastid-localized protein of unknown function (Fig. 1A). The identity of *ANU7* was confirmed by phenotypic rescue of the *anu7-1* mutant and reconfirmed by allelism tests involving *anu7-1* and two insertional lines, which we dubbed a*nu7-2* and *anu7-3* and have been reported to be embryo-lethal[4,5]. Despite their embryonic lethality, we were able to grow *anu7-2* plants in GM medium supplemented with sucrose. In these conditions, *anu7-2* plants were capable to complete their life cycle, exhibiting an albino, dwarf phenotype (Fig. 1E). The expression pattern of the *ANU7* gene was analyzed using *ANU7*$_{pro}$*:GUS* transgenic plants: we found the reporter gene expressed in all the tissues studied (Fig. 3). We are presently characterizing transgenic plants to determine the subcellular localization of the ANU7 protein by means of translational fusions with different fluorescent proteins.

We performed a microarray analysis of *anu7-1* RNA. The results show that the expression of several nuclear genes encoding transcription factors is deregulated, suggesting a possible role for *ANU7* in retrograde signaling[6].

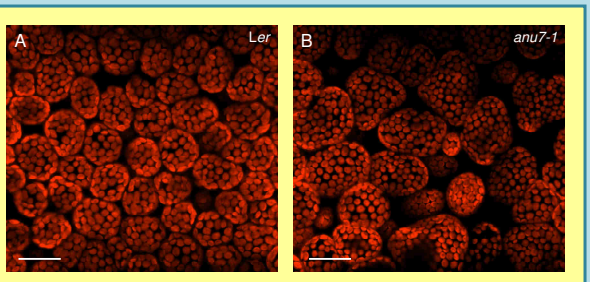


Figure 2.- Confocal micrographs of the subepidermal layer of mesophyll cells from (A) L*er* and (B) *anu7-1*, showing chlorophyll autofluorescence of chloroplasts. Pictures were taken from third-node leaves, collected 20 das. Scale bars: 30 µm.

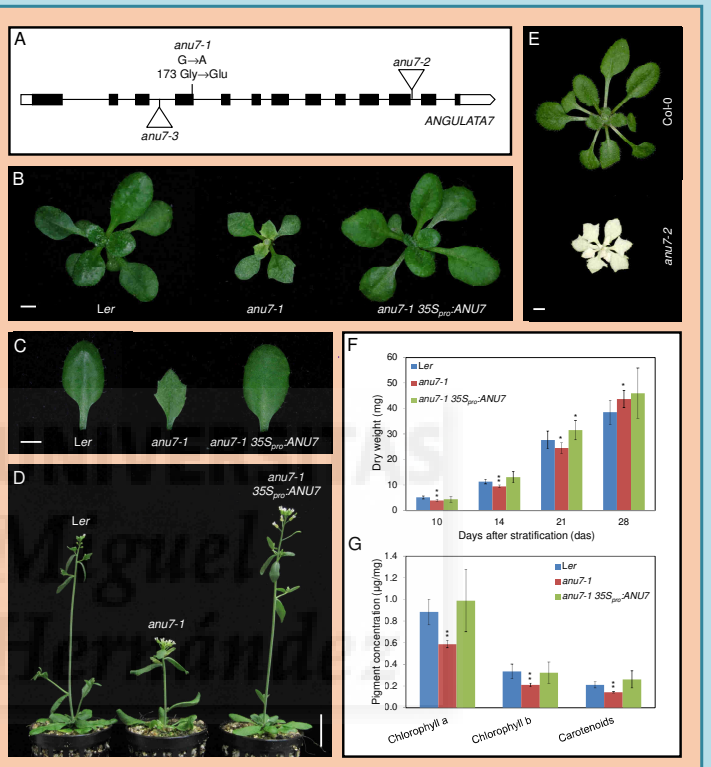


Figure 1.- Structure of the *ANU7* gene and phenotype of the *anu7* mutants. (A) Scheme of the *ANU7* gene, indicating the nature and position of *anu7* mutations. Boxes and lines indicate exons and introns, respectively. White boxes represent UTRs. Triangles indicate T-DNA insertions. (B) Rosettes, (C) third-node leaves and (D) adult plants of L*er*, *anu7-1* and the transgenic line *anu7-1 35S*$_{pro}$*:ANU7*. (E) Rosettes of Col-0 and *anu7-2*. (F) Dry weight variation throughout vegetative development, and (G) pigment content determined 12 days after stratification (das) in L*er*, *anu7-1* and *anu7-1 35S*$_{pro}$*:ANU7* transgenic plants. Pictures were taken (B, C) 16 (E) 21 and (D) 33 das. Scale bars: 2 mm. Asterisks indicate values significantly different from those of L*er* in a Mann-Whitney U test: *p<0.05, and **p<0.01.

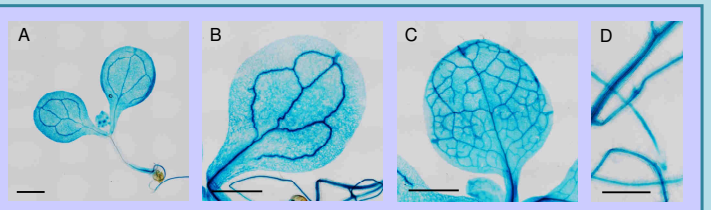


Figure 3.- Visualization of *ANU7*$_{pro}$*:GUS* transgene activity in a L*er* background. (A) Seedling, (B) cotyledon , (C) first vegetative leaf and (D) roots. Pictures were taken (A) 7 and (B, C, D) 12 das. Scale bars: (A, B, C) 1 mm, and (D) 250 µm.

REFERENCES

1.- Berná, G., Robles, P., and Micol, J.L. (1999). A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. *Genetics* **152,** 729-742.

2.- Casanova-Sáez, R., Mateo-Bonmatí, E., Kangasjärvi, S., Candela, H., and Micol., J.L. (2014). *Arabidopsis* ANGULATA10 is required for thylakoid biogenesis and mesophyll development. *J Exp Bot* **65,** 2391-2404.

3.- Mateo-Bonmatí, E., Casanova-Sáez, R., Candela, H., and Micol., J.L. (2014). Rapid identification of *angulata* leaf mutations using next-generation sequencing. *Planta* **240,** 1113-1122.

4.- Tzafrir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., David Patton, D., and Meinke, D. (2003). The *Arabidopsis* SeedGenes Project. *Nucl Acids Res* **31,** 90-93.

5.- Meinke, D., Sweeney, C., and 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.

6.- Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nature Rev Mol Cell Biol* **14,** 787–802.

ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M. is supported by grants from the Ministerio de Economía y Competitividad of Spain (BIO2014-53063-P), and the Generalitat Valenciana (PROMETEOII/2014/006). T.M.N. holds a VALi+d fellowship from the Generalitat Valenciana (ACIF/2013/273).

The DnaJ-like ANGULATA7 protein is required for plastid gene expression and thylakoidal membrane organization in Arabidopsis

Tamara Muñoz-Nortes, Tamara González-Costa, María Rosa Ponce,
Héctor Candela, José Luis Micol

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Spain

The characterization of mutants with altered leaf shape and pigmentation has previously allowed the identification of nuclear genes that encode plastid-localized proteins that perform essential functions in leaf growth and development[1]. A large-scale screen previously allowed us to isolate ethyl methanesulfonate (EMS) induced mutants with small rosettes and pale green leaves with prominent marginal teeth, which were assigned to a phenotypic class that we dubbed Angulata[2]. The molecular characterization of the twelve genes assigned to this phenotypic class should help us to advance our understanding of the still poorly understood relationship between chloroplast biogenesis and leaf morphogenesis[3].

Here we report the phenotypic and molecular characterization of the *angulata7-1* (*anu7-1*) mutant of Arabidopsis, which we found to carry a novel hypomorphic allele of the *EMB2737* gene, which was previously known only for its embryonic lethal mutations. *ANU7* encodes a plant-specific protein containing a domain with conserved cysteine and glycine residues that is similar to an incomplete central cysteine-rich domain, which accounts for the disulfide isomerase activity of DnaJ proteins. DnaJ proteins normally function as chaperones, either alone or in combination with heat-shock protein 70, and have been proposed to participate in the folding, unfolding, assembly and degradation of proteins, maintaining protein homeostasis under normal or stress conditions[4].

Although the complete loss of *ANU7* function causes embryonic lethality, our EMS-induced alleles are hypomorphic and viable, and have allowed us to demonstrate that *ANU7* is required for the accumulation of photosynthetic pigments and the correct organization of the thylakoid membrane system. The observed genetic interaction of *anu7-1* with a loss-of-function allele of *GENOMES UNCOUPLED1*[5] suggests that the *anu7-1* mutation triggers a retrograde signal that is at least in part responsible for the observed phenotypic defects. Our microarray expression studies show that many genes

that normally function in the chloroplasts are upregulated in *anu7-1* rosettes, with a significant overrepresentation of those required for the expression plastid genome genes.

1.- Micol, J.L. (2009). *Current Opinion in Plant Biology* 12, 9-16.

2.- Berná, G., Robles, P., and Micol, J.L. (1999). *Genetics* 152, 729-742.

3.- Casanova-Sáez, R., Mateo-Bonmatí, E., Kangasjärvi, S., Candela, H., and Micol, J.L. (2014). *Journal of Experimental Botany* 65, 2391-2404.

4.- Craig, E.A., Huang, P., Aron, R., and Andrew, A. (2006). *Reviews of Physiology, Biochemistry and Pharmacology* 156, 1-21.

5.- Cottage, A., Mott, E.K., Kempster, J.A., and Gray, J.C. (2010). *Journal of Experimental Botany* 61, 3773-3786.

2016
XIII Reunión de Biología Molecular de Plantas
Oviedo
Póster

The DnaJ-like ANGULATA7 protein is required for plastid gene expression and thylakoidal membrane organization in Arabidopsis

Tamara Muñoz-Nortes, Tamara González-Costa, María Rosa Ponce, Héctor Candela and José Luis Micol

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
tmunoz@umh.es jlmicol@umh.es http://genetics.umh.es

The characterization of mutants with altered leaf shape and pigmentation has previously allowed the identification of nuclear genes that encode plastid-localized proteins that perform essential functions in leaf growth and development[1]. A large-scale screen previously allowed us to isolate ethyl methanesulfonate (EMS)-induced mutants with small rosettes and pale green leaves with prominent marginal teeth, which were assigned to a phenotypic class that we dubbed Angulata[2]. The molecular characterization of the twelve genes assigned to this phenotypic class should help us to advance our understanding of the still poorly understood relationship between chloroplast biogenesis and leaf morphogenesis[3,4].

Here we report the phenotypic and molecular characterization of the *angulata7-1* (*anu7-1*) mutant of Arabidopsis, which we found to carry a novel hypomorphic allele of the *EMB2737* gene, previously known only for its embryonic lethal mutations (Fig. 2-4, 6, 7). *ANU7* encodes a plant-specific protein containing a domain with conserved cysteine and glycine residues that is similar to an incomplete central cysteine-rich domain (Fig. 1), which accounts for the disulfide isomerase activity of DnaJ proteins. DnaJ proteins normally function as chaperones, either alone or in combination with heat-shock protein 70, and have been proposed to participate in the folding, unfolding, assembly and degradation of proteins, maintaining protein homeostasis under normal or stress conditions[5].

Although the complete loss of *ANU7* function causes embryonic lethality, our EMS-induced allele is hypomorphic and viable, and has allowed us to demonstrate that *ANU7* is required for the accumulation of photosynthetic pigments and the correct organization of the thylakoid membrane system (Fig. 4). The observed genetic interaction of *anu7-1* with a loss-of-function allele of *GENOMES UNCOUPLED1*[6] suggests that the *anu7-1* mutation triggers a retrograde signal that is at least in part responsible for the observed phenotypic defects (Fig. 5). Our microarray and qRT-PCR expression studies show that many genes that normally function in the chloroplasts are upregulated in *anu7-1* rosettes, with a significant overrepresentation of those required for the expression of plastid genome genes such as many subunits of plastid transcriptionally active chromosome complexes (pTAC; Fig. 2).

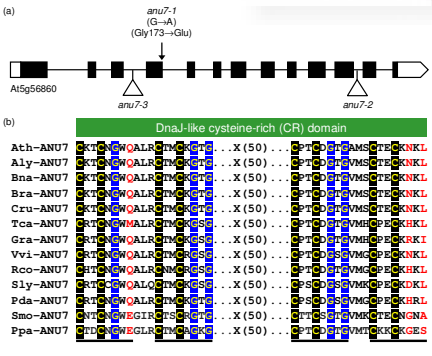


Figure 1.- Structure of the *ANU7* gene and conservation of the ANU7 protein. (a) Structure of *ANU7*, with indication of the nature and position of the *anu7* mutations. Boxes and lines between boxes indicate exons and introns, respectively. White boxes represent the 5'- and 3'-UTRs. Triangles indicate T-DNA insertions. (b) Detail of an alignment of the ANU7 protein and some of its orthologs, showing the DnaJ-like CR domain. The alignment was generated with T-Coffee. The four zinc binding motifs of the DnaJ-like CR domain are underlined in black. Conserved cysteines and glycines are shaded black and blue, respectively. Plant species used for the multiple sequence alignment: Ath, *Arabidopsis thaliana*; Aly, *Arabidopsis lyrata*; Bna, *Brassica napus*; Bra, *Brassica rapa*; Cru, *Capsella rubella*; Tca, *Theobroma cacao*; Gra, *Gossypium raimondii*; Vvi, *Vitis vinifera*; Rco, *Ricinus communis*; Sly, *Solanum lycopersicum*; Pda, *Phoenix dactylifera*; Smo, *Selaginella moellendorffii*; and Ppa, *Physcomitrella patens*.

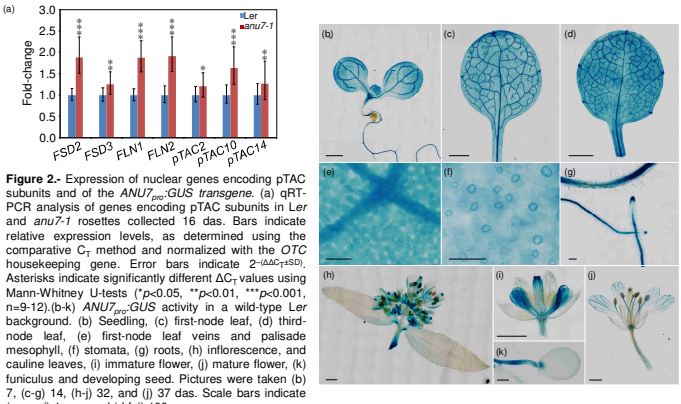


Figure 2.- Expression of nuclear genes encoding pTAC subunits and of the *ANU7*$_{pro}$*:GUS transgene*. (a) qRT-PCR analysis of genes encoding pTAC subunits in *Ler* and *anu7-1* rosettes collected 16 das. Bars indicate relative expression levels, as determined using the comparative C_T method and normalized with the *OTC* housekeeping gene. Error bars indicate $2^{-(\Delta\Delta C_T \pm SD)}$. Asterisks indicate significantly different ΔC_T values using Mann-Whitney U-tests (*p<0.05, **p<0.01, ***p<0.001, n=9-12).(b-k) *ANU7*$_{pro}$*:GUS* activity in a wild-type *Ler* background. (b) Seedling, (c) first-node leaf, (d) third-node leaf, (e) first-node leaf veins and palisade mesophyll, (f) stomata, (g) roots, (h) inflorescence, and cauline leaves, (i) immature flower, (j) mature flower, (k) funiculus and developing seed. Pictures were taken (b) 7, (c-g) 14, (h-j) 32, and (j) 37 das. Scale bars indicate (a-c, g-i) 1 mm, and (d-f, j) 100 µm.

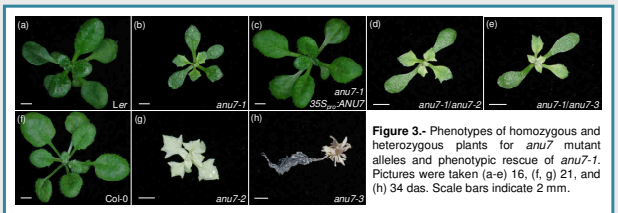


Figure 3.- Phenotypes of homozygous and heterozygous plants for *anu7* mutant alleles and phenotypic rescue of *anu7-1*. Pictures were taken (a-e) 16, (f, g) 21, and (h) 34 das. Scale bars indicate 2 mm.

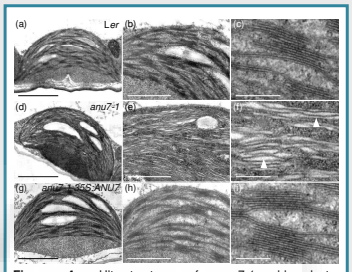


Figure 4.- Ultrastructure of *anu7-1* chloroplasts. Transmission electron micrographs of palisade mesophyll cell chloroplasts from (a-c) *Ler*, (d-f) *anu7-1*, and (g-i) *anu7-1 35S$_{pro}$:ANU7* plants. Pictures were taken from third-node leaves collected 21 das. Scale bars indicate (a, d, g) 2 µm, (b, e, h) 1 µm, and (c, f, i) 0.25 µm. Arrowheads in (f) indicate unappressed thylakoidal membranes in the *anu7-1* mutant.

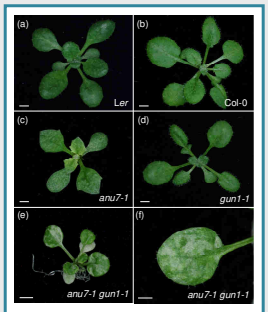


Figure 5.- Leaf phenotype of the *anu7-1 gun1-1* double mutant. Pictures were taken 16 das. Scale bars indicate (a-e) 2 mm, and (f) 1 mm.

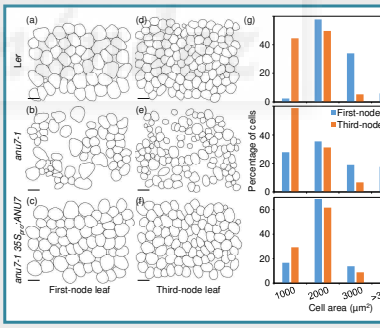


Figure 6.- Morphometry of *anu7-1* palisade mesophyll cells from *Ler*, *anu7-1*, and *anu7-1 35S$_{pro}$:ANU7* plants. (a-f) Representative diagrams of the subepidermal layer of palisade mesophyll cells from (a-c) first- and (d-f) third-node leaves, obtained by drawing the outlines of the cells from micrographs of cleared leaves. (g) Distribution of palisade mesophyll cell sizes (n≥600) in first and third-node leaves. Cell sizes were distributed in four groups. Bars indicate the percentage of cells belonging to each group.

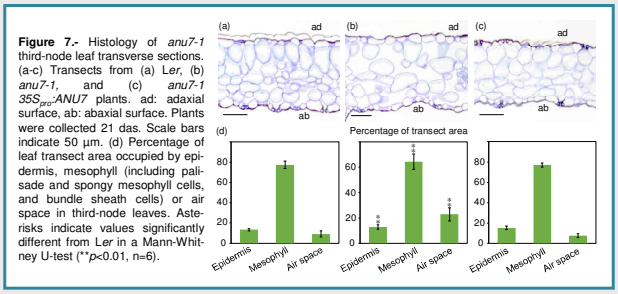


Figure 7.- Histology of *anu7-1* third-node leaf transverse sections. (a-c) Transects from (a) *Ler*, (b) *anu7-1*, and (c) *anu7-1 35S$_{pro}$:ANU7* plants. ad: adaxial surface, ab: abaxial surface. Plants were collected 21 das. Scale bars indicate 50 µm. (d) Percentage of leaf transect area occupied by epidermis, mesophyll (including palisade and spongy mesophyll cells), or bundle sheath cells) or air space in third-node leaves. Asterisks indicate values significantly different from *Ler* in a Mann-Whitney U-test (**p<0.01, n=6).

REFERENCES

1.- Micol, J.L. (2009). *Current Opinion in Plant Biology* **12,** 9-16.
2.- Berná, G., Robles, P., and Micol, J.L. (1999). *Genetics* **152,** 729-742.
3.- Casanova-Sáez, R., Mateo-Bonmatí, E., Kangasjärvi, S., Candela, H., and Micol, J.L. (2014). *Journal of Experimental Botany* **65,** 2391-2404.
4.- Mateo-Bonmatí, E., Casanova-Sáez, R., Candela, H., and Micol, J.L. (2014). *Planta* **240,** 1113-1122.
5.- Craig, E.A., Huang, P., Aron, R., and Andrew, A. (2006). *Reviews of Physiology, Biochemistry and Pharmacology* **156,** 1-21.
6.- Cottage, A., Mott, E.K., Kempster, J.A., and Gray, J.C. (2010). *Journal of Experimental Botany* **61,** 3773-3786.

ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M. is supported by grants from the Ministerio de Economía y Competitividad of Spain (BIO2014-53063-P), and the Generalitat Valenciana (PROMETEOII/2014/006). T.M.N. held a VALi+d fellowship from the Generalitat Valenciana (ACIF/2013/273). H.C. held a Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073).

The *ANGULATA7* gene encodes a DnaJ-like zinc-finger-domain protein involved in chloroplast function and leaf development in Arabidopsis

Muñoz-Nortes, T., Pérez-Pérez, J.M., Ponce, M.R., Candela, H., and Micol, J.L.

Instituto de Bioingeniería, Universidad Miguel Hernández,
Campus de Elche, 03202 Elche, Alicante, Spain

The characterization of mutants with altered leaf shape and pigmentation has allowed the identification of nuclear genes that encode plastid-localized proteins with essential functions in leaf growth and development. A large-scale screen previously allowed us to isolate ethyl methanesulfonate (EMS)-induced mutants with small rosettes and pale green leaves with prominent marginal teeth, which were assigned to a phenotypic class that we dubbed Angulata. The molecular characterization of the *ANGULATA* genes should help us to advance our understanding of the relationship between chloroplast biogenesis and leaf morphogenesis.

Here we report the phenotypic and molecular characterization of the *angulata7-1* (*anu7-1*) mutant of Arabidopsis, which we found to be a hypomorphic allele of the *EMB2737* gene, previously known only for its embryonic-lethal mutations. *ANU7* encodes a plant-specific protein containing a domain similar to the central cysteine-rich (CR) domain of DnaJ proteins. DnaJ proteins normally function as chaperones, either alone or in combination with heat-shock protein 70, and have been proposed to participate in the folding, unfolding, assembly and degradation of other proteins.

We have found that ANU7 is necessary for the accumulation of photosynthetic pigments and the correct organization of the thylakoid membrane system. Our microarray and qRT-PCR expression studies show that many genes which normally function in the chloroplasts are upregulated in *anu7-1* rosettes, with a significant overrepresentation of those required for the expression of plastid genome genes such as many subunits of plastid transcriptionally active chromosome complexes (pTAC). The synergistic interaction between the *anu7-1* mutation and a loss-of-function mutation of *GENOMES UNCOUPLED1* (*GUN1*) points to a functional relationship between *ANU7* and *GUN1*, both of which have been isolated in the nucleoid fraction of plastids.

2017
Plant Organ Growth Symposium
Elche
Póster

The *ANGULATA7* gene encodes a DnaJ-like zinc finger domain protein involved in chloroplast function and leaf development in Arabidopsis

Tamara Muñoz-Nortes, José Manuel Pérez-Pérez, María Rosa Ponce, Héctor Candela and José Luis Micol

Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Alicante, Spain.
tmunoz@umh.es jlmicol@umh.es http://genetics.umh.es

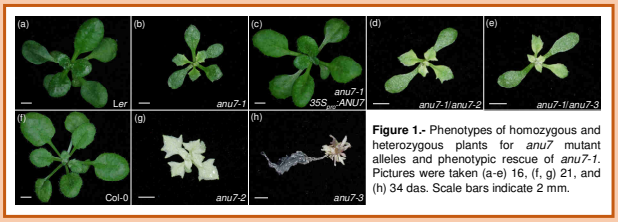


Figure 1.- Phenotypes of homozygous and heterozygous plants for *anu7* mutant alleles and phenotypic rescue of *anu7-1*. Pictures were taken (a-e) 16, (f, g) 21, and (h) 34 das. Scale bars indicate 2 mm.

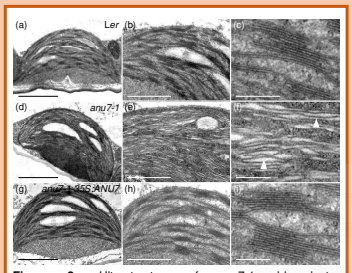


Figure 2.- Ultrastructure of *anu7-1* chloroplasts. Transmission electron micrographs of palisade mesophyll cell chloroplasts from (a-c) *Ler*, (d-f) *anu7-1*, and (g-i) *anu7-1 35S$_{pro}$:ANU7* plants. Pictures were taken from third-node leaves collected 21 das. Scale bars indicate (a, d, g) 2 µm, (b, e, h) 1 µm, and (c, f, i) 0.25 µm. Arrowheads in (f) indicate unappressed thylakoidal membranes in the *anu7-1* mutant.

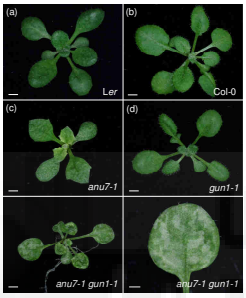


Figure 3.- Leaf phenotype of the *anu7-1 gun1-1* double mutant. Pictures were taken 16 das. Scale bars indicate (a-e) 2 mm, and (f) 1 mm.

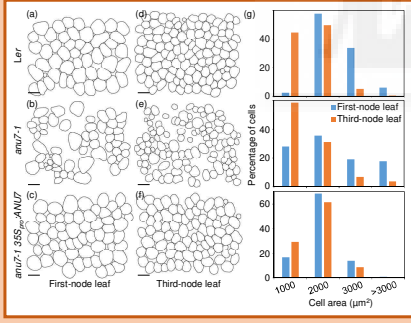


Figure 4.- Morphometry of *anu7-1* palisade mesophyll cells from *Ler*, *anu7-1*, and *anu7-1 35S$_{pro}$:ANU7* plants. (a-f) Representative diagrams of the subepidermal layer of palisade mesophyll cells from (a-c) first- and (d-f) third-node leaves, obtained by drawing the outlines of the cells from micrographs of cleared leaves. (g) Distribution of palisade mesophyll cell sizes (n≥600) in first and third-node leaves. Cell sizes were distributed in four groups. Bars indicate the percentage of cells belonging to each group.

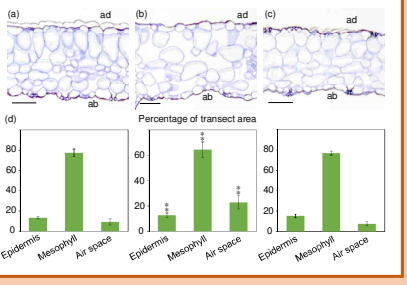


Figure 5.- Histology of *anu7-1* third-node leaf transverse sections. (a-c) Transects from (a) *Ler*, (b) *anu7-1*, and (c) *anu7-1 35S$_{pro}$:ANU7* plants. ad: adaxial surface, ab: abaxial surface. Plants were collected 21 das. Scale bars indicate 50 µm. (d) Percentage of leaf transect area occupied by epidermis, mesophyll (including palisade and spongy mesophyll cells, and bundle sheath cells) or air space in third-node leaves. Asterisks indicate values significantly different from *Ler* in a Mann-Whitney U-test (**p<0.01, n=6).

The characterization of mutants with altered leaf shape and pigmentation has allowed the identification of nuclear genes that encode plastid-localized proteins with essential functions in leaf growth and development[1]. A large-scale screen previously allowed us to isolate ethyl methanesulfonate (EMS)-induced mutants with small rosettes and pale green leaves with prominent marginal teeth, which were assigned to a phenotypic class that we dubbed Angulata[2]. The molecular characterization of the *ANGULATA* genes should help us to advance our understanding of the relationship between chloroplast biogenesis and leaf morphogenesis[3,4].

Here we report the phenotypic and molecular characterization of the *angulata7-1* (*anu7-1*) mutant of Arabidopsis, which we found to be a hypomorphic allele of the *EMB2737* gene, previously known only for its embryonic-lethal mutations (Fig. 1, 2, 4-7). *ANU7* encodes a plant-specific protein containing a domain similar to the central cysteine-rich (CR) domain of DnaJ proteins (Fig. 6)[5]. DnaJ proteins normally function as chaperones, either alone or in combination with heat-shock protein 70, and have been proposed to participate in the folding, unfolding, assembly and degradation of other proteins[6].

We have found that ANU7 is necessary for the accumulation of photosynthetic pigments and the correct organization of the thylakoid membrane system (Fig. 2). Our microarray and qRT-PCR expression studies show that many genes which normally function in the chloroplasts are upregulated in *anu7-1* rosettes, with a significant overrepresentation of those required for the expression of plastid genome genes such as many subunits of plastid transcriptionally active chromosome complexes (pTAC; Fig. 7). The synergistic interaction between the *anu7-1* mutation and a loss-of-function mutation of *GENOMES UNCOUPLED1* (*GUN1*)[7] points to a functional relationship between *ANU7* and *GUN1* (Fig. 3), both of which have been isolated in the nucleoid fraction of plastids[8-9].

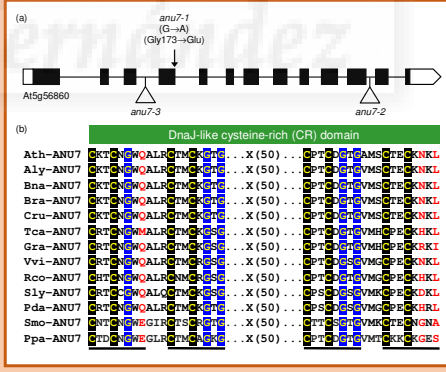


Figure 6.- Structure of the *ANU7* gene and conservation of the ANU7 protein. (a) Structure of *ANU7*, with indication of the nature and position of the *anu7* mutations. Boxes and lines between boxes indicate exons and introns, respectively. White boxes represent the 5'- and 3'-UTRs. Triangles indicate T-DNA insertions. (b) Detail of an alignment of the ANU7 protein and some of its orthologs, showing the DnaJ-like CR domain. The alignment was generated with T-Coffee. The four zinc binding motifs of the DnaJ-like CR domain are underlined in black. Conserved cysteines and glycines are shaded black and blue, respectively. Plant species used for the multiple sequence alignment: Ath, *Arabidopsis thaliana*; Aly, *Arabidopsis lyrata*; Bna, *Brassica napus*; Bra, *Brassica rapa*; Cru, *Capsella rubella*; Tca, *Theobroma cacao*; Gra, *Gossypium raimondii*; Vvi, *Vitis vinifera*; Rco, *Ricinus communis*; Sly, *Solanum lycopersicum*; Pda, *Phoenix dactylifera*; Smo, *Selaginella moellendorffii*; and Ppa, *Physcomitrella patens*.

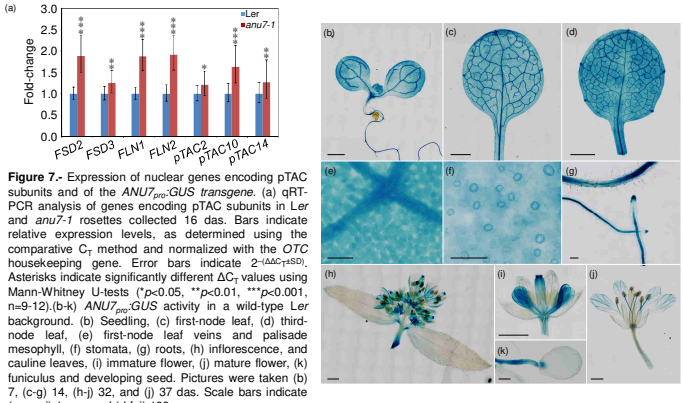


Figure 7.- Expression of nuclear genes encoding pTAC subunits and of the *ANU7$_{pro}$:GUS transgene*. (a) qRT-PCR analysis of genes encoding pTAC subunits in *Ler* and *anu7-1* rosettes collected 16 das. Bars indicate relative expression levels, as determined using the comparative C_T method and normalized with the *OTC* housekeeping gene. Error bars indicate $2^{-(\Delta\Delta C_T \pm SD)}$. Asterisks indicate significantly different ΔC_T values using Mann-Whitney U-tests (*p<0.05, **p<0.01, ***p<0.001, n=9-12).(b-k) *ANU7$_{pro}$:GUS* activity in a wild-type *Ler* background. (b) Seedling, (c) first-node leaf, (d) third-node leaf, (e) first-node leaf veins and palisade mesophyll, (f) stomata, (g) roots, (h) inflorescence, and cauline leaves, (i) immature flower, (j) mature flower, (k) funiculus and developing seed. Pictures were taken (b) 7, (c-g) 14, (h-j) 32, and (j) 37 das. Scale bars indicate (a-c, g-i) 1 mm, and (d-f, j) 100 µm.

REFERENCES

1.- Micol, J.L. (2009). *Current Opinion in Plant Biology* **12,** 9-16.
2.- Berná, G., Robles, P., and Micol, J.L. (1999). *Genetics* **152,** 729-742.
3.- Casanova-Sáez, R., Mateo-Bonmatí, E., Kangasjärvi, S., Candela, H., and Micol, J.L. (2014). *Journal of Experimental Botany* **65,** 2391-2404.
4.- Mateo-Bonmatí, E., Casanova-Sáez, R., Candela, H., and Micol, J.L. (2014). *Planta* **240,** 1113-1122.
5.- Muñoz-Nortes, T., Pérez-Pérez, J.M., Ponce, M.R., Candela, H., and Micol, J.L. (2017). *Plant Journal* **89,** 870-884.
6.- Craig, E.A., Huang, P., Aron, R., and Andrew, A. (2006). *Reviews of Physiology, Biochemistry and Pharmacology* **156,** 1-21.
7.- Cottage, A.J, Mott, E.K., Kempster, J.A., and Gray, J.C. (2010). *Journal of Experimental Botany* **61,** 3773-3786.
8.- Cottage, A.J., Mott, E.K., Wang, J.-H., Sullivan, J.A., MacLean, D., Tran, L., Choy, M.-K, Newell, C., Kavanagh, T.A., Aspinall, S., and Gray, J.C. (2008). *Photosynthesis. Energy From the Sun: 14th International Congress on Photosynthesis.*
9.- Huang, M., Friso, G., Nishimura, K., Qu, X., Olinares, P.D., Majeran, W., Sun, Q., and van Wijk, K.J. (2013). *Journal of Proteome Research* **12,** 491-504.

ACKNOWLEDGEMENTS

Research in the laboratory of J.L.M. is supported by grants from the Ministerio de Economía y Competitividad of Spain (BIO2014-53063-P) and the Generalitat Valenciana (PROMETEOII/2014/006). T.M.N. held a VALi+d fellowship from the Generalitat Valenciana (ACIF/2013/273). H.C. held a Marie Curie International Reintegration Grant (PIRG03-GA-2008-231073).

UNIVERSITAS
Miguel
Hernández

VII.- AGRADECIMIENTOS

UNIVERSITAS
Miguel
Hernández

VII.- AGRADECIMIENTOS

La realización de esta Tesis ha sido posible gracias a la financiación del trabajo que se realiza en el laboratorio de José Luis Micol por la Comisión Europea (AGRON-OMICS, LSHG-CT-2006-037704), la Generalitat Valenciana (PROMETEOII/2014/006) y el Ministerio de Economía y Competitividad (BFU2011-22825 y BIO2014-53063-P). Durante parte de mi periodo predoctoral he sido beneficiaria de un contrato del programa VALi+d (ACIF/2013/273) de la Generalitat Valenciana.

Gracias a mi director de Tesis, José Luis Micol, por depositar su confianza en mí y permitirme llevar a cabo esta Tesis Doctoral en su laboratorio.

A mi codirector de Tesis, Héctor Candela, por no dejar que me asfixie y por esas sentadas para escribir grandes historias. Sin ti no habría sido capaz.

A María Rosa Ponce, por su interés en mi trabajo y sus comentarios.

Al Profesor Pedro Robles, por su gran amistad y sus buenos consejos, y a los Profesores Víctor Quesada y José Manuel Pérez, por su apoyo y cercanía.

A los técnicos de laboratorio Paqui, Tania, Bea, Leila, Diana, Aurelia, José María, María José, Antonio y Juan. A José Manuel Serrano, el técnico ideal, por su amistad, por ser el motor del laboratorio y por estar siempre ahí.

A los “predocs” y “postdocs” que formaron parte del laboratorio durante mi Tesis (Silvia, Rafa, “las Almus”, Raquel, David Esteve, Rubén, Rebeca), por prestarme su ayuda en mis primeros años, porque aprendí mucho de ellos, y porque a día de hoy todavía siento que los tengo para lo que necesite, estén en Elche o en el Polo Norte.

A David Wilson, por ser un gran amigo y compañero de viajes. Por escucharme siempre y no juzgarme. Por las risas, las películas y las canciones.

A Rosa Micol, por preocuparse de cómo me siento en cada momento y cuidarme. Por tantísimos detalles. Por contar conmigo siempre.

A la siguiente generación de predoctorales, Edu (¡te cedo el relevo!), Carla, Alejandro, Sergio, Adrián, Lucía, Uri, Riad y Samuel, porque con ellos el laboratorio queda en buenas manos.

A mis alumnos y alumnas Lorena, Pili, Oliver, Adrián, Alejandro y Tamara, por su ayuda cada verano y por hacerme dudar a veces con sus preguntas, pues eso me animó a intentar mejorar cada día y tratar de aprender a transmitir lo que sé.

A los compañeros de Toxicología, por compartir tantos “tuppers” en el Hélike. También a las conserjes, a las camareras de la cantina y al personal de limpieza y mantenimiento, por el buen trato y por tener siempre una sonrisa que ofrecer.

A mi padre, la persona a la que más admiro. Por hacer todo lo necesario para sacar a la familia adelante. Por ser fuerte y aguantar con buen humor las zancadillas que te ha ido poniendo la salud en los últimos años. Gracias a ti me he convertido en la persona que soy. Te quiero papá.

A mi madre, porque sé que lo único que has querido siempre es protegerme, aunque tu manera de hacerlo a veces me resulte algo incomprensible. Te quiero mamá.

A mi iaia Asunción y a mi tío Pepito, porque siempre me pusieron de ejemplo ante todo el mundo y sé que les hubiera gustado mucho ver esto.

A Patricia, por hacerme feliz, quererme y apoyarme como nadie lo ha hecho. Por convertir las curvas en rectas. Por demostrarme que el amor entre dos personas se basa en el respeto. Me has traído la paz que necesitaba. Te amo, mi mujer de verde.

A mi familia política, por tratarme tan bien y hacerme sentir como una más.

A todos los profesores que me han formado a lo largo de mi vida académica, desde Alicia Arévalo hasta Pepe Nieto, pues sin ellos hoy no estaría escribiendo esto.

A mis compis Gyo (¡Te quiero mandurriala!), Cristina, Joan, Néstor, Diana (¡Mi arma!), Emma y Cris, por compartir piso conmigo durante la Tesis y aguantar mis quejas, mis ataques de limpieza, de risa y de cantar “a grito pelado”.

A mis pamplonicas Ainhoa y Kai, porque cuando necesito escapar del desierto siempre estáis ahí esperándome para un Rummy, un concierto o un enigma. Toda una vida juntos (¡te copio!). Sabéis que me gustaría estar más cerca. Os adoro. También a Elena y Paula, por cuidaros y haceros felices.

A mis luchadoras Espe y Miri, por ser grandes ejemplos de superación y enseñarme a valorar más la vida. Os tengo presentes.

A mis bailarinas Nere y Carmen, por dar ritmo y alegría a esta Tesis y por haberos quedado en mi vida.

A Cristian y Víctor. Lo más bonito de Granada no es la Alhambra...sois vosotros.

A Yolanda Pérez, por poner nombre y apellidos a las cosas que pienso y siento, y por ayudarme a buscar soluciones para luchar contra mí misma, mi entorno y mis fantasmas del pasado y del presente.

A mis animalitos, por alegrarme la vida cada día al llegar a casa.

A todos los que en algún momento se han interesado por mí y por mi Tesis durante estos años.