Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

Estudio transcriptómico y proteómico de la biogénesis del cloroplasto en el modelo albino de *Agave angustifolia* Haw.

Tesis que presenta

Mauricio Andrade Marcial

En opción al título de

DOCTOR EN CIENCIAS

(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán, México

Septiembre, 2023

CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.
POSGRADO EN CIENCIAS BIOLÓGICAS

RECONOCIMIENTO

Por medio de la presente, hago constar que el trabajo de tesis de **Mauricio Andrade Marcial** titulado "**Estudio transcriptómico y proteómico de la biogénesis del cloroplasto en el modelo albino de *Agave angustifolia* Haw.",** fue realizado en la Unidad de Biotecnología, en la línea de investigación de Agrobiotecnología, en el Laboratorio de Epigenética del Centro de Investigación Científica de Yucatán, A.C. bajo la dirección de la **Dra. Clelia De la Peña Seaman**, dentro de la opción de Biotecnología, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Atentamente

Dra. Cecilia Hernández Zepeda
Directora de Docencia

Mérida, Yucatán, México, a 19 de septiembre de 2023

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en las secciones de: Materiales y Métodos, Resultados y Discusión de este documento, proviene de las actividades de investigación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que, a razón de lo anterior, y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley Federal de Protección a la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y se regirán, en todo caso, por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley Federal de Protección a la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.

Firma: _____

M. en C. Mauricio Andrade Marcial

Este trabajo se llevó a cabo en la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, A.C., con el financiamiento del Consejo Nacional de Humanidades, Ciencias y Tecnologías (Proyecto CB2016-285898, Ciencia Básica 2016) bajo la dirección de la Dra. Clelia De la Peña Seaman.

AGRADECIMIENTOS

Al **Consejo Nacional de Humanidades, Ciencias y Tecnologías**, por otorgarme la beca nacional con número 687732.

Al **Centro de Investigación Científica de Yucatán, A.C.** y a la **Unidad de Biotecnología**, por la oportunidad de realizar mis estudios de doctorado y por facilitarme todo lo necesario para su culminación.

A mi asesora, la **Dra. Clelia De la Peña Seaman**, porque desde mi ingreso al doctorado hasta el día de hoy, siempre mostró el liderazgo, la confianza, la paciencia y el compromiso en la culminación de mi proyecto de investigación.

A mi comité tutoral y revisores de tesis: **Dra. Elsa Góngora**, **Dr. José Ramón Pacheco**, **Dr. Enrique Castaño**, **Dr. Victor Manuel Loyola**, **Dra. Virginia Aurora Herrera** y **Dr. Luis Alfonso Sáenz**, por sus recomendaciones para mejorar mi proyecto de investigación.

Al **Dr. Eliel Ruiz May** y al **M. en C. José Miguel Elizalde Contreras** por su apoyo durante mi estancia en el Laboratorio de Proteómica de la Red de Estudios Moleculares Avanzados del Instituto de Ecología, A. C.

A los técnicos **Lic. Eduardo Castillo Castro**, **Lic. María Rosalina Rodríguez Román** y **Lic. Verónica Limones Briones** por su apoyo y orientación en el laboratorio.

A mis compañeros del laboratorio de epigenética: **Alexis Salazar**, **Alexis Cadena**, **Ligia Aguilar**, **Edder Aguilar**, **Fernando de Jesús Colorado**, **Alan Barragán**, **Jade Pereyda** y **Luis Núñez** por los entrañables momentos que vivimos durante nuestra estancia en el CICY. Siempre los recordaré como hombres y mujeres que como yo, se embarcaron en esta travesía de estudiar una carrera científica. Les deseo toda la suerte del mundo.

DEDICATORIAS

A mis padres, **Alfredo y Leticia**. No tengo palabras que plasmen total y fielmente lo que representan para mí. Después de 30 años, he descubierto que cada logro académico que alcanzo es mi manera de agradecerles el dar la vida día a día por mí y por mis hermanos. Hoy entiendo que este es NUESTRO logro. Aún con mis errores y limitaciones, siempre llevaré conmigo su naturalidad, humildad, honradez, nobleza, fidelidad y libertad. Los amo.

A **Diego, Valeria** y a **mi otro hermano**. Gracias por estos años caminando juntos. Me reconforta saber que, a pesarde mis errores y omisiones, el amor que nos inculcaron nuestros padres sigue presente y nos supera. Continúo creyendo que su aparición tan repentina en mi vida ha sido el mayor regalo que me ha dado Dios. Siempre, cerca o lejos, serán mis compañeros de vida.

A **mi abuelo Toño**. Hoy que es imposible verte de nuevo al llegar a tu casa, me doy cuenta de lo mucho que te extraño. Durante años, creí que mi presencia no significaba nada para ti. Justo en el ocaso de tu vida, me demostraste que tu amor era igual de fuerte que tu templanza y carácter en momentos difíciles. Vuelvo a confirmar que parte de lo que soy fue forjado por tus manos. Te amo abuelo, y espero que algún día pueda volver a encontrame contigo y con **mis abuelas Lupe y Josefina**. Siempre admiraré su apertura a la vida.

A **David y Rebeca**. Gracias por su presencia a lo largo de estos diez años. Sé que ambos me conocen en todas las facetas de mi vida y, en cada circunstancia, han sacado lo mejor de mí, aunque a veces me cueste tanto demostrarlo. Gracias Rebeca por ser mi amiga más cercana y confiable en esta carrera científica que emprendimos casi al mismo tiempo. David, te agradezco tu perseverancia y paciencia conmigo. Hoy puedo ver que tu irrupción en mi historia siempre trajo consigo el enfrentarme a quien soy y a replantearme que quiero en mi futuro. Les amo demasiado.

LISTA DE LOS PRODUCTOS GENERADOS

I. Artículos científicos:

- Hernández-Castellano, S., **Andrade-Marcial, M.**, Aguilar-Méndez, E. D., Loyola-Vargas, V. M., de Folter, S., De-la-Peña, C. (2022). MiRNA expression analysis during somatic embryogenesis in *Coffea canephora*. *Plant Cell Tissue and Organ Culture*, 150, 177–190.

- **Andrade-Marcial, M.**, Pacheco-Arjona, R., Góngora-Castillo, E., De-la-Peña, S. (2022). Chloroplastic pentatricopeptide repeat proteins (PPR) in albino plantlets of *Agave angustifolia* Haw. reveal unexpected behavior. *BMC Plant Biology*, 22, 352.

- **Andrade-Marcial, M.**, Ruíz-May, E., Elizalde-Contreras, J.M., Pacheco, N., Herrera-Pool, E., De-la-Peña, C. (2023). Proteome of *Agave angustifolia* Haw.: uncovering metabolic alterations, over-accumulation of amino acids, and compensatory pathways in chloroplast-deficient albino plantlets. *Plant Physiology and Biochemistry*, 201, 107902.

II. Capítulo de libro:

- Limones-Briones, V., **Andrade-Marcial, M.**, Aguilar-Méndez, E.D., De-la-Peña, C. (2023). Luces, fotosíntesis, acción: el despertar de los cloroplastos, en: Descubriendo mi talento 2022, De-Gante-Ayora, F.M., Herrera-Alamillo, M.Á. (Comps.). Centro de Investigación Científica de Yucatán, A.C., pp-52-67.

III. Dirección de tesis:

- Che-Aguilar, L. (2023). **Expresión diferencial de los genes cloroplásticos en variantes somaclonales de *Agave angustifolia* Haw.** Tesis de Licenciatura. Ingeniería Bioquímica. Instituto Tecnológico de Mérida. Tecnológico Nacional de México. Yucatán, México. En codirección con la Dra. Clelia De la Peña Seaman.

IV. Actividades de divulgación:

- **Visitas universitarias en el CICY.** Centro de Investigación Científica de Yucatán, A.C. Mérida, Yucatán, México. Enero-Junio de 2022.

- **Club Ambiental. Sesión 2: ¿Cómo ver los microplásticos en el agua? ¿Son un peligro para el zooplancton?** Centro de Investigación Científica de Yucatán, A.C. Mérida, Yucatán, México. Marzo de 2022.

- **Curso: Luces, fotosíntesis, acción: el despertar de los cloroplastos.** Programa Talento CICY 2022-Preparatoria. Centro de Investigación Científica de Yucatán, A.C. Mérida, Yucatán, México. Agosto de 2022.

- **¡Doctor! ¡Mi agave albino es adicto a los LEGO!** Mi tesis en 10 minutos. 3º Evento anual de charlas de divulgación científica 2022. Revista Ciencia Cakotanú, A.C. Octubre de 2022.

V. Congresos:

- **Andrade-Marcial, M.**, Ruíz-May, E., Elizalde-Contreras, J.M., Pacheco, N., Herrera-Pool, E. y De-la-Peña, C. **Proteome of *Agave angustifolia* Haw.: insights about amino acids metabolism in albino plantlets**. XXXIII Congreso Nacional de Bioquímica. 16-21 de octubre de 2022. Mérida, Yucatán, México.

- **Andrade-Marcial, M.**, Pacheco-Arjona, R., Góngora-Castillo, E. y De-la-Peña, C. **Chloroplastic pentatricopeptide repeat proteins (PPR) in albino plantlets of *Agave angustifolia* Haw. reveal unexpected behavior**. XXXIII Congreso Nacional de Bioquímica. 16-21 de octubre de 2022. Mérida, Yucatán, México.

VI. Distinciones

- Mención honorífica en el concurso “Mi tesis en 10 minutos” con la presentación **¡Doctor! ¡Mi *Agave* albino es adicto a los LEGO!** 3º Evento anual de charlas de divulgación científica 2022. Revista Ciencia Cakotanú, A.C. Octubre de 2022.

ÍNDICE

LISTADO DE FIGURAS

LISTADO DE TABLAS

ABREVIATURAS

2,4-D	2,4-diclorofenoxiacético
A	Plántulas albinas
AA o AAs	Aminoácidos
AaALS	Enzima acetolactato sintasa
AaAST	Enzima aspartato transaminasa
AaBCAT	Enzima aminotransferasa de cadena ramificada
AaCGS	Enzima cistationina gama sintasa
AaCM	Enzima corismato mutasa
AaIPMS	Enzima isopropilmalato sintasa
AaMS	Enzima metionina sintasa
AaSHMT	Enzima serina hidroximetil transferasa
ADN o DNA	Ácido desoxirribonucleico
AL	Hoja albina
AM	Meristemo albino
ANOVA	Análisis de varianza
ARN o RNA	Ácido ribonucleico
ARNm o RNAm	ARN mensajero
ARNr o RNAr	ARN ribosomal
ARNt o RNAt	ARN de transferencia
ATP	Adenosín trifosfato
BA o BAP	6-benciladenina
BCAA	Aminoácidos de cadena ramificada
CAM	Metabolismo ácido de las crasuláceas
CDS	Secuencia codificante
CTV	Cultivo de tejidos vegetales
DAP	Proteína diferencialmente acumulada
DEG	Gen diferencialmente expresado
DYW	Dominio Aspartato-Tirosina-Triptofano
G	Plántulas verdes
GL	Hoja verde
GM	Meristemo verde

LHC	Complejo de cosecha de luz
MS	Espectrometría de masas
NADPH	Nicotinamida adenina dinucleótido fosfato
NEP	RNA polimerasa codificada por el núcleo
NGS	Secuenciación de nueva generación
ORF	Marco de lectura abierto
PEP	RNA polimerasa codificada por el plástido
PPR	Repeticiones pentatricopeptídicas
PSI	Fotosistema I
PSII	Fotosistema II
qRT-PCR	PCR cuantitativa de transcripción inversa
RNA-seq	Secuenciación de ARN
RuBisCO	Ribulosa-1,5-bisfosfato carboxilasa/oxigenasa
SDS-PAGE	Electroforesis en gel de poliacrilamida con dodecilsulfato de sodio
SNP	Polimorfismo de nucleótido único
TCA	Ciclo de los ácidos tricarboxílicos
TPR	Repeticiones tetratricopeptídicas
V	Plántulas variegadas
VS	Variación somaclonal

RESUMEN

Las técnicas de cultivo *in vitro* de tejidos (CTV) han revolucionado la forma de hacer investigación en plantas. Uno de los grupos taxonómicos que más se han beneficiado de estas herramientas son los *Agaves*. Durante la micropropagación de plántulas verdes de *Agave angustifolia* Haw. en el Laboratorio de Epigenética del Centro de Investigación Científica de Yucatán, A.C., un evento espontáneo y no provocado por la presencia de agentes mutagénicos conocidos, originó la aparición de una variante somaclonal albina. Esto dio paso a una nueva línea de investigación enfocada al estudio del *Agave* albino, el cual se caracteriza por la ausencia de clorofila, un metabolismo heterotrófico y un bloqueo en la biogénesis del cloroplasto. La ausencia de cloroplastos funcionales le plantea a la planta retos metabólicos y fisiológicos ya que este organelo es indispensable para la sobrevivencia de la planta. En este estudio, se generaron datos transcriptómicos y proteómicos para profundizar en las alteraciones en la biogénesis del cloroplasto en el *Agave* albino.

Nuestros resultados transcriptómicos sugieren que las plántulas de *Agave* albino presentan una reprogramación transcripcional caracterizada por una mayor expresión de genes vinculados con la división del plástido, la importación de proteínas, la regulación de la actividad transcripcional y traduccional, el metabolismo de tetrapirroles y la fotosíntesis. De particular interés es la intensa transcripción de genes con repeticiones pentatricopeptídicas (*PPR*), los cuales son claves en el metabolismo de ARN en el plástido. En conjunto esto sugiere que los plástidos albinos se mantienen en un estadio juvenil de desarrollo, muy similar al del proplástido. Además, es posible que las rutas de señalización retrógrada/anterógrada se encuentren comprometidas, por lo que la activación de los genes *PPR* podría formar parte de una ruta compensatoria que intenta corregir y restaurar la correcta biogénesis del cloroplasto. Por otra parte, los resultados proteómicos revelaron una fuerte activación del metabolismo de aminoácidos (AA), particularmente de enzimas biosintéticas así como un incremento en la concentración de AA de cadena ramificada, aromáticos, Asp, Lys, Met y Pro. Esto revela que en respuesta a la ausencia de cloroplastos, estas plántulas modifican el metabolismo de AA para poder mantener sus funciones esenciales activas y responder a las condiciones ambientales en ausencia de su metabolismo autotrófico.

ABSTRACT

In vitro tissue culture techniques (CTV) have revolutionized plant research. One taxonomic group that has benefited the most from these tools are the *Agaves*. During micropropagation of green *Agave angustifolia* Haw. Plantlets in the Epigenetics Laboratory of the Centro de Investigación Científica de Yucatán, A.C., a spontaneous event not caused by the presence of known mutagenic agents gave rise to an albino somaclonal variant. This led to a new line of research focused on the study of albino *Agave*, which is characterized by the absence of chlorophyll, a heterotrophic metabolism, and a blockage in chloroplast biogenesis. The absence of functional chloroplasts poses metabolic and physiological challenges to the plant, as this organelle is essential for survival. This study generated transcriptomic and proteomic data to gain a deeper understanding of the alterations in chloroplast biogenesis in albino *Agave*.

The transcriptomic results suggest that albino *Agave* plantlets exhibit transcriptional reprogramming, characterized by increased expression of genes linked to plastid division, protein import, regulation of transcriptional and translational activity, tetrapyrrole metabolism, and photosynthesis. Of particular interest is the intense transcription of pentatricopeptide repeat (*PPR*) genes, which are key in RNA metabolism in the plastid. This suggests that albino plastids remain in a juvenile state of development, very similar to that of the proplastid. It is also possible that the retrograde/anterograde signaling pathways are compromised, so the activation of the *PPR* genes could be part of a compensatory pathway that attempts to correct and restore correct chloroplast biogenesis. On the other hand, the proteomic results revealed a strong activation of AA metabolism, particularly of biosynthetic enzymes, as well as an increase in the concentration of branched-chain AAs (BCAAs), aromatic AAs, Asp, Lys, Met, and Pro. This suggests that, in response to the absence of chloroplasts, these plantlets modify their AA metabolism to maintain their essential functions active and respond to environmental conditions in the absence of their autotrophic metabolism.

INTRODUCCIÓN

Una de las herramientas biotecnológicas más socorridas en la investigación en plantas es el cultivo *in vitro* de tejidos vegetales (CTV). Este es definido como el cultivo de cualquier parte de una planta, ya sean células, tejidos u órganos, en un medio artificial, en condiciones estériles y bajo condiciones ambientales controladas. De entre las diferentes herramientas del CTV, una de las más extendidas y utilizadas comercialmente es la micropropagación (Loyola-Vargas y Ochoa-Alejo, 2018).

Aunque uno de los principios del CTV es que todo explante debe regenerar clonas genotípica y fenotípicamente uniformes, en ocasiones esta supuesto no se cumple y aparecen variaciones somaclonales (VS). La VS es definida como la variabilidad que puede surgir en plantas propagadas a través de técnicas de CTV (Duta-Cornescu *et al.*, 2023; Weckx *et al.*, 2019).

En el contexto de la micropropagación de plantas de *A. angustifolia* con fenotipo verde en el Laboratorio de Epigenética del Centro de Investigación Científica de Yucatán, A.C., un evento espontáneo de VS del cual aún se desconocen su(s) causa(s), provocó la aparición de dos poblaciones de *Agaves* con características morfológicas únicas y detectables a simple vista: una de ellas presentaba un fenotipo variegado (V) y la otra uno albino (A).

El albinismo vegetal se caracteriza por I) la ausencia total o parcial de la capacidad de producir pigmentos fotosintéticos como la clorofila y los carotenoides, II) un bloqueo en la biogénesis del cloroplasto, en la que los proplástidos no logran diferenciarse en cloroplastos y III) un pobre o nulo desarrollo de las membranas tilacoidales que alojan los fotosistemas (Satou *et al.*, 2014; Kumari *et al.*, 2009).

Los organismos fotótrofos son esenciales para sustentar vida en la tierra debido a su habilidad para realizar fotosíntesis. Actualmente es bien aceptado que este organelo se originó a partir de la fagocitosis de una protocianobacteria fotosintética por una eucariota heterótrofa primitiva, en un proceso conocido como endosimbiosis primaria (Maréchal, 2018; Archibald, 2015). Con el paso del tiempo la cianobacteria fagocitada dio paso a lo que hoy conocemos como cloroplasto, el cual es el centro metabólico por excelencia de la

célula vegetal (Füssy y Oborník, 2018; Daniell *et al.*, 2016). La ausencia de estos organelos en modelos albinos explica su incapacidad de mantener un metabolismo fotosintético y autotrófico activo.

En plantas con fenotipos albinos se tiene conocimiento que los proplástidos tienen problemas durante su proceso de diferenciación en cloroplastos maduros (Satou *et al.*, 2014). El proceso en el que los proplástidos, después de una serie de eventos de división en respuesta a la percepción de la luz, se diferencian en cloroplastos funcionales es conocido como biogénesis del cloroplasto (Cackett *et al.*, 2022). Recientemente se ha demostrado que una familia de proteínas codificadas en el núcleo y conocidas como proteínas con repeticiones pentatricopeptídicas (PPR) exhiben funciones críticas en el metabolismo del ARN y por consiguiente, en la conformación y funcionalidad del cloroplasto (Pogson y Albrecht, 2011).

Los nuevos avances tecnológicos que se aplican hoy en día en la investigación con plantas, entre las que se incluyen las herramientas ómicas como la genómica, transcriptómica, proteómica y metabolómica, han revolucionado la forma de entender como las plantas responden a condiciones específicas que afectan su desarrollo y crecimiento (Ahmad *et al.*, 2016; Ahsan *et al.*, 2010).

En este proyecto de investigación haremos uso de dos herramientas ómicas, la transcriptómica y la proteómica, para abordar el estudio de genes claves en la biogénesis del cloroplasto en las variantes somaclonales de *A. angustifolia*. Este enfoque novedoso permitirá obtener un panorama general de las principales alteraciones moleculares que presenta el *Agave* albino y que impiden su correcta diferenciación en cloroplastos maduros. Adicionalmente, haremos un estudio más profundo de la familia de genes *PPR*, la cual parece ser uno de los protagonistas en la comunicación entre el núcleo y el cloroplasto durante la biogénesis de este organelo. Por último, abordaremos el estudio del metabolismo de AA desde una perspectiva proteómica, con el fin de identificar las principales alteraciones que exhibe este modelo albino

CAPÍTULO I

ANTECEDENTES

1.1 ASPECTOS GENERALES DEL GÉNERO *Agave*

El género *Agave sensu lato* es uno de los nueve géneros que conforman a la familia *Asparagaceae* y cuenta con alrededor de 211 especies descritas (Flores-Abreu *et al.*, 2019). Este grupo de plantas angiospermas monocotiledóneas y de ciclos de vida largos, se caracteriza por ser monocárpico, tolerar la sequía, presentar una morfología arrosetada, una larga inflorescencia central en forma de espiga, así como un síndrome de polinización por murciélagos (Eguiarte *et al.*, 2021; Flores-Abreu *et al.*, 2019).

México es considerado el centro de origen y diversidad de los *Agaves*, donde se distribuye en ecosistemas áridos y semiáridos del altiplano. Alrededor del 79% de especies de este género son encontradas en México. Sin embargo, su presencia se ha extendido desde el sur de los Estados Unidos de América hasta países de Centroamérica -incluyendo las Islas del Caribe- y Sudamérica (Eguiarte *et al.*, 2021; Good-Avila *et al.*, 2006).

Como resultado de su distribución espacial, los *Agaves* presentan características y adaptaciones morfo-fisiológicas y metabólicas distintivas que en conjunto, les han permitido hacer frente y sobrevivir a ambientes extremos. Por ejemplo, uno de los elementos claves para entender el éxito de su nicho ecológico es que son plantas suculentas -con altos niveles de agua acumulándose en sus células-, hojas ricas en fibras resistentes y con cutículas cerosas que evitan su deshidratación durante el día. Otro rasgo característico es que han desarrollado un tipo de fotosíntesis conocido como metabolismo ácido de las crasuláceas (*crassulacean acid metabolism* o CAM, por sus siglas en inglés), en el que la planta cierra sus estomas durante el día y los abre por la noche, para evitar la pérdida de H_2O debido a las altas temperaturas y la sequía (Álvarez-Chávez *et al.*, 2021; Davis *et al.*, 2019; Stewart, 2015).

Debido a que las condiciones ambientales están cambiando como consecuencia del cambio climático, el tema de los desafíos para el cultivo de plantas con importancia agrícola ha preocupado a la comunidad científica internacional. Actualmente se considera que los

Agaves, más que otras plantas CAM de importancia comercial –como la piña o la tuna–, son plantas que responderán mejor a los efectos del cambio climático (Stewart, 2015). Adicionalmente, los *Agaves* han sido utilizados en otros ámbitos de la vida diaria como en la alimentación, la industria cosmética y de fibras textiles, la construcción, la producción de biomateriales y biocombustibles, la nutrición, la farmacéutica, como plantas ornamentales y en la producción de bebidas alcohólicas y no alcohólicas (Álvarez-Chávez *et al.*, 2021; Pérez-Zavala *et al.*, 2020; Davis *et al.*, 2019; Stewart, 2015). Dentro de las especies de este género, una de las que sobresale por su importancia en el ámbito económico y comercial es *Agave angustifolia* Haw. (**Figura 1.1**).

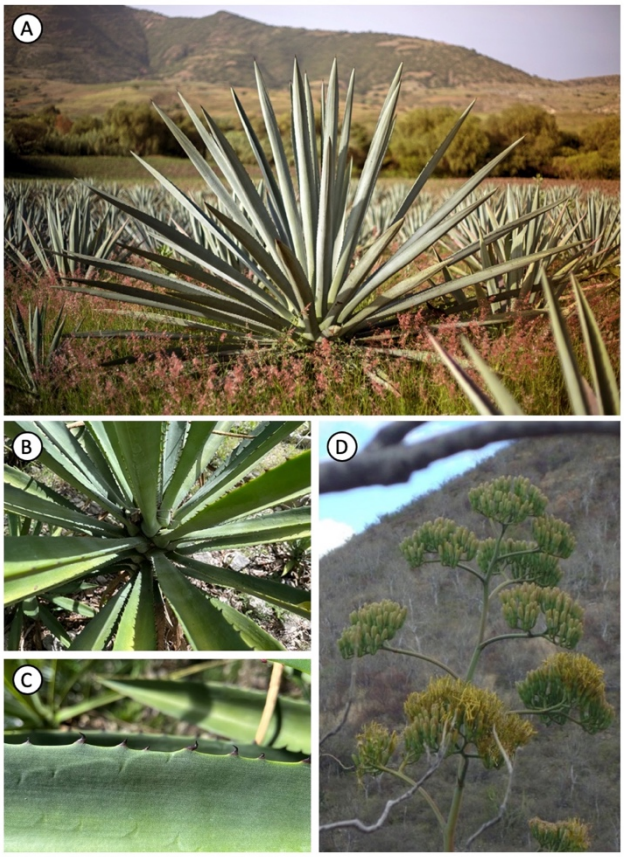


Figura 1.1. Crecimiento en condiciones naturales de *Agave angustifolia* Haw. A) Fotografía de la planta de *A. angustifolia* tomada de https://mezcologia.mx/agave-angustifolia/. Foto tomada por Santiago Barreido. **B)** y **C)** Arquitectura de la roseta y espinas características de la especie. Fotos tomadas por Mauricio Andrade. **D)** Inflorescencia. Foto tomada y modificada de SAGARPA (2016).

1.2. EL CASO PARTICULAR DE *Agave angustifolia* Haw.

Denominado popularmente como “maguey delgado” o “espadín” (haciendo referencia a sus características hojas lanceoladas y rígidas), *A. angustifolia* es reconocido como la principal materia prima para la producción del mezcal conocido como bacanora en diversos estados de la República Mexicana, principalmente Sonora y Oaxaca (**Figura 1.1**) (Barrientos-Rivera *et al.*, 2019; Arzate-Fernández *et al.*, 2016; SAGARPA, 2016).

En el último par de décadas, la popularidad y demanda en el consumo de las bebidas alcohólicas obtenidas a partir de *Agaves* ha incrementado considerablemente en México y en el ámbito internacional. Tan solo de 2003 a 2016, la producción potencial de *Agave* en nuestro país se incrementó cerca del 280%. Para el caso particular del mezcal, las exportaciones durante el periodo 2014-2016 tuvieron un crecimiento acumulado del 267%, siendo EUA el principal importador del producto mexicano con cerca del 70% del volumen total (SAGARPA, 2017).

La búsqueda por lograr aumentar la producción de mezcal para satisfacer la demanda del mercado, trajo consigo que la materia prima obtenida a partir de poblaciones silvestres haya sido sobreexplotada. Este hecho puso a las poblaciones de *A. angustifolia* en riesgo de sufrir erosión genética debido a la pérdida de las variedades (SAGARPA, 2016; Núñez-Noriega y Salazar-Solano, 2009). Esto quedó descrito en la producción potencial de agave mezcalero, la cual se redujo alrededor del 31% (SAGARPA, 2017).

Como sucede con otros miembros del género *Agave*, diversos aspectos de la reproducción sexual de *A. angustifolia* añaden dificultades técnicas para su preservación y variabilidad genética en condiciones naturales. Entre ellas podemos incluir su ciclo de vida de larga duración (de alrededor de 6 a 8 años), una única floración, polinización limitada, dificultades en el establecimiento y germinación de las semillas así como en la supervivencia de las plántulas (SAGARPA, 2016; Núñez-Noriega y Salazar-Solano, 2009)

Como una alternativa, las plantaciones de *Agave* han sido reproducidas asexualmente en la naturaleza de dos formas: a través de brotes producidos por los rizomas basales de la planta o de bulbillos producidos a partir de las inflorescencias (Robert *et al.*, 2006). El hecho que los *Agaves* se reproduzcan vegetativamente ha permitido a los

productores hacer uso de las poblaciones silvestres como materia prima (SAGARPA, 2016). Sin embargo, los descendientes de las plantas madre aunque son clonas, suelen presentar afectaciones a nivel de tamaño y vigor y son susceptibles a la transmisión vertical de enfermedades. Otro de los problemas asociados a la reproducción por rizomas, es que una planta produce alrededor de 25 rizomas en un periodo de cinco años (Robert *et al.*, 2006). Esto pone de manifiesto la dificultad de satisfacer la demanda creciente del mercado, de una manera sustentable, con las técnicas agrícolas convencionales. La búsqueda de una alternativa viable, de bajo costo y eficiente, han colocado al CTV como una posible herramienta a utilizar.

1.3. EL CULTIVO *IN VITRO* COMO HERRAMIENTA BIOTECNOLÓGICA PARA LA MICROPROPAGACIÓN DE *A. angustifolia*

El CTV ha sido definido como el cultivo de cualquier parte de una planta (células, tejidos u órganos) en un medio artificial, en condiciones estériles y bajo condiciones ambientales controladas. Entre las diferentes herramientas del CTV, una de las más extendidas y utilizadas comercialmente es la micropropagación (Loyola-Vargas y Ochoa-Alejo, 2018). Esta herramienta es definida como la propagación clonal y masiva de plantas a partir de pequeños tejidos, bajo condiciones *in vitro* (Kaur et al., 2018). Para los *Agaves*, la micropropagación se puede realizar a partir de dos tipos de explantes: embriones cigóticos o brotes creciendo a partir de la planta madre (Rodríguez-Garay y Rodríguez-Domínguez, 2018).

Considerando las limitantes en la reproducción de *A. angustifolia* y en otras especies del género *Agave*, y aunado a la demanda comercial por materia prima, la micropropagación ofrece ventajas únicas. Entre ellas podemos mencionar que su aplicación permite una rápida propagación de material vegetal de alta calidad, la generación de miles de clonas de fácil mantenimiento, poca variabilidad genética y libres de enfermedades (Robert *et al.*, 2006). En los últimos 30 años han surgido un número considerable de protocolos que permiten la micropropagación de diferentes especies de *Agave* (Delgado-Aceves *et al.*, 2022; Rodríguez-Garay y Rodríguez-Domínguez, 2018; Puente-Garza *et al.*, 2015; Chen *et al.*, 2014; Robert *et al.*, 2006; Das, 1992; Robert *et al.*, 1992). En nuestro laboratorio, la micropropagación de *A. angustifolia* se ha venido realizando a partir del

cultivo de brotes surgiendo periódicamente de plantas madre utilizando el protocolo -con ligeras modificaciones- propuesto por Robert *et al.* (2006); Robert *et al.* (1992).

Si bien el CTV es una herramienta biotecnológica por excelencia, su aplicación puede someter al modelo vegetal de interés a diversos tipos de estrés, los cuales pueden inducir algún tipo de variación en las clonas obtenidas respecto al explante inicial (Weckx *et al.*, 2019; Manchanda *et al.*, 2018).

1.4. LA VARIACIÓN SOMACLONAL EN EL CULTIVO DE TEJIDOS VEGETALES

A pesar de que uno de los principios del CTV es que todo explante utilizado debe regenerar clonas genotípica y fenotípicamente uniformes, esto no es una regla que se cumpla. Uno de los fenómenos que presenta una alta frecuencia de aparición en condiciones *in vitro* es la VS, la cual es considerada un fenómeno ubicuo durante el CTV. Su aparición puede significar para algunos una fuente de diversidad en programas de mejoramiento genético y para otros, un obstáculo si esto provoca pérdida de la fidelidad genética (Ferreira *et al.*, 2023; Manchanda *et al.*, 2018; Krishna *et al.*, 2016).

La VS es definida como la variabilidad que puede surgir en plantas propagadas a través de técnicas de CTV. Su aparición está fuertemente vinculada a las condiciones de estrés a las que se someten estas plantas por las diferentes herramientas del CTV (Duta-Cornescu *et al.*, 2023; Weckx *et al.*, 2019). Algunas de las causas de estrés más comunes son los procesos de esterilización y tipo de explante, las técnicas de propagación, daño mecánico en el tejido, condiciones de iluminación, número y duración de los ciclos de subcultivo, cambios en los componentes de los medios, entre los que destacan los reguladores de crecimiento y sus concentraciones, etc. (Ferreira *et al.*, 2023; Manchanda *et al.*, 2018; Sahijram, 2015). En consecuencia, se ha propuesto que la VS puede tener un origen genético o epigenético, sin que esto se exprese obligatoriamente en el fenotipo de la planta (Duta-Cornescu *et al.*, 2023).

Muchos ejemplos han sido ofrecidos para elucidar qué mecanismos genéticos podrían estar asociados a la aparición de VS. Del lado del aspecto genético los más populares son las roturas cromosómicas, cambios en el número y en la estructura cromosómica, poliploidía, entrecruzamiento cromosómico, mutaciones en el ADN,

activación de elementos transponibles, entre otros (Manchanda *et al.*, 2018; Sahijram, 2015). Por el lado epigenético, la aparición de VS suele estar relacionada con la metilación en el ADN y en las histonas (Karim *et al.*, 2016).

Uno de los aspectos más importantes al evaluar la aparición de VS es utilizar un método eficiente para su detección. Las técnicas clásicas para ello incluyen análisis morfológicos/fenotípicos, citológicos, fisiológicos, bioquímicos y moleculares (Bairu *et al.*, 2011). De éstas, la detección morfológica ha sido una de las más exitosas y sencillas, ya que puede ser detectada preliminarmente solo con una inspección visual (Krishna *et al.*, 2016). Esta se ha enfocado en anormalidades en la pigmentación, cambios en el crecimiento vegetal y arquitectura vegetal, hojas y flores con formas modificadas, entre otros (Chen y Henny, 2006). Sin embargo, la VS no siempre se asocia con evidencia morfológica suficiente para concluir la ocurrencia de un evento de esta naturaleza (Bairu *et al.*, 2011), lo que hace necesario análisis adicionales para caracterizar este fenómeno.

1.5. VARIACIÓN SOMACLONAL EN PLÁNTULAS DE *A. angustifolia*: EL ORIGEN DE LOS FENOTIPOS VARIEGADO Y ALBINO

En el contexto de la micropropagación de plantas de *A. angustifolia* con fenotipo verde (G) en el Laboratorio de Epigenética del Centro de Investigación Científica de Yucatán, A.C., un evento espontáneo de VS del cual aún se desconocen su(s) causa(s), provocó la aparición de dos poblaciones de *Agaves* con características morfológicas únicas y detectables a simple vista: una de ellas presentaba un fenotipo variegado (V) y la otra uno albino (A) (**Figura 1.2**). Los datos disponibles hasta este momento indican que esta VS podría tener un origen epigenético en función de los resultados observados a nivel de metilación del ADN y de modificaciones a nivel de histonas (Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016). Otro dato que le da sustento a esta hipótesis, es que al estudiar la variabilidad genética mediante AFLP (amplified fragment length polymorphism, por sus siglas en inglés) se obtuvo un coeficiente de similitud muy alto (0.97) entre las variantes fenotípicas, lo que indica una estabilidad genética entre los fenotipos G, V y A (Us-Camas *et al.*, 2017).

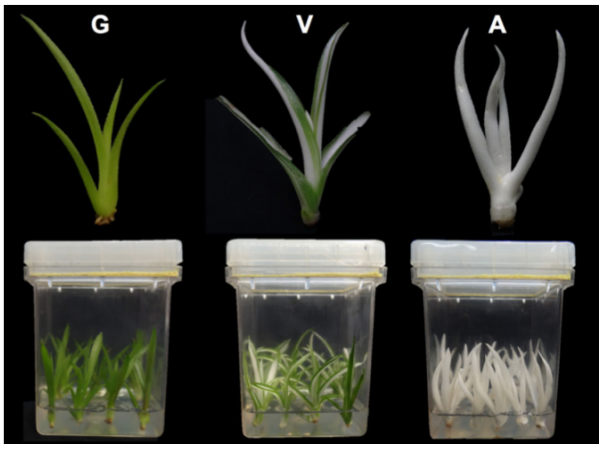


Figura 1.2. Variantes somaclonales de *A. angustifolia*. Dos variantes somaclonales de la especie fueron obtenidos a partir de sucesivas rondas de micropropagación de plantas verdes. Las variantes presentaron alteraciones en su pigmentación, una con un fenotipo albino (A) y la otra uno variegado (V). En la imagen se presentan fotografías de plántulas de cada fenotipo y la forma en la que son cultivadas en cajas magenta. Imagen tomada de Duarte-Aké *et al.* (2016).

Se han realizado algunos estudios morfológicos, fisiológicos, celulares, moleculares y bioquímicos en los tres fenotipos de *A. angustifolia* para comprender mejor los cambios extrínsecos e intrínsecos generados por este evento de VS (Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016). Los resultados más sobresalientes reportados hasta este momento, son descritos en la **Tabla 1.1**.

Tabla 1.1. Características generales de los tres fenotipos de *A. angustifolia* obtenidos por micropropagación

	Fenotipo			
				
	Verde (G)	**Variegado (V)**		**Albina (A)**
		GV*	**AV***	
MORFOLOGÍA GENERAL				
Color de la planta	Verde	Regiones verdes y blancas a lo largo de la lámina foliar		Blanco

Roseta		Clásica	Más abierta respecto al fenotipo G		Más abierta respecto al fenotipo G
Meristemo		Normal	Normal		No diferenciado y con base tipo callo
ESTOMAS[+]					
Forma		Redondeados con presencia de ceras epicuticulares alrededor de células subsidiarias no desarrolladas.	Hundidos y carentes de ceras epicuticulares. Deformación en el sitio donde debían desarrollarse células subsidiarias.		Elípticos y las células epidérmicas presentaban mayor elongación
# de células adyacentes		4	4		2
Densidad estomática	**Ad**	+++	++	+	+
	Ab	+++	+	+	++
Longitud	**Ad**	+++	++	+	+
	Ab	+++	++	+	++
Anchura	**Ad**	+++	++	+	+
	Ab	+++	+	++	+++
Apertura	**Ad**	++	+++	+	+
	Ab	+++	++	+	+
PIGMENTOS FOTOSINTÉTICOS					
Clorofila *a*		+++	++		+
Clorofila *b*		++	+++		+
Clorofila total		+++	++		+
Carotenoides		+++	++		+
Clorofila *a*/*b*		+++	++		+
Fitoeno		-	+		+++
Licopeno		-	-		-
β-caroteno		+++	+		-
Zeaxantina		+++	+		-
ABA		+++	++		+
Luteína		+++	++		+
CRECIMIENTO Y DESARROLLO[§]					
Peso fresco		+++	+++		+
Longitud		+++	+++		+
Número de brotes		++	+++		+
DATOS EPIGENÉTICOS					
Metilación de ADN global		+++	++		+
H3K4me2		+	+		+++
H3K36me2		++	+		+++
H3K9ac		+	++		+++
H3K9me2		+	+		+++
H3K27me3		+	+		+++
OTRAS VARIABLES					
Malato (al término del periodo de oscuridad)		++	+++		+
Fenotipos que puede regenerar		G y V	G, V y A		A

*AV= región albina del fenotipo variegado; GV= región verde del fenotipo variegado. En el caso de disponer información de cada una de esas regiones, se anotó en el cuadro por separado. En caso contrario se registró como un dato del fenotipo V.Ad, adaxial; Ab, abaxial; [+], Datos obtenidos por microscopía electrónica de barrido; [§], los datos corresponden a plantas del quinto subcultivo. Información tomada de Duarte-Aké *et al*., (2016) y Us-Camas *et al*., (2017).

1.6. EL ALBINISMO, UNA ALTERACIÓN EN LA PIGMENTACIÓN VEGETAL

Usualmente, las plantas se caracterizan por presentar una única coloración verde en su superficie foliar. Sin embargo, algunas especies tanto mono como eudicotiledóneas, poseen alteraciones en la pigmentación de su tejido foliar. Entre estas destacan la clorosis (reducción uniforme en la pigmentación verde) (Li *et al.*, 2021), la variegación (Zhang *et al.*, 2020; Li *et al.*, 2019a) y el albinismo (**Figura 1.3**). Las plantas variegadas presentan regiones verdes acompañadas de patrones regulares o irregulares de diversos colores a lo largo de la superficie foliar (Zhang *et al.*, 2020). Por su parte, el albinismo se caracteriza por I) la ausencia total o parcial de la capacidad de producir pigmentos fotosintéticos como la clorofila y los carotenoides, II) un bloqueo en la biogénesis del cloroplasto, en la que los proplástidos no logran diferenciarse en cloroplastos, y III) un pobre o nulo desarrollo de las membranas tilacoidales que alojan los fotosistemas (Satou *et al.*, 2014; Kumari *et al.*, 2009).

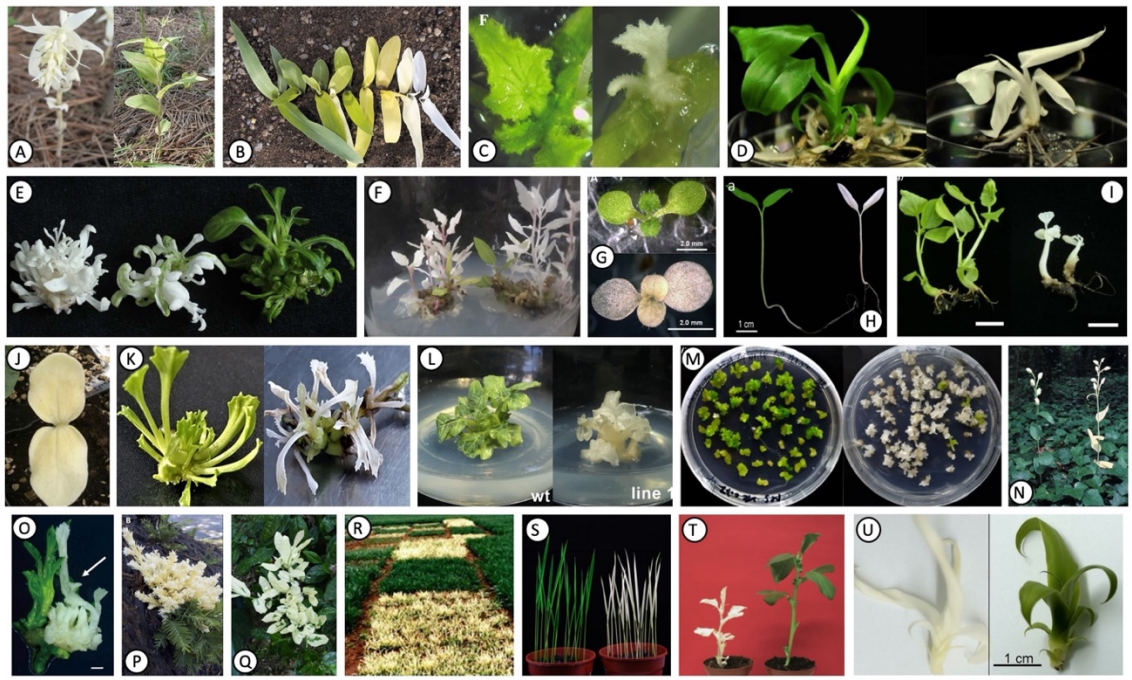


Figura 1.3. Plantas con fenotipo albino creciendo tanto en condiciones *in vitro* como *in vivo*. En la figura se presentan imágenes de una **A)** orquídea micoheterotrófica (*Epipactis helleborine*) (Suetsugu *et al.*, 2017), **B)** maíz (*Zea mays*) (Stern *et al.*, 2004), **C)** melón (*Cucumis melo*) (Hooghvorst *et al.*, 2019), **D)** banana (*Musa acuminata* Cavendish cv. Williams) (Naim *et al.*, 2018), **E)** petunia (*Petunia hybrid*) (Zhang *et al.*, 2016), **F)** álamo (*Populus tomentosa* Carr.) (Fan *et al.*, 2015), **G)** *Arabidopsis thaliana* (Fucile *et al.*, 2008), **H)** tomate (*Solanum lycopersicum*) (García-

Alcázar *et al.*, 2017), l) algodón (*Gossypium hirsutum*) (Wang *et al.*, 2018a), **J)** calabaza (*Cucurbita pepo*) (García *et al.*, 2018), **K)** col verde china (*Brassica oleracea* var. *alboglabra*) (Sun *et al.*, 2018), **L)** sandía (*Citrullus lanatus*) (Tian *et al.*, 2017), **M)** tabaco (*Nicotiana tabacum*) (Lin *et al.*, 2018) **N)** orquídea mixotrófica (*Cephalanthera damasonium*) (Julou *et al.*, 2005) **O)** manzana (*Malus prunifolia* (Wild.) Borkh. 'Seishi' × *M. pumila* Mill. var. *paradisiaca* Schneid. 'M.9') (Nishitani *et al.*, 2016), **P)** secuoya (*Sequoia sempervirens*) (Pittermann *et al.*, 2018), **Q)** tea plant (*Camellia sinensis*) (Lu *et al.*, 2019), **R)** Trigo (*Triticum aestivum*) (Hou *et al.*, 2009), **S)** arroz (*Oryza sativa*) (Su *et al.*, 2012), **T)** haba (*Vicia faba*) (Roelfsema *et al.*, 2006) y **U)** piña (*Ananas comosus* var. *bracteatus*) (Lin *et al.*, 2019). Las imágenes incluídas fueron tomadas de los artículos citados en cada inciso.

Centrándonos en el albinismo, esta es una variación poco común en el reino vegetal (Silva *et al.*, 2020). Si bien hay reportes de plantas albinas creciendo en condiciones naturales (Almagro-Armenteros *et al.*, 2019; Lu *et al.*, 2019; Pittermann *et al.*, 2018; Suetsugu *et al.*, 2017; Julou *et al.*, 2005), su presencia es mucho más frecuente como resultado de la aplicación de técnicas de CTV, particularmente durante la androgénesis de cereales (Canonge *et al.*, 2021). Existen reportes que indican que, en experimentos de esta naturaleza y con cultivares específicos, se ha obtenido una progenie casi 100% albina. Tal es el caso del cultivar de trigo primaveral Cork (Caredda *et al.*, 2000).

En las últimas décadas se ha tratado de elucidar los factores que podrían determinar la aparición de fenotipos albinos. Entre los factores asociados a su aparición podemos mencionar: el genotipo o cultivar (Caredda *et al.*, 2000), factores ambientales como la luz y la temperatura (Jiang et al., 2020; Yan et al., 2020), cambios en las concentraciones de reguladores del crecimiento (Dewir *et al.*, 2018), incompatibilidad entre el genoma del núcleo y el cloroplástico (Yao y Cohen, 2000), mutaciones en el ADN del plástido (Park *et al.*, 2022), alteraciones en las rutas de biosíntesis de la clorofila (Wang *et al.*, 2020a; Li *et al.*, 2018b; Shi *et al.*, 2017), entre otros. Sin embargo, a pesar de los avances hechos en el tema del albinismo vegetal, es evidente que aún se desconocen la(s) causa(s) que dan lugar a la aparición de esta anormalidad en la pigmentación.

Lo que queda claro es que las plantas con alteraciones en su pigmentación, particularmente las plantas albinas, tienen un desbalance severo sobre la tasa fotosintética (Satou *et al.*, 2014; Julou *et al.*, 2005). La evidencia de esto la tenemos tanto en ejemplares creciendo en la naturaleza como en condiciones *in vitro*. Para el caso de ejemplares albinos

en la naturaleza, la supervivencia de estas plantas depende de otros organismos. Ejemplos de esto lo tenemos en orquídeas albinas completamente micoheterotróficas del género *Epipactis* and *Cephalanthera*, cuya supervivencia es enteramente dependiente de los suministros de carbono obtenidos de la simbiosis con hongos micorrízicos (Suetsugu *et al.*, 2017; Julou *et al.*, 2005). Otro ejemplo es el de *Sequoia sempervirens*, que ocasionalmente genera follaje albino creciendo como holoparásito en otros troncos (Pittermann *et al.*, 2018). Para el caso particular de plantas albinas creciendo en condiciones *in vitro*, algunas de ellas pueden ser letales como la mutante de arroz *lethal albinic seedling 1* (*las1*) obtenida a través de CRISPR/CAS9 (Liu *et al.*, 2020). El daño en el metabolismo fotosintético y el bloqueo en la biogénesis del cloroplasto provoca que estas plantas mueran en etapas tempranas del desarrollo y solo logren sobrevivir con un medio de cultivo adecuado suministrando los nutrientes necesarios (Makowska y Oleszczuk, 2014; Kumari *et al.*, 2009). Los dos fenotipos con alteraciones en su pigmentación surgidos durante la micropropagación de plántulas de *A. angustifolia*, representan un modelo de estudio ideal para comprender aspectos relacionados con la regulación de la fotosíntesis, la biogénesis y el desarrollo del cloroplasto. Su idoneidad para esto se debe a que no son fenotipos letales, se micropropagan con cierta facilidad y no son el resultado de eventos de mutagénesis inducida.

1.7. EL CLOROPLASTO: EL PRINCIPAL CENTRO METABÓLICO DE LA CÉLULA VEGETAL

Los organismos fototróficos son esenciales para sustentar la vida en la tierra debido a su habilidad para realizar fotosíntesis. Para llevar a cabo esta función, las plantas cuentan con centros metabólicos activos conocidos como cloroplastos (Füssy y Oborník, 2018; Daniell *et al.*, 2016). Estos organelos alojan complejos de captación de luz (*light harvesting complex* o LHC, por sus siglas en inglés), que transfieren la energía obtenida de los fotones a través del aparato fotosintético alojado en las membranas tilacoidales, creando un gradiente de protones, NADPH y ATP para la fijación de CO_2 (Füssy y Oborník, 2018; Solymosi *et al.*, 2018).

1.7.1. Características estructurales del cloroplasto vegetal

Los cloroplastos son organelos en forma de lente que tienen un diámetro promedio de 5 μm y un grosor de 2.5 μm (Staehelin, 2003). Un cloroplasto maduro está delimitado por tres sistemas de membranas: la membrana externa, la membrana interna y el sistema de membranas tilacoidales (Kirchhoff, 2019). La región delimitada por la membrana externa y la interna es conocida como espacio intermembranal. Ambas membranas son claves en la importación de proteínas del núcleo al cloroplasto y se encuentran conformadas principalmente por galactolípidos, y en menor medida por sulfo- y fosfolípidos. (Solymosi *et al.*, 2018). Entre la membrana interna y las membranas tilacoidales se encuentra el estroma. Esta matriz acuosa es metabólicamente muy activa, alojando las reacciones metabólicas independientes de luz, destacando las enzimas del Ciclo de Calvin-Benson-Bassham y la ribulosa-1,5-bisfosfato carboxilasa/oxigenasa (RuBisCO) para la fijación del CO_2. Aquí también ocurre la asimilación de azufre y nitrógeno. El estroma también aloja los ribosomas 70S, al nucleoide y los gránulos de almidón, entre otros (Kirchhoff, 2019; Solymosi *et al.*, 2018).

El cloroplasto también exhibe una serie de membranas internas conocidas como tilacoides. Estas membranas se componen de lípidos, proteínas y pigmentos como las clorofilas y los carotenoides que proveen la plataforma para las reacciones dependientes de la luz de la fotosíntesis. Estas reacciones convierten la luz solar en un transporte activo de electrones y un gradiente de protones para producir NADPH y ATP, los cuales se acumulan en el estroma para las reacciones metabólicas que ahí se realizan (Kirchhoff, 2019; Pribil *et al.*, 2014). En un estadio maduro del cloroplasto, las membranas tilacoidales tienden a apilarse en estructuras conformando la grana, las cuales se interconectan a través de las lamelas. Es por esto que se reconoce que la membrana de los tilacoides conforma una estructura en forma de red tridimensional continua. Por último, las membranas tilacoidales delimitan un compartimento acuoso conocido como espacio lumenal o lumen, cuya principal función es acumular el gradiente electroquímico que permite la síntesis de ATP (Kirchhoff, 2019; Solymosi *et al.*, 2018; Pribil *et al.*, 2014) (**Figura 1.4A**). Cuando las membranas tilacoidales no logran conformar la grana como sucede en el etioplasto, estas adquieren una conformación diferente que se caracteriza por presentar cuerpos prolamelares y protilacoides (Floris y Kühlbrandt, 2021; Solymosi *et al.*, 2018). Una

alteración en el proceso de diferenciación de los plástidos que afecte su ultraestructura final, puede ser un factor determinante que comprometa su metabolismo autotrófico.

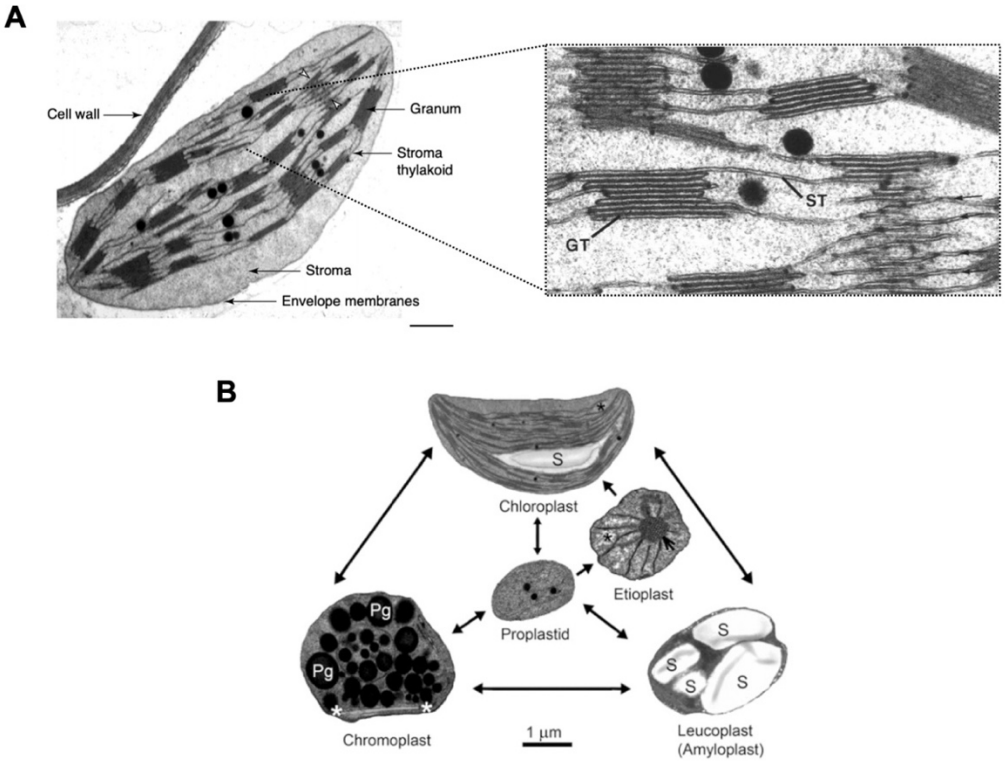


Figura 1.4. Ultraestructura del cloroplasto y miembros de la familia de los plástidos. A) En la imagen de la izquierda se presenta una microfotografía electrónica del cloroplasto de una planta superior señalando las diversas estructuras que lo componen. Escala de la barra= 0.5 µm. En la imagen de la derecha se amplifica una porción donde se distingue la grana (GT) y los tilacoides del estroma (ST). Imágenes tomadas y editadas de Mustárdy y Garab (2003); Staehelin (2003). **B)** Ultraestructura típica e interconversiones de los principales tipos de plástidos. Pg: plastoglóbulos, S: granos de almidón, los asteriscos blancos están localizados en los dos puntos terminales de un cristal de ß-caroteno, asteriscos negros: regiones nucleoides, flechas: cuerpos prolamelares. Imágenes tomadas a través de un microscopio electrónico de transmisión. Imagen tomada de Solymosi *et al.* (2018).

1.7.2. Funciones generales del cloroplasto vegetal

El cloroplasto es conocido por su protagonismo en el metabolismo fotosintético y su clásica coloración verde asociada a la presencia de clorofila. Sin embargo, su importancia no se reduce solo esto, ya que también desempeña papeles claves en la fisiología de la planta como lo son: el metabolismo del nitrógeno y el azufre, la biosíntesis de compuestos

esenciales como AA, nucleótidos, ácidos grasos, tetrapirroles, isoprenoides, fitohormonas, vitaminas, además de que participa en la respuesta al estrés abiótico (respuestas a calor, sequía, sal, luz, etc.) y biótico (en la defensa contra patógenos) (Cackett *et al.*, 2022; Füssy y Oborník, 2018; Daniell *et al.*, 2016).

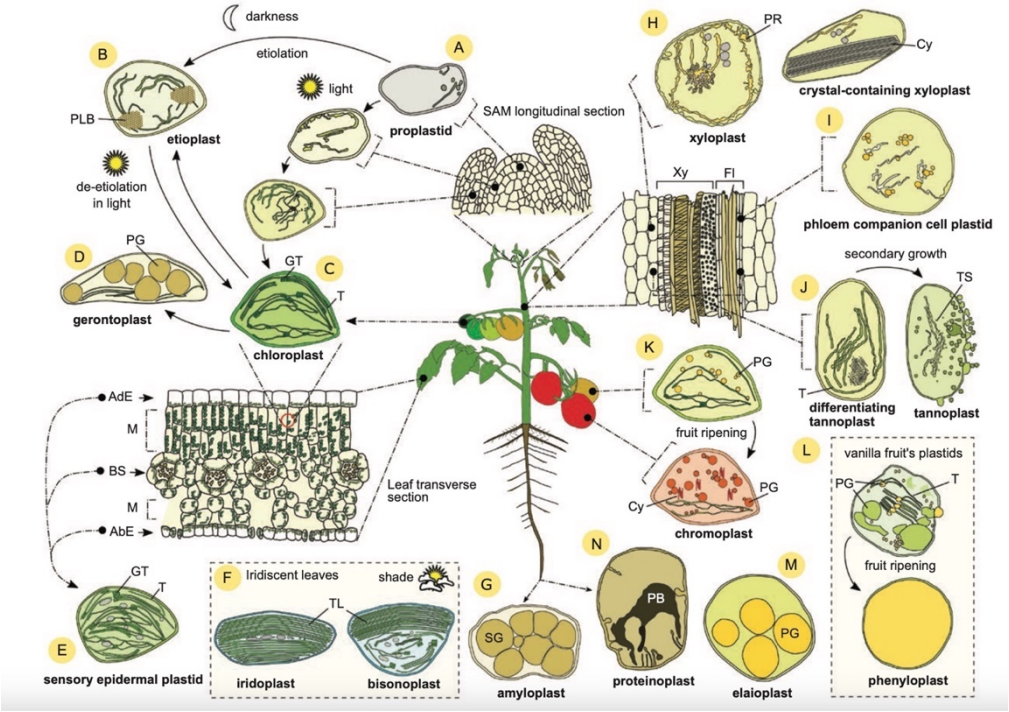


Figura 1.5. Miembros de la familia de los plástidos en plantas y sus interconversiones. A) Proplástidos, B) etioplastos, C) cloroplastos, D) gerontoplastos, E) plástidos epidérmicos sensoriales, F) iridoplastos, G) amiloplastos, H) xiloplastos, I), J) tanoplastos, K) cromoplastos, L) feniloplastos, M) elaioplastos y N) proteinoplastos. PLB: cuerpos prolamelares, PG: plastoglobulos, GT: grana, T: tilacoides, TL: lamelas tilacoidales, SG: granos de almidón, PB: cuerpos proteicos, Cy: cristales, TS: estructura del tanosoma, Xy: xilema secundario, Fl: floema y PR: estructura del retículo periférico. Imagen tomada de Sierra *et al.* (2023).

El cloroplasto solo es uno de los numerosos miembros que conforman la familia de los plástidos. Cada uno de los miembros de esta familia presenta una morfología y funciones metabólicas especializadas que se adecuan a momentos claves del desarrollo vegetal y de las condiciones ambientales (Sierra *et al.*, 2023). Entre ellos podemos mencionar a los cromoplastos, leucoplastos, amiloplastos, proteinoplastos, elaioplastos, etioplastos, gerontoplastos, entre otros. Para más detalles sobre la diversidad de miembros

de esta familia y sus funciones, consultar a Sierra *et al.* (2023) (**Figura 1.5**). Sin embargo, todo ellos derivan de un plástido inmaduro, no diferenciado, incapaz de realizar la fotosíntesis, incoloro y muy abundante en los tejidos meristemáticos de las plantas (Sierra *et al.*, 2023; Rolland *et al.*, 2018; Solymosi *et al.*, 2018). Si bien, cada tipo de plástido tiene funciones únicas y específicas en ciertos momentos del desarrollo vegetal, su capacidad de interconversión en respuesta a señales internas y externas añade una mayor complejidad en la caracterización de sus funciones (Sierra *et al.*, 2023; Solymosi *et al.*, 2018) (**Figura 1.4B**).

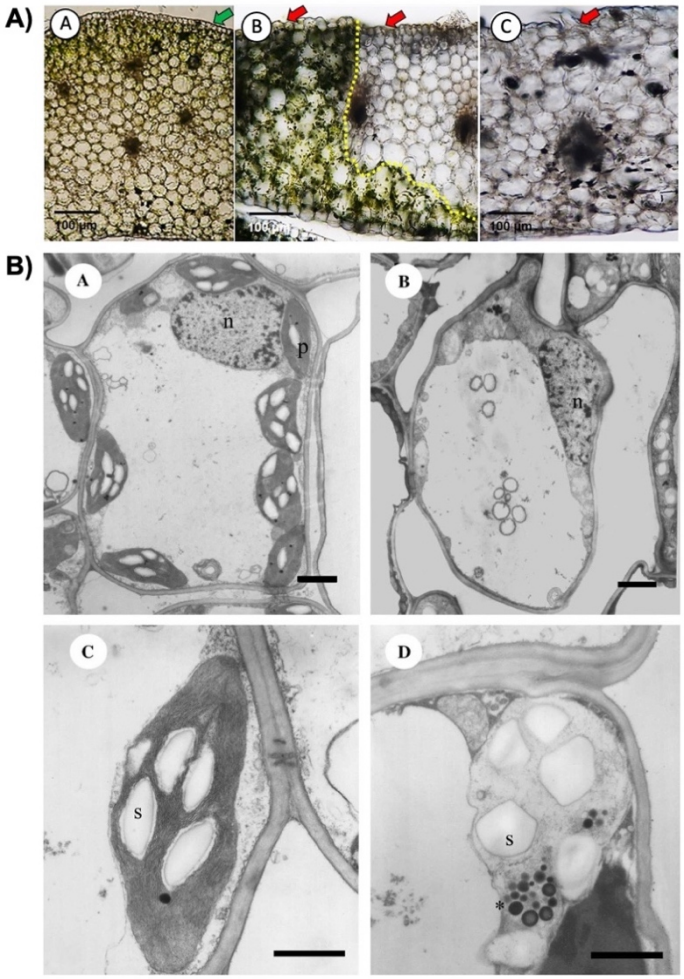


Figura 1.6. Distribución y ultraestructura de plástidos en hojas de las tres variantes somaclonales de *A. angustifolia* y *Bambusa edulis*. A) Presencia de los plástidos en cortes de hojas de los fenotipos verde (G), variegado (V) y albino (A) de *A. angustifolia*. En el fenotipo G es posible distinguir que los cloroplastos se distribuyen a lo largo de la hoja por el característico color

verde dado por la clorofila. En el fenotipo V los cloroplastos se encuentran distribuidos solo en la sección verde y no en la albina. Por último, en el fenotipo A no fue posible identificar cloroplastos maduros. Imagen tomada de (Hernández-Castellano *et al.*, 2020). **B)** Morfología de los plástidos presentes en hoja verde (A) y albina de (B) *B. edulis* vistos en el microscopio electrónico de transmisión. Las hojas verdes contienen células con núcleo (n) y cloroplastos (p) con tilacoides bien organizados en grana así como gránulos de almidón (S) y plastoglóbulos (*). Las hojas albinas contienen plástidos sin tilacoides formados y se observan gránulos de almidón y un mayor contenido de plastoglóbulos. Barras= 1 μm. Imagen tomada de Lin *et al.* (2008).

En plantas con fenotipos albinos se tiene conocimiento que los proplástidos tienen problemas durante su proceso de diferenciación, resultando en cloroplastos anormales o ausentes y con pocas o sin membranas tilacoidales, lo que les impide mantener su metabolismo fotosintético cotidiano y su crecimiento autotrófico (**Figura 1.6 A** y **B**). Por el contrario, los plástidos de plantas albinas contienen un mayor número de plastoglóbulos, reducción en los gránulos de almidón, una estructura lamelar desorganizada y un mayor volumen vacuolar (Park *et al.*, 2022; Liu *et al.*, 2020; Zhou *et al.*, 2017; Satou *et al.*, 2014; Clarke *et al.*, 2011; Lin *et al.*, 2008; Liu *et al.*, 2007).

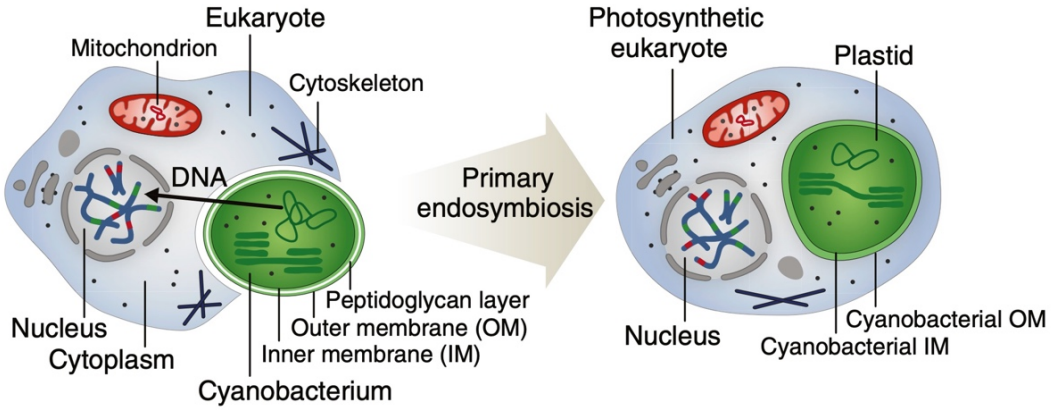


Figura 1.7. Endosimbiosis primaria de una cianobacteria primitiva por una eucariota no fotosintética. Imagen tomada y modificada de Archibald (2015).

1.7.3. El origen del cloroplasto: la hipótesis de la endosimbiosis

El cloroplasto es considerado un organelo "semiautónomo" dentro de la célula eucariota vegetal (Dobrogojski *et al.*, 2020). Esta cierta autonomía se sustenta con el hecho que este organelo contiene su propio ADN genómico –conocido como plastoma– que en el

caso de *A. angustifolia* presenta un tamaño de alrededor de 157 Kb. Como la mayoría de los plastomas de angiospermas, este codifica genes claves para la fotosíntesis, la transcripción y la traducción del plástido (Qin *et al.*, 2021). Adicionalmente, los plástidos tienen la capacidad de dividirse de una manera similar a la fisión binaria vista en organismos procariotas, en un modo independiente de la división de la célula (Solymosi *et al.*, 2018). Estas características parecen revelar una cierta autonomía de los plástidos. No obstante, estos organelos son incapaces de llevar a cabo todas sus funciones independientemente.

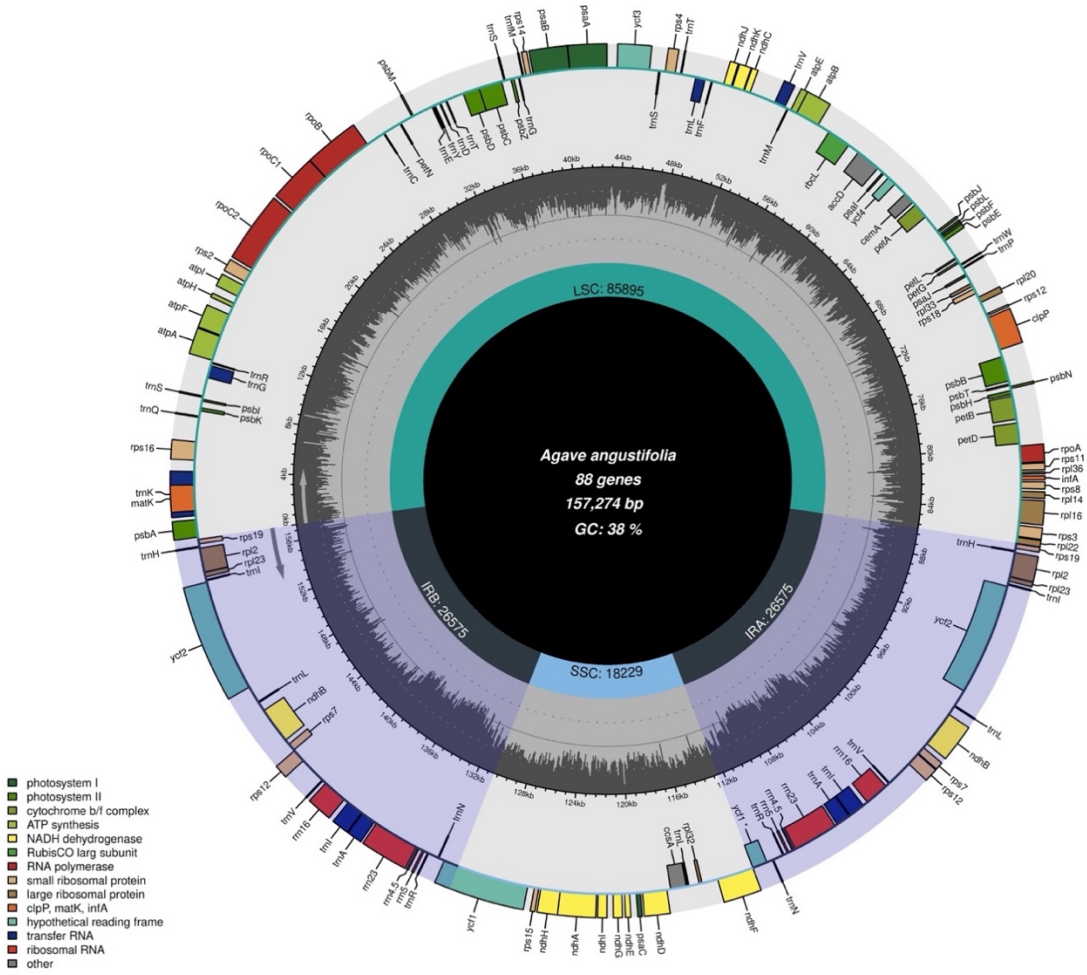


Figura 1.8. Mapa del genoma del cloroplasto de *A. angustifolia*. Genes con funciones similares son agrupados con un mismo color. El esquema fue generado a partir del número de accesión MW540498.1 (GenBank-NCBI) que corresponde al genoma del cloroplasto de *A. angustifolia* reportado por Qin *et al.* (2021) utilizando el programa CHLOROPLOT (Zheng *et al.*, 2020). Los números después de las regiones del plastoma indican su longitud en pb. SSC: región corta de copia

única, IRA y IRB: regiones repetidas invertidas A y B, LSC: región larga de copia única, bp: pares de bases, GC: porcentaje de guanina y citosina, *: pseudogen.

La vida como se conoce actualmente, solo puede entenderse como resultado de un singular evento que sucedió hace alrededor de 1,200 millones de años, el cual fue crucial en el proceso evolutivo de los organismos eucariotas, particularmente el de los autótrofos fotosintéticos (Bölter, 2018; Zimorski *et al.*, 2014). Este evento, conocido como endosimbiosis primaria tuvo como protagonista a una célula eucariota heterótrofa con núcleo y mitocondrias, estas últimas adquiridas previamente por la fagocitosis de una α-proteobacteria (Maréchal, 2018; Archibald, 2015). Esta hipótesis plantea que esta eucariota unicelular primigenia incorporó en su citoplasma a una cianobacteria fototrófica antigua de vida libre (con aparato fotosintético completo), la cual se estableció en el interior de la célula, y con el paso de los años evolucionaría en el cloroplasto (**Figura 1.7**) (Archibald, 2015). Las principales evidencias que apoyan el origen endosimbiótico de los cloroplastos son: la transferencia horizontal de genes de la cianobacteria primitiva al núcleo de la célula hospedera, la similitud entre las secuencias de ARNr cloroplásticos y bacterianos, los coeficientes de sedimentación de 70S de sus ribosomas, y la presencia de un ADN genómico propio y desnudo (sin histonas) con alta similitud al de los genomas de las cianobacterias (Allen, 2018; Maréchal, 2018; Bock, 2017).

Aunque los cloroplastos conservaron su propio genoma aún después del evento endosimbiótico, esto no significó una independencia genética y metabólica respecto de la célula hospedera. Aunque no existen estudios que estimen el número de proteínas cloroplásticas en *A. angustifolia*, en otras especies modelo como *A. thaliana* se estima que este se compone de entre 2,100 y 3,600 proteínas, aunque estudios más recientes fijan este número en alrededor de los 3,000. De este número, más del 90% son proteínas codificadas por el núcleo e importadas postraduccionalmente al cloroplasto (Zoschke y Bock, 2018; Yu *et al.*, 2008; Leister, 2003).

Recientemente se secuenció el plastoma de *A. angustifolia* y se encontró que este codifica un total de 132 genes, de los cuales 86 codifican proteínas, 38 ARNt y 8 ARNr. Entre los genes que codifican proteínas se encuentran aquellos que participan en el ensamblaje de los fotosistemas I y II (PSI y PSII) (*photosystem I* y *II*, por sus siglas en inglés), los complejos del citocromo b/f, ATP sintasa, NADH deshidrogenasa, la subunidad

mayor de la RuBisCO, subunidades de la ARN polimerasa codificada por el plástido (PEP) (*plastid-encoded RNA polymerase*, por sus siglas en inglés), genes ARN de la subunidad menor (SSU) (*small subunit*, por sus siglas en inglés) y mayor (LSU) (*large subunit*, por sus siglas en inglés) del ribosoma así como otros de función desconocida (**Figura 1.8**) (Qin *et al.*, 2021). El hecho que las proteínas cloroplásticas sean codificadas por genes alojados tanto en el genoma nuclear como en el plastoma, confirma la fuerte codependencia entre ambos sitios subcelulares.

Posterior a la transferencia de genes al núcleo, fue necesario un perfeccionamiento entre la comunicación entre el genoma nuclear y el plastídico. Este proceso coevolutivo implicó la pérdida de información genética prescindible y redundante, el establecimiento de mecanismos de importación de proteínas y metabolitos citoplásmicos al plástido, así como mecanismos de señalización desde el núcleo al plástido y viceversa (Bock, 2017).

1.7.4 Comunicación entre el núcleo y el cloroplasto

La comunicación que existe entre el núcleo celular y el cloroplasto es un proceso que debe ser finamente regulado. Es por esto que el intercambio continuo de señales entre ambos sitios involucra vías de comunicación que van en ambas direcciones, es decir, desde el núcleo al cloroplasto y viceversa (Fey *et al.*, 2005).

La primera de estas vías de señalización, conocida como señalización anterógrada, tiene su origen en el núcleo y se dirige al cloroplasto. Esta involucra reguladores de eventos transcripcionales, postranscripcionales y traduccionales en el plástido. Entre estos podemos mencionar: I) los factores sigma (SIG), II) la ARN polimerasa codificada por el núcleo (*nuclear-encoded RNA polymerase* o *NEP*, por sus siglas en inglés), III) ciertos genes que codifican para proteínas asociadas al nucleoide plastídico (*plastid nucleoid-associated proteins* o *pTACs*, por sus siglas en inglés), IV) genes nucleares asociados a la fotosíntesis (*photosynthesis-associated nuclear genes* o *PhANGs*, por sus siglas en inglés), V) proteínas ribosomales, V) proteínas con repeticiones tetratricopeptídicas (*tetratricopeptide repeats* o *TPR*, por sus siglas en inglés), VI) proteínas con repeticiones pentatricopeptídicas (*pentatricopeptides repeats* o *PPR*, por sus siglas en inglés), entre otros (Jan *et al.*, 2022; Berry *et al.*, 2013).

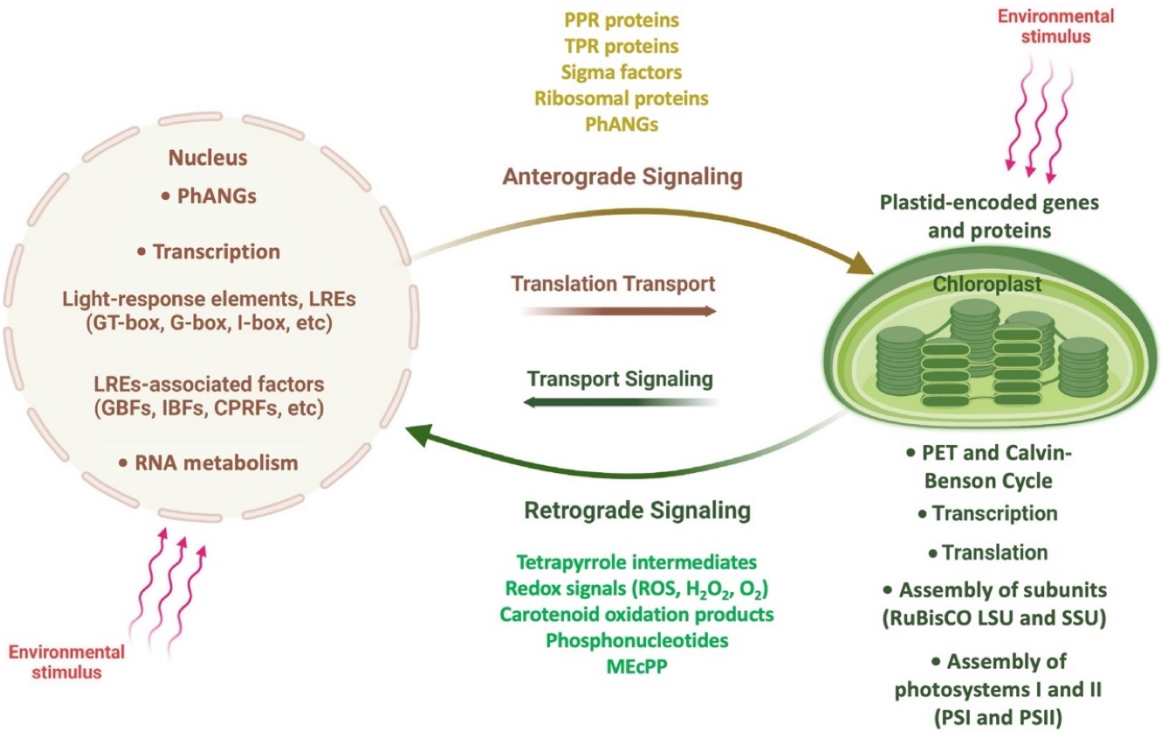


Figura 1.9. Mecanismos de señalización anterógrada y retrógrada actuando entre el núcleo y el plástido. Imagen tomada y modificada de Jan *et al.* (2022).

Por su parte, la señalización retrógrada es un proceso de retroalimentación que va del plástido al núcleo. En esta ruta, la transcripción de genes nucleares es regulada a partir de señales plastídicas que comunican al núcleo el estado metabólico, de desarrollo y de estrés en el que se encuentra el organelo (Jung y Chory, 2009; Shyjan y Buttow, 1993). Las principales rutas de señalización retrógrada reportadas son: I) la acumulación de intermediarios de la ruta de tetrapirroles, II) ciertos productos de oxidación de la ruta de biosíntesis de carotenoides, III) señales generadas de la expresión de genes plastídicos, IV) las alteraciones en el estado redox del plástido y la producción de especies reactivas de oxígeno, V) la ruta de señalización SAL1-PAP (3-fosfoadenosina 5´-fosfato) y VI) el precursor de isoprenoides 2-C-metil-D-eritritol 2,4-ciclodifosfato (MEcPP), derivado de la vía del metileritrol fosfato (*methylerythritol phosphate pathway* o *MEP*, por sus siglas en inglés), entre otros (Jan *et al.*, 2022; Jiang y Dehesh, 2021; Chi *et al.*, 2015a). Recientemente se ha propuesto una tercera ruta de señalización conocida como corrección

anterógrada núcleo-cloroplasto o comunicación retro-anterógrada. Esta tercera ruta fue descrita en la mutante abigarrada *cue8* de *A. thaliana* y se caracteriza por activarse en respuesta a una falla en la biogénesis del cloroplasto (Loudya *et al.*, 2020). Más detalles de estas vías de señalización núcleo-cloroplasto serán discutidas en el siguiente capítulo.

Es reconocido que tanto la transcripción como la maduración del ARN, la traducción de proteínas y sus modificaciones tienen un impacto en la biogénesis y desarrollo del cloroplasto (Pogson y Albrecht, 2011). Un ejemplo muy claro sobre la importancia de la co-evolución núcleo-cloroplasto se determinó al comparar los genomas de los ancestros cianobacterianos respecto al de los cloroplastos modernos. Esta comparación reveló que el aparato transcripcional y los mecanismos postranscripcionales de los cloroplastos de las plantas superiores son más complejos que en procariotas (Börner *et al.*, 2015). Esto parece estar relacionado con las características de sus ARNs, ya que para su maduración y para poder ser traducidos pueden requerir: I) cis y/o trans-procesamiento para eliminar intrones del grupo I o II, II) edición en sus secuencias, III) protección contra ARNasas mediante la formación de horquillas en sus extremos 3’ o 5’, y IV) cortes en sus extremos terminales que permitan su maduración (en caso de que hayan sido transcritos en un operon en forma de ARN policistrónico y necesiten ser cortados en ARNs monocistrónicos antes de su traducción) (Börner *et al.*, 2015). Recientemente se ha demostrado que una familia de proteínas codificadas en el núcleo y conocidas como proteínas PPR, exhiben muchas de las funciones mencionadas anteriormente, las cuales son específicas y críticas en el metabolismo del ARN y por consiguiente, en la conformación y funcionalidad del cloroplasto (Pogson y Albrecht, 2011).

1.9. PROTEÍNAS CON REPETICIONES DE PENTATRICOPEPTÍDICAS COMO ACTORES CLAVE EN LA BIOGÉNESIS DEL CLOROPLASTO

1.9.1. Descubrimiento y evolución de las proteínas PPR

Con el nacimiento de los proyectos de secuenciación de genomas eucariotas a finales del siglo XX y principios del XXI, una gran cantidad de familias de genes desconocidos para la comunidad científica internacional empezaron a volverse el centro de atención (Small y Peeters, 2000). Bajo este nuevo enfoque ómico se logró generar el genoma completo de *A. thaliana*, lo que permitió el descubrimiento de una familia de genes

hipotetizada –y años más tarde confirmada– como la más numerosa en plantas, de funciones desconocidas y con características estructurales que le dieron el adjetivo de "misteriosa" (Lurin *et al.*, 2004). Fueron Small y Peeters (2000) quienes bautizaron a esta familia como genes con repeticiones pentatricopeptídicas o simplemente genes PPR (*pentatricopeptide repeats*, por sus siglas en inglés). En *A. thaliana* esta familia representa alrededor del 2% de los genes que codifican para proteínas (Nakamura *et al.*, 2012; Lurin *et al.*, 2004). Si bien estas proteínas se encuentran ampliamente distribuidas en el genoma de los eucariotas, el número de proteínas PPR presentes en organismos vegetales es mucho mayor que en cualquier otro grupo (Cushing *et al.*, 2005).

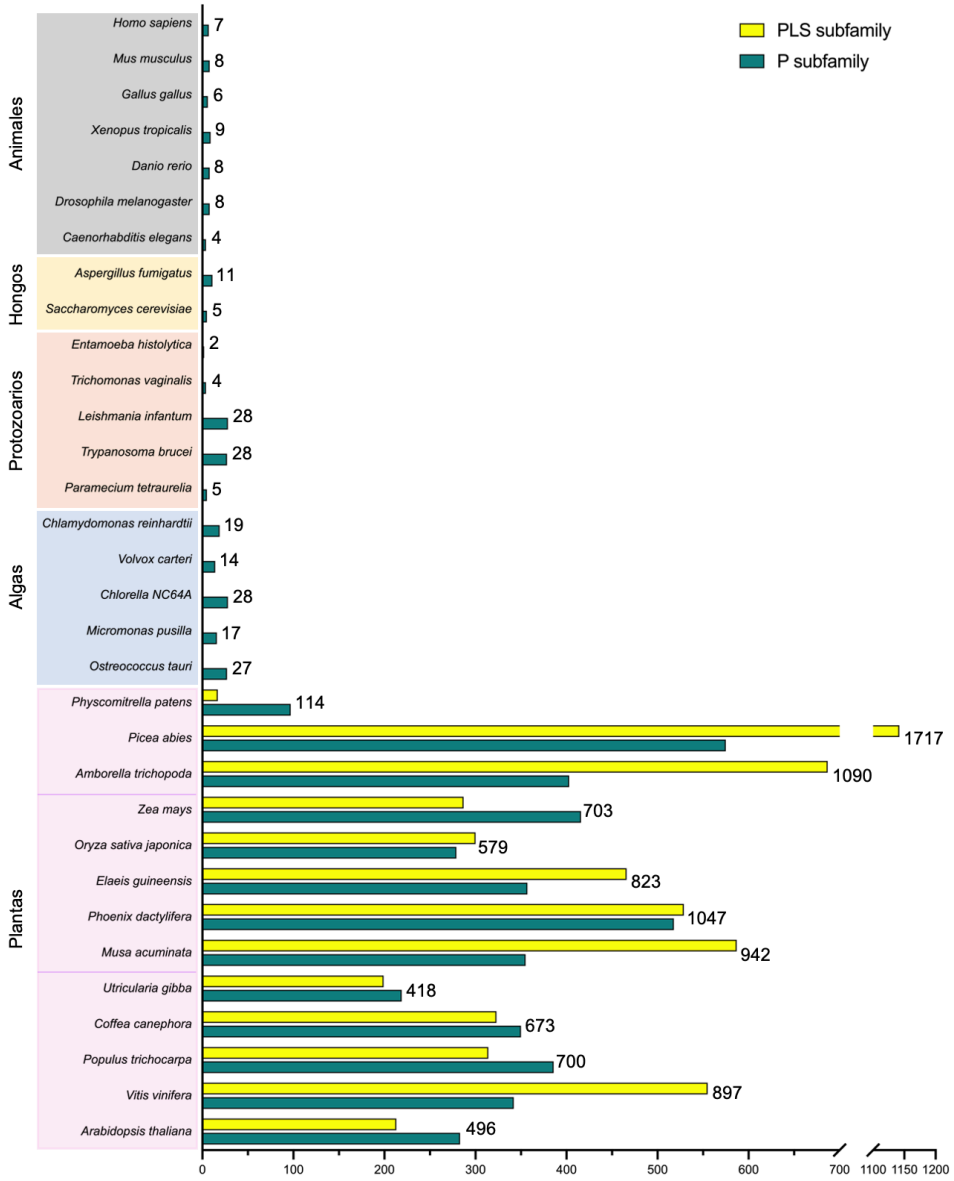


Figura 1.10. Distribución de genes PPR en genomas de diversos eucariotas. El número de proteínas PPR de la subfamilia P son indicadas con barras color amarillo mientras que el de la subfamilia PLS con barras color azul. El total de genes se indica del lado derecho de las barras. Información tomada de Cheng *et al.* (2016).

El proceso evolutivo que ha dado lugar a la expansión inicial de esta familia de genes en plantas superiores, aunque enigmática, parece indicar que la expansión y diversificación de esta familia ocurrió a partir del origen de las plantas terrestres (hace 475 millones de años) pero antes de la divergencia de espermatofitas y angiospermas modernas de las briofitas (hace 140 millones de años). En este punto específico numerosos eventos de retrotransposición parecen ser la causa de esta expansión en las angiospermas (Gutmann *et al.*, 2020; O'Toole *et al.*, 2008).

El número de proteínas PPR en hongos, animales y algunos protozoarios ronda entre 2 y 28 miembros. En microalgas, estos números parecen incrementarse alcanzando hasta 28 miembros en algunas especies. Sin embargo, en las plantas se registra una fuerte expansión en el número de proteínas PPR. Por ejemplo, para el musgo *Physcomitrella patens se* determinaron 114 miembros, 496 para *A. thaliana* y 579 para *O. sativa* var. *japonica*. Algo que llama la atención es el hecho que, a pesar de que las angiospermas tienen variaciones marcadas entre el tamaño de sus genomas y el contenido total de genes, el número de proteínas PPR es relativamente constante (entre los 500 y los 900 miembros) (**Figura 1.9**). Excepciones a esta regla la tenemos en la licofita *Selaginella moellendorffii* con alrededor de 5,115 proteínas PPR (Cheng *et al.*, 2016). En las especies del género *Gossypium* los niveles de ploidía también afectan el contenido final de genes PPR. Especies diploides como *G. arboreum* y *G. raimondii* exhiben la mitad de genes PPR respecto a las especies tetraploides *G. hirsutum* y *G. barbadense* (Han *et al.*, 2020).

Una característica notable para los genes de esta familia es que rara vez presentan intrones dentro de su secuencia codificante. Por ejemplo, *A. thaliana* exhibe alrededor de 80% de sus proteínas PPR con un solo exón y sin ningún intrón, el 12% contiene un intrón y el 8% dos o más intrones. Por su parte, el 65% de los genes PPR en *O. sativa* se conforman de un exón sin intrones, 16% tienen un intrón y el 19% más de uno (Chen *et al.*, 2018a). La ausencia de intrones en la mayoría de los miembros de esta familia de genes

permite explicar su corta longitud (<2 kb desde el ATG al codón de paro) que contrasta con una mayor longitud a nivel de proteínas (642 AA en promedio) (Lurin *et al.*, 2004).

1.9.2. Características estructurales de las proteínas PPR

Las proteínas PPR se caracterizan por presentar repeticiones en tándem de motivos específicos de esta familia. Esta región es conocida como motivo pentatricopeptídico, una región degenerada conformada de 35 residuos de AA (Lurin *et al.*, 2004; Small y Peeters, 2000). Estudios recientes proponen que, dependiendo del tipo de motivo PPR, su longitud varía en un rango de entre 30 y 40 residuos (Cheng *et al.*, 2016). Las proteínas PPR presentan de dos a veintiséis motivos PPR, con un promedio de 12 motivos por proteína, en un único arreglo en tándem que abarca al menos 2/3 partes de la proteína (Lurin *et al.*, 2004).

Desde su descubrimiento, se suscitó cierta controversia debido a que los motivos presentes en proteínas PPR tenían una alta similitud con las proteínas con repeticiones tetratricopeptídicas (*tetratricopeptide repeats* o TPR, por sus siglas en inglés). Estas últimas presentan motivos en tándem y cada motivo ensambla un par de α-hélices antiparalelas (conocidas como hélices A y B) cercando un surco capaz de regular interacciones proteína-proteína (Bohne *et al.*, 2016; Small y Peeters, 2000). Las proteínas PPR comparten las características estructurales mencionadas anteriormente, ya que ambas conforman una estructura tipo solenoide (Bohne *et al.*, 2016). A diferencia de las proteínas TPR, las proteínas PPR no están presentes en los procariotas y predominan en los eucariotas, principalmente en plantas. Otra diferencia se identifica a nivel estructural en la hélice A de las PPR, la cual está dirigida hacia el interior del surco y participa en el sitio de unión al ligando. Sobre este último punto destaca que en las proteínas TPR los residuos de la hélice A varían considerablemente lo que podría reflejar la variedad de ligandos de unión. Por su parte, en la hélice A de las PPR, las cadenas laterales que recubren el surco central son casi exclusivamente hidrofílicas, lo que explica su unión a nucleótidos de ARN y en menor proporción, de ADN. Otra característica distintiva entre ambas familias, es que las proteínas PPR tienen, en promedio, el doble de repeticiones respecto a las TPR (las cuales alojan entre 3 y 16 motivos) (Cheng *et al.*, 2016; Saha *et al.*, 2007; Small y Peeters, 2000).

Aunque el origen de genes *PPR* aún genera contradicciones, existe más información para los genes *TPR* cuya presencia en genomas procariotas y eucariotas sustenta un origen antiguo, el cual parece preceder a la aparición de otras proteínas similares. Por la similitud entre los motivos PPR y TPR, se cree que los primeros han surgido como resultado de la divergencia de motivos TPR (Sharma y Pandey, 2016; Small y Peeters, 2000).

1.9.3. Clasificación de las proteínas PPR en función de la diversidad de sus motivos

Para su clasificación, las proteínas PPR han sido divididas en dos subfamilias en función de la organización y variabilidad de los motivos que presentan: la subfamilia P (que conforma el modelo clásico) y la subfamilia PLS, la cual es únicamente encontrada en plantas terrestres (**Figura 1.10**) (Saha *et al.*, 2007; Rivals *et al.*, 2006; Lurin *et al.*, 2004). Actualmente se han definido 10 variantes del motivos PPR (Cheng *et al.*, 2016).

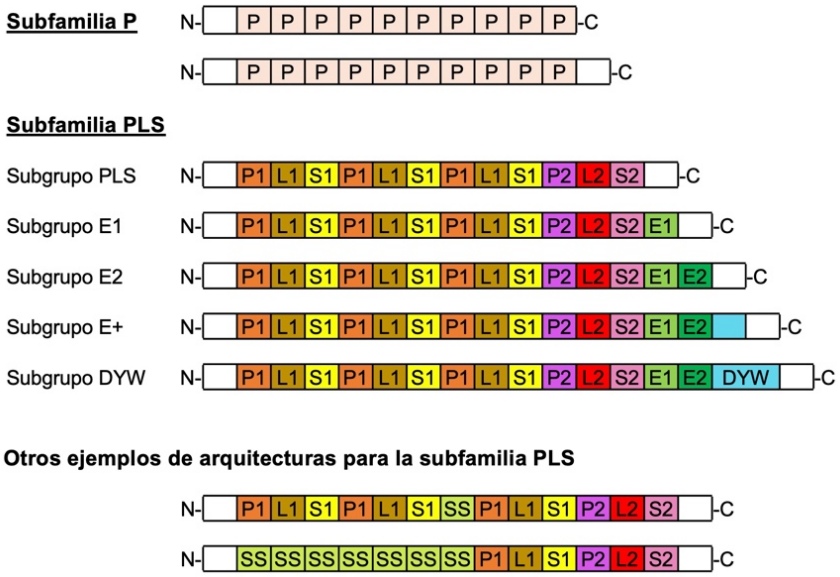


Figura 1.11. Representación estructural de las proteínas PPR. La subfamilia P compromete repeticiones en tándem de motivos P. La subfamilia PLS contiene tripletes característicos de motivos P1, L1 ("L" de long, generalmente de 36 residuos) y S1 ("S" de short, generalmente de 31 residuos) y dominios C-terminales característicos. El dominio E (del inglés *extended*) encontrado en la mayoría de las PPR del subgrupo PLS, se divide en motivos más pequeños conocidos como E1 y E2 (aunque algunos autores los conglomeran como una región entera). El dominio DYW en el C-terminal (bautizado así por su característico tripéptido aspartato-tirosina-triptofano en el C-terminal) se encuentra únicamente en un subgrupo de la subfamilia PLS. Imagen modificada de Cheng *et al.* (2016).

La subfamilia P se caracteriza por presentar únicamente motivos P clásicos adyacentes uno de otro, sin espacios entre ellos y sin motivos o dominios adicionales en el C-terminal. Esta subfamilia es la más antigua y se encuentra presente en todos los eucariotas (Gutmann *et al.*, 2020; Barkan y Small, 2014; Saha *et al.*, 2007). Por su parte, la subfamilia PLS tiene una organización modular más compleja que la anterior ya que se compone de un triplete de motivos conocidos como P-L-S (**Figura 1.10**) (Saha *et al.*, 2007; Lurin *et al.*, 2004). El arreglo clásico de una proteína de esta subfamilia sigue el patrón $(P_1\text{-}L_1\text{-}S_1)_n\text{-}P_2\text{-}L_2\text{-}S_2$ (Cheng *et al.*, 2016; Rivals *et al.*, 2006). Dos variantes del motivo P clásico han sido descritos en la subfamilia PLS: P1 y P2 (ambos de una longitud ~35 AA). Los motivos P1 no muestran una distribución en tándem como en la subfamilia P, pues entre ellos aparece una región de alrededor de 65 a 70 residuos de AA, donde se ubican dos variantes de motivos conocidos como S1 y L1 (**Figura 1.11**). El motivo S1 tiene una longitud aproximada de 31 residuos de aminoácidos (Cheng *et al.*, 2016; Lurin *et al.*, 2004). Visto en conjunto, el bloque P1, L1 y S1 se organiza de tal forma que se presenta en tándem de hasta siete copias por proteína. En la parte final de esta región, suele aparecer un sitio homólogo al bloque P1-L1-S1 que difiere en secuencia y en su origen evolutivo y que se conoce como P2-L2-S2. En este último bloque, el motivo L2 tiene una longitud de alrededor de 36 AA y el S2 de cerca de 32 (Cheng *et al.*, 2016; Rivals *et al.*, 2006). Justo después de un motivo S suele aparecer un motivo tipo S de alrededor de 31 residuos de AA, conocido como SS. (**Figura 1.11**). En la mayoría de los casos el motivo SS se encuentra entre los motivos S1 o SS y SS, P1 o P2 o en tándem de varios motivos SS (Cheng *et al.*, 2016).

En la región C-terminal de las proteínas PPR de la subfamilia PLS y posterior a la presencia del bloque P2-L2-S2, pueden aparecer de uno a tres regiones no relacionadas a los motivos PPR clásicos, conocidos como E, E+ y DYW. Estas, aunque no se presentan en tándem ni en copias múltiples, si tienen un orden de aparición ordenada, colinear y consecutiva en las proteínas (E–E+–DYW). Sin embargo, en algunas proteínas es posible encontrar individualmente la región E y la combinación E y E+ (Lurin *et al.*, 2004). La última actualización en la organización estructural de estas proteínas redefinió esta región como E1, E2 y DYW. Los motivos E1 y E2 tienen una extensión de 34 residuos (**Figura 1.11**) mientras que el DYW una longitud de 136. Este último debe su nombre a la presencia de los aminoácidos aspartato (D), tirosina (Y) y triptofano (W). El motivo DYW presenta tres regiones conservadas: la caja PG, el sitio activo y el C-terminal (**Figura 1.12**). Sin embargo,

los motivos DYW suelen estar truncados o degenerados en el extremo C-terminal, alrededor del décimo y sexagésimo residuo (Gutmann *et al.*, 2020; Cheng *et al.*, 2016). En la **Figura 1.11** se presentan las secuencias consenso de las 10 variantes de motivos PPR y sus alineamientos respecto al motivo P.

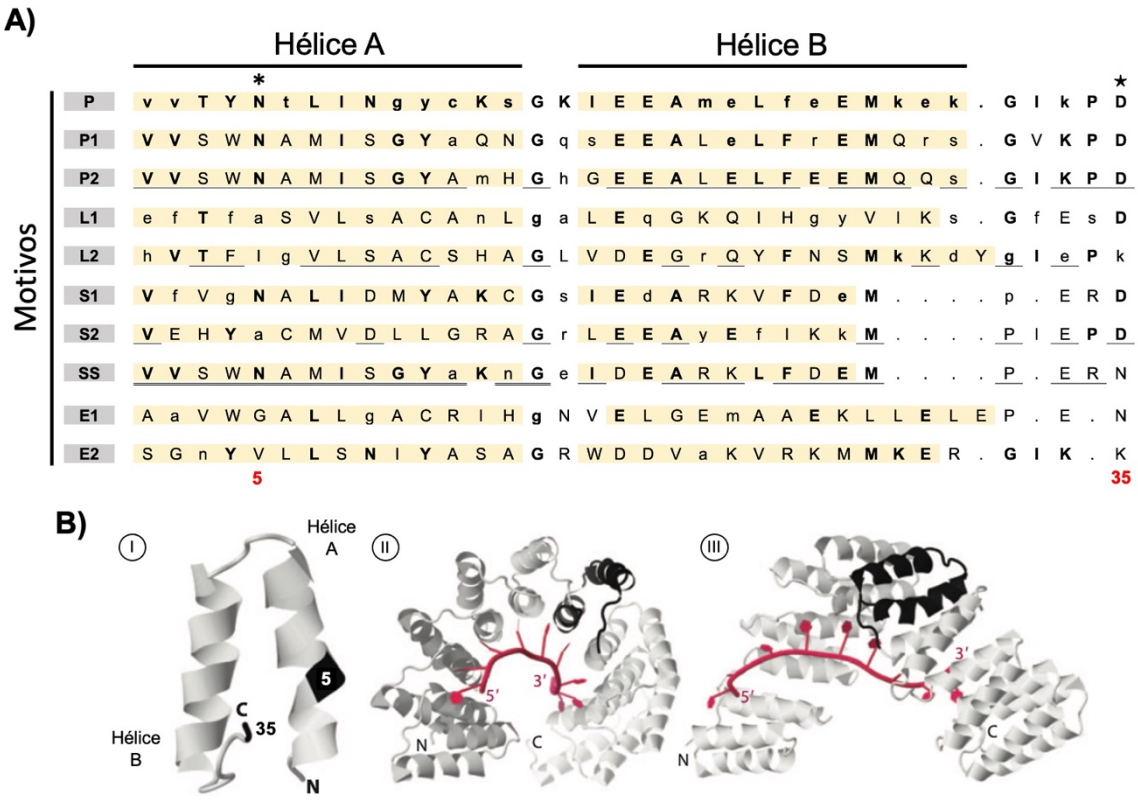


Figura 1.12. Características de los diferentes motivos PPR y estructura tridimensional de una proteína PPR. A) Alineamiento de los diferentes motivos PPR con el motivo P canónico presente en las proteínas de la clase P y estructura tridimensional Las secuencias consenso fueron obtenidas mediante el uso de secuencias de motivos PPR de 41 genomas representativos. Los residuos en letras mayúsculas están altamente conservados dentro de cada motivo. Residuos en negritas coinciden con el motivo P canónico. Residuos subrayados indican identidad entre P2 y P1, L2 y L1, S2 y S1, SS y P1 (dos líneas) o SS y S1 (una línea). Las dos hélices vistas en cada una de las estructuras modelo se encuentran marcadas en amarillo y señaladas en la parte superior. El asterisco indica el aminoácido en la posición 5 y la estrella el aminoácido en la posición 36, los cuales son claves en el reconocimiento y unión a una base nucleotídica. Imagen modificada de Cheng *et al.* (2016). **B)** Estructura tridimensional de la proteína proteinaceous RNase P1 (PRORP1, AT2G32230)

de *A. thaliana*. I) Motivo individual PPR mostrando en negro los aminoácidos 5 y 35 claves en la unión específica a un nucleótido. II) Modelo de diez motivos PPR (gris) unidas a 9 nucleótidos de ARN (indicados en color magenta). III) Mismo modelo anterior visto lateralmente para mostrar la interacción de bases con los bordes del solenoide. Imagen tomada y modificada de Barkan y Small (2014)

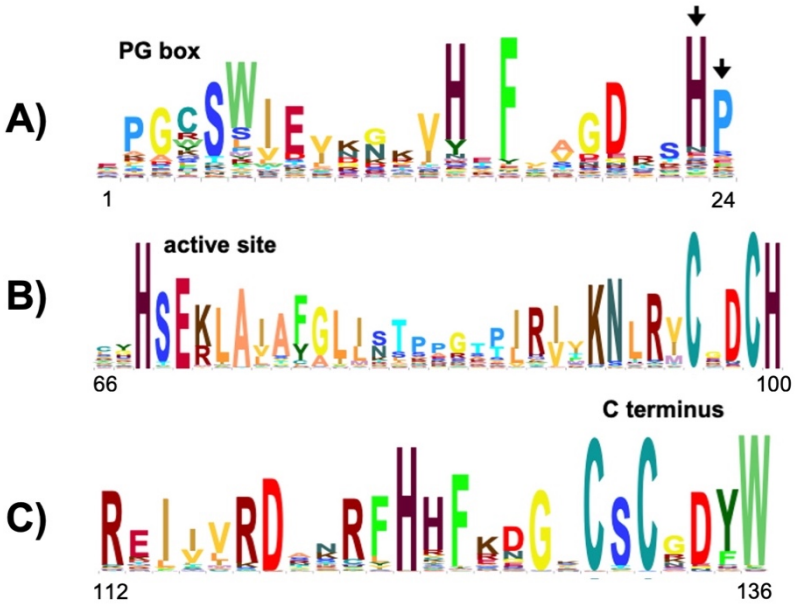


Figura 1.13. Secuencias logos de las regiones conservadas del motivo DYW. Se muestran las tres regiones conservadas del motivo DYW: A) la caja PG, B) el sitio activo y C) el C-terminal. Imagen tomada y modificada de Gutmann *et al.* (2020).

1.9.4. Función de las proteínas PPR en el metabolismo del ARN

Las proteínas PPR juegan un papel central en la regulación post-transcripcional y traduccional tanto en plástidos como en mitocondrias. Se ha descrito que las proteínas PPR de la subfamilia P están envueltas en funciones relacionadas con la maduración del ARN, las cuales abarcan la estabilización de las regiones 5' y 3', protección contra nucleasas, corte del transcrito en las regiones intergénicas, procesamiento alternativo y en la regulación del inicio de la traducción. Por otro lado, los miembros de la subfamilia PLS están particularmente asociados con la edición de ARN (Gutmann *et al.*, 2020; Cheng *et al.*, 2016; Shikanai y Fujii, 2013). En la **Tabla 1.2** se enlistan algunas funciones clásicas descritas para PPR previamente estudiadas.

Tabla 1.2. Mecanismos de acción desempeñados por diferentes proteínas PPR en plantas

Gen	Subfamilia PPR	Organismo	Sitio de acción	Mecanismo de acción	Función de la proteína	Bibliografía
SSA1	P	*O. sativa*	CL	-Edición del ARN -Procesamiento alternativo del ARN	-Edición del ARN *ndhB* -Procesamiento del intrón en el ARN *ycf3*	(Wang *et al.*, 2022)
DG409	P	*A. thaliana*	CL/MT	-Edición de ARN	-Edición de los ARN *clpP*, *rpoA*, *accD*, *nad7* and *rps3*.	(Wang *et al.*, 2023)
AES	P	*A. thaliana*	CL	-Procesamiento alternativo del ARN	-Procesamiento del operón *psbB* (*psbB-psbT-psbH-petB-petD*), *ycf3* y *ndhA*.	(An *et al.*, 2023)
PpPPR_32	P	*P. patens*	CL	-Regulación postranscripcional y traduccional	-Regula la expresión del gen *psaC* así como la acumulación de la proteína que codifica.	(Suzuki *et al.*, 2022)
TPJ1	P	*Z. mays*	CL	-Regulación de la traducción	-Participa en la traducción del ARNm *psbJ*.	(Williams-Carrier *et al.*, 2019)
LPE1	P	*A. thaliana*	CL	-Regulación de la traducción	-Unión al 5' UTR del ARNm *psbA*. -Facilita la asociación de HCF173 con el ARNm *psbA* para regular la traducción la proteína D1. -Participa en la traducción de los ARNm *psbJ* and *psbN*.	(Williams-Carrier *et al.*, 2019) (Jin *et al.*, 2018)
PPR287	P	*A. thaliana*	CL	-Estabilidad de ARNr	-Regula la estabilidad de los precursores y productos maduros de los ARNr 4.5S, 5S, 16S and 23S.	(Lee *et al.*, 2019b)
PPR10	P	*Z. mays*	CL	Estabilidad del ARNm	Protección de regiones 5' o 3' de ARNm frente a la actividad de exonucleasas.	(Pfalz *et al.*, 2009)
PpPPR_21	P	*P. patens*	CL	-Estabilidad del ARNm -Regulación de la traducción.	-Estabilización de los ARNm dicistrónicos *psbI-ycf12* por la unión a las regiones 5' traducibles y no traducibles del ARNm *psbI*. -Bloqueo de la degradación exonucleolítica 5'→3' y regulación de su traducción.	(Ebihara *et al.*, 2019)
PGL1	PLS	*O. sativa*	MT/CL	Edición del ARN	-Relacionado con la edición del ARN *ndhD* cloroplástico y del ARN	(Xiao *et al.*, 2018)

		ccmFc mitocondrial vía editosoma (en interacción con proteínas MORF).
MIT= mitocondria, CL= cloroplasto		

1.9.5. Las proteínas PPR y el albinismo

Uno de los aspectos más interesantes al estudiar la familia de proteínas PPR es su estrecha relación con la aparición de fenotipos albinos y su impacto en el desarrollo del cloroplasto. En la especie *Camellia sinensis* L. cv. AnJiBaiCha se tiene registro de un fenotipo albino que aparece al descender la temperatura ambiental. Con el fin de conocer mejor los mecanismos que rigen este albinismo periódico se identificaron genes diferencialmente expresados mediante la técnica de cDNA-AFLP. Como resultado del análisis se encontraron 127 fragmentos derivados de transcritos, de los cuales 60 mostraron alta similitud a secuencias con funciones conocidas. Dentro de los genes relacionados con la biogénesis del cloroplasto, se detectaron tres transcritos de genes *PPR* (Yuan *et al.*, 2015). Recientemente en esta misma especie se determinó que 22 proteínas PPR podrían estar relacionadas con los procesos de etiolación y albinismo (Zhang *et al.*, 2022). En mutantes albinas o verde pálido (*albino or pale-green* o *apg*, por sus siglas en inglés) de *A. thaliana* también se ha encontrado una intensa actividad transcripcional de los genes PPR (Satou *et al.*, 2014).

En arroz, se ha identificado a la proteína nuclear OsPPR1 como un elemento clave en el desarrollo del cloroplasto. Al utilizar una estrategia transgénica antisentido para suprimir la expresión de *OsPPR1*, se obtuvieron mutantes albinas y letales. Si bien, el análisis citológico comparativo entre la mutante y el fenotipo silvestre no reveló cambios en la estructura foliar, sí se detectaron anormalidades en la forma y el número de los plástidos de la mutante. Un análisis más fino demostró que la mutante *osppr1* presentaba estructuras tipo proplástido que carecían de membranas tilacoidales (Gothandam *et al.*, 2005). Algo similar sucede con la mutante letal *asl3* (*albino seedling lethality3*) de arroz, que alojó la mutación en el gen *ASL3* que codifica una proteína PPR cloroplástica. Este fenotipo albino se asoció a una reducción en el contenido de pigmentos fotosintéticos, una alteración en el desarrollo temprano del cloroplasto así como con severos cambios transcripcionales de genes asociados a la maquinaria traduccional y fotosintética (Lin *et al.*, 2015).

También se han obtenido mutantes en genes *PPR* de *A. thaliana*, cuyos fenotipos tienden a ser albinos y/o cloróticos. Por ejemplo, la mutante *pigmented-defective mutant3* (*pdm3*), la cual exhibe un fenotipo albino en hojas, eventualmente conduce a una letalidad en las plántulas bajo condiciones de crecimiento autotróficas. Sus principales efectos en el cloroplasto parecen estar relacionados con la regulación en los niveles de transcritos dependientes de las polimerasas codificadas en el plástido y en el núcleo, así como en el procesamiento alternativo de los genes *trnA*, *ndhB* y *clpP-1*, lo que revela una participación activa en el desarrollo del cloroplasto (Zhang *et al.*, 2017).

Considerando el posible bloqueo de la biogénesis del cloroplasto en las plántulas de *Agave* albino, así como el papel crucial que podrían protagonizar las proteínas PPR por su vínculo con la aparición de fenotipos con alteraciones en su pigmentación, en este proyecto de investigación profundizaremos el estudio de las variantes somaclonales de *Agave angustifolia* a través de diferentes herramientas ómicas.

1.10. EL USO DE HERRAMIENTAS ÓMICAS EN EL ESTUDIO DE MODELOS VEGETALES

Los nuevos avances tecnológicos que se aplican hoy en día en las investigaciones con modelos de estudio vegetal, entre las que se incluyen las herramientas ómicas como la genómica, la transcriptómica, la proteómica y la metabolómica, han revolucionado la forma de entender cómo las plantas responden a condiciones específicas que afectan su desarrollo y crecimiento. Parte de su éxito se debe a que su aplicación permite tener un panorama más amplio de las redes genéticas y moleculares de un modelo de estudio (Ahmad *et al.*, 2016; Ahsan *et al.*, 2010). En este proyecto de investigación haremos uso de dos herramientas ómicas, el análisis transcriptómico y proteómico, para el estudio de las variantes somaclonales de *A. angustifolia*.

1.10.1. El análisis transcriptómico

La publicación del método de secuenciación Sanger a finales de la década de los años setenta (Sanger *et al.*, 1977; Sanger y Coulson, 1975) revolucionó la forma de hacer investigación biológica (Mardis, 2017). Fue a inicios del siglo XXI cuando el avance en esta área alcanzó un mayor apogeo con las tecnologías de secuenciación de siguiente generación (NGS) (*next-generation sequencing*, por sus siglas en inglés). Estas nuevas

tecnologías tuvieron dos ventajas claves para su éxito: un manejo masivo de datos producidos por las plataformas de secuenciación a través de corridas de miles de reacciones simultáneas, lo que a su vez tuvo un impacto en la reducción de costos por base secuenciada (Slatko *et al.*, 2018; Mardis, 2017). Uno de los retos para el manejo correcto de los datos masivos ha sido la disponibilidad de nuevos y sofisticados programas que provean las bases para simplificar el análisis de tal información (Mardis, 2017; Garber *et al.*, 2011). A la aplicación de herramientas computacionales al estudio de sistemas bióticos se le conoce como bioinformática (Bayat, 2002). Con esto en mente no podemos olvidar que estas tecnologías de secuenciación y la capacidad de análisis bioinformático siguen respondiendo a un único objetivo: facilitar y mejorar el entendimiento de los sistemas biológicos de interés (Mardis, 2017).

Hoy en día, entre las tecnologías de NGS, Illumina se coloca como la más popular en el mercado de las plataformas de secuenciación. Parte de su éxito se debe a su alto rendimiento por corrida que alcanza los 750 Gb, al número de lecturas que en los equipos más modernos ronda los 3,000 millones, a su baja tasa de error (<0.1%) y al bajo costo por base secuenciada (Goodwin *et al.*, 2016). La gran ventaja de esta tecnología es que puede ser aplicado a un sin fin de protocolos que incluyen secuenciación de genomas, exomas, metilomas, metagenomas, inmunoprecipitación de la cromatina acoplada a secuenciación (*ChIP-seq*, por sus siglas en inglés), secuenciación de ARN (*RNA-seq*, por sus siglas en inglés), entre otros (Slatko *et al.*, 2018).

El concepto de RNA-seq refiere al conjunto de procedimientos experimentales que genera lecturas de secuencias de ADN derivado de ARN mediante la aplicación de tecnologías de NGS (Garber *et al.*, 2011). Actualmente, el método RNA-seq se ha convertido en una técnica poderosa y robusta que revolucionó la manera en la que se abordan problemas biológicos específicos, particularmente la forma de hacer análisis transcriptómicos (Hölzer y Marz, 2019; Wang *et al.*, 2019; Garber *et al.*, 2011).

En términos generales, un transcriptoma es definido como el juego completo de transcritos de una muestra (incluyendo tanto ARN codificantes como no codificantes) en un espacio y momento específico (Moreton *et al.*, 2016). Su uso permite a) reconstruir las secuencias completas de los ARNm generados a partir del genoma, b) mejorar la estimación del número de transcritos producidos a partir de un genoma, c) detectar

isoformas obtenidas por eventos de procesamiento alternativo, d) revelar información sobre la variación alélica vinculada con el SNP (*single-nucleotide polymorphism*, por sus siglas en inglés), e) cuantificar los cambios en los niveles de expresión de los genes entre las condiciones estudiadas, entre otros (Wang *et al.*, 2019; Moreton *et al.*, 2016; Haas *et al.*, 2013). Actualmente existen dos estrategias para la obtención de un transcriptoma: el ensamble a partir de un genoma de referencia y el ensamble *de novo*. En el primer caso, se utiliza como referencia un genoma secuenciado de alta calidad para mapear las lecturas y reconstruir el transcriptoma. En el segundo caso, para organismos que no cuentan con un genoma secuenciado, la estrategia a seguir es el ensamble a partir de la sobreposición de lecturas mediante la construcción de grafos de Bruijn (Moreton *et al.*, 2016). Considerando que *A. angustifolia* carece de un genoma de referencia, la estrategia que seguiremos en este proyecto será el ensamble *de novo*.

1.10.2. El análisis proteómico

Al estudiar el genoma de cualquier organismo tenemos que considerar que al ser la estructura básica de la herencia biológica, este presenta características poco cambiantes al transmitirse de generación en generación (Kosová *et al.*, 2018). Por su parte, el estudio de transcriptomas como lo mencionamos en la sección anterior, ofrece un panorama informativo más amplio de los cambios en la expresión de genes de interés en una condición y en un momento dado. Sin embargo, uno de los principales inconvenientes es que el nivel de expresión de un gen respecto a la abundancia de su respectiva proteína codificante puede mostrar discrepancias al ser comparados (Lyu *et al.*, 2021; Chen *et al.*, 2017). Esto ha sido vinculado a mecanismos reguladores que van desde las modificaciones postranscripcionales, traduccionales y postraduccionales (Tian *et al.*, 2004). Por lo tanto, los análisis a nivel molecular en plantas no pueden limitarse solo a un nivel transcripcional (Benešová *et al.*, 2012). Es aquí donde la proteómica se coloca como una herramienta que podría ofrecernos un acercamiento más profundo e informativo.

La proteómica es definida como la disciplina cuyo objeto de estudio es el proteoma (Jorrin-Novo, 2021). La primera vez que se usó el término proteoma, este fue definido como el complemento proteico completo expresado por un genoma, un tipo de célula o de tejido (Wilkins *et al.*, 1996). Años más tarde, la definición de proteoma fue ampliada para incluir también al conjunto de proteínas presentes en extractos, fracciones subcelulares, órganos,

individuos y ecosistemas (Jorrin-Novo, 2021). Desde su concepción, el proteoma no fue visto como una entidad uniforme ni estática. Este puede variar en función del tipo de tejido y responde al desarrollo de los organismos estudiados, al nivel de homeostasis, así como a las señales del ambiente (Kosová *et al.*, 2018; Adams, 2008; Wilkins *et al.*, 1996). A pesar de ser un producto directo del genoma, el número de proteínas puede ser mayor al número de genes debido a los diversos eventos postranscripcionales y postraduccionales de la célula (Jorrin-Novo, 2021; Wilkins *et al.*, 1996).

En este estudio y como un primer acercamiento proteómico a las variantes somaclonales de *A. angustifolia*, se realizó un análisis proteómico cuantitativo mediante marcaje isobárico. Con esto se pretende tener un proteoma de referencia que complemente el estudio transcriptómico, profundice en aspectos del metabolismo y permita distinguir proteínas claves para el establecimiento y sobrevivencia de estos fenotipos albinos.

JUSTIFICACIÓN

El hecho que las plantas de *Agave* presenten un metabolismo fotosintético tipo CAM y una tolerancia a vivir en condiciones extremas, las coloca como un candidato a modelo relevante si consideramos las condiciones climáticas actuales (Stewart, 2015). Hoy en día, la gran mayoría de estudios en plantas se han limitado a modelos como *A. thaliana* y *O. sativa,* así como a cultivos convencionales y comercialmente importantes, cuya incapacidad de hacer frente a las nuevas condiciones ambientales ha sido probada. Si bien las mejoras realizadas a estas especies han impactado en diversas áreas de la vida humana, sus alcances han sido limitados. Esto coloca a las especies del género *Agave* como modelos potenciales para hacer frente a las nuevas necesidades mundiales (Stewart, 2015).

Bajo el contexto descrito, es un hecho que la biotecnología del cloroplasto se posiciona como una de las nuevas alternativas para enfrentarse al cambio climático. En este sentido se vuelve indispensable dilucidar los mecanismos genéticos y moleculares de este organelo, los cuales podrían ser usados en la mejora de cultivos agrícolas en un futuro cercano, particularmente a través de la mejora en el rendimiento fotosintético (De-la-Peña *et al.*, 2022).

La existencia de los *Agaves* verde y albino y un sistema de micropropagación *in vitro* rápido y eficiente, pone a nuestro modelo en una posición única como posible alternativa para abordar el estudio de la biogénesis del cloroplasto y otros aspectos metabólicos ligados completa o parcialmente al cloroplasto. En este sentido, un modelo de *Agave* albino con las características descritas previamente, permite evaluar las alteraciones en mecanismos moleculares específicos, así como la reprogramación metabólica que las plantas exhiben al cambiar su metabolismo autotrófico por uno heterotrófico. Entre estos podemos destacar el papel clave de la reprogramación en el metabolismo de los AA.

El uso de herramientas ómicas permitirá tener un primer acercamiento global a lo que sucede en el contexto de la biogénesis del cloroplasto. De particular interés para nosotros es la familia de genes PPR, la cual recientemente ha emergido como una posible herramienta biotecnológica en la manipulación del ARN, particularmente a nivel de su edición (Lee y Kang, 2023).

PREGUNTAS DE INVESTIGACIÓN

- ¿La plántula de *Agave* albina presenta cambios en los perfiles de expresión de genes implicados en la biogénesis del cloroplasto respecto a la plántula de *Agave* verde?

- ¿Cómo se expresan y cuáles son los ARN blanco hipotéticos de un conjunto de genes *PPR* encargados de la regulación postranscripcional en el plástido de la plántula de *Agave* albina?

- ¿Hay diferencias en la acumulación de enzimas del metabolismo de AA y la concentración de los mismos, entre las variantes somaclonales albina y variegada de *A. angustifolia?*

OBJETIVO GENERAL

- Analizar los cambios transcriptómicos y proteómicos relacionados con la biogénesis del cloroplasto y el metabolismo de aminoácidos en las variantes somaclonales verde, variegada y albina de *A. angustifolia* Haw.

OBJETIVOS ESPECÍFICOS

- Generar un transcriptoma y un proteoma de referencia de las variantes somaclonales de *A. angustifolia*.

- Analizar los perfiles de expresión de un conjunto de genes implicados en la biogénesis del cloroplasto en el transcriptoma de las variantes somaclonales de *A. angustifolia*.

- Determinar con base en el transcriptoma de referencia el total de transcritos PPR y su estructura a nivel de motivos y dominios.

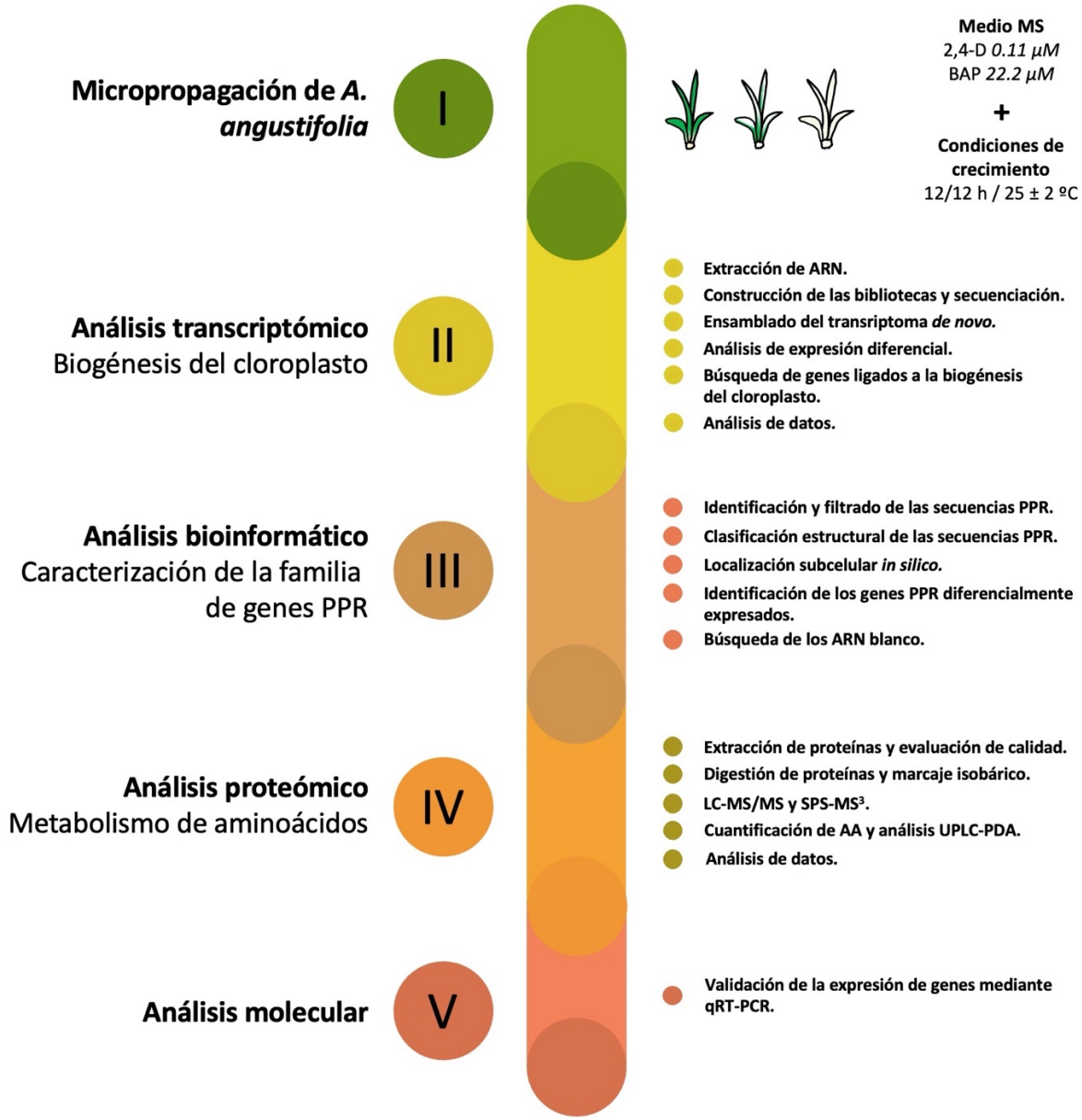
- Identificar los posibles blancos de acción de un conjunto de genes PPR.

- Analizar los cambios a nivel proteómico en la concentración de enzimas claves en el metabolismo de aminoácidos en las variantes somaclonales de *A. angustifolia*.

- Cuantificar el contenido de aminoácidos libres presentes en las variantes somaclonales de *A. angustifolia*.

- Validar los niveles de expresión de genes claves en la biogénesis del cloroplasto y en el metabolismo de aminoácidos en las variantes somaclonales de *A. angustifolia*.

ESTRATEGIA EXPERIMENTAL



CAPÍTULO II

Transcriptome analysis reveals molecular mechanisms underlying chloroplast biogenesis in albino *Agave angustifolia* plants

Manuscrito por someter (fecha tentativa octubre de 2023).

Autores: Andrade-Marcial, M., Che-Aguilar*, L., Hernández-Castellano, S., Pacheco-Arjona, R., De-la-Peña, C.

2.1. INTRODUCTION

Since the scientific community began searching for new plant biotechnology platforms (Ma *et al.*, 2022), the chloroplast quickly became a multifaceted alternative. Chloroplast biotechnology has influenced many aspects of modern life, such as insect pest control (He, 2022), human disease control (Legen *et al.*, 2023), value nutritional compounds (Tanwar *et al.*, 2022), energy production (Gomez-Casati *et al.*, 2022), nanotechnology (Newkirk *et al.*, 2021) plant environmental stress response (Singhal *et al.*, 2023), improved photosynthetic efficiency (Głowacka *et al.*, 2023), among others. In the context of the plant cell, the chloroplast is considered the most efficient metabolic factory, sustaining life on the planet (Kirchhoff, 2019). This "*small green factory*" imports CO_2 and H_2O and captures light energy to convert them into O_2 and chemical energy, in a process known as photosynthesis (Jan *et al.*, 2022). However, the chloroplast cannot be reduced to only its photosynthetic function, as it also comprises other metabolic aspects such as C, N and S assimilation and the biosynthesis of sugars, AAs, chlorophyll, carotenoids, fatty acids, lipids, vitamins, cofactors, and more (Rolland *et al.*, 2018). Although the chloroplast is a semi-autonomous organelle with its own genome (known as the plastome), it is not capable of performing all of the above functions independently.

The lack of functional autonomy of chloroplasts can only be understood in the light of the endosymbiotic theory, which proposes the origin of the chloroplast (Margulis, 1970). This theory suggests that plastids evolved from free-living ancient photosynthetic cyanobacterium that was taken up by a heterotrophic eukaryotic cell that already contained nucleus and mitochondria (Archibald, 2009). One strong piece of evidence supporting the lack of chloroplast autonomy is the horizontal gene transfer events that occurred from the

primitive cyanobacterial genome to the host cell nucleus (Abdallah *et al.*, 2000; Martin *et al.*, 1998). Currently, the chloroplast proteome is estimated to contain around 3000 proteins, of which more than 95% are encoded in the nucleus, translated in the cytoplasmic ribosomes, and imported into the chloroplast (Zoschke y Bock, 2018; Martin *et al.*, 1998). However, despite the minimal contribution of the plastome to the final number of proteins, it is still essential for the proper functioning of this organelle.

A hallmark of terrestrial plant plastomes is their conserved structure, organization, and gene distribution. These genomes typically contain around 130 genes, of which between 60-70% are protein-coding genes and the rest are tRNAs and rRNAs (Daniell *et al.*, 2016). Among the genes that encode proteins, those involved with photosynthetic metabolism dominate. These include genes that encode proteins of photosystems I (PSI) and II (PSII), as well as subunits of the cytochrome b6/f complex, NADH dehydrogenase, ATP synthase and RuBisCO large subunit. Additionally, there are genes that encode the subunits of the plastid-encoded RNA polymerase (PEP), of the 70S ribosome, and others of unknown function (Dobrogojski *et al.*, 2020). However, an additional aspect that must be taken into account to understand the plastids physiology is the fine-tuned regulation that takes place during their development and maturation.

The plastid family is a group of heterogeneous organelles, both structurally and functionally. While the chloroplast is the best-known member for its photosynthetic capacity, the other members also perform key functions for the plant cell (Sierra *et al.*, 2023). All plastids derive from small, non-photosynthetic, and undifferentiated plastids very numerous in the meristematic tissues known as proplastids (Sierra *et al.*, 2023; Loudya *et al.*, 2021). The process by which proplastids, after a series of division events and in response to early light perception, differentiate in to functional chloroplasts is known as chloroplast biogenesis (Cackett *et al.*, 2022).

Considering that most chloroplast proteins are encoded in the cell nucleus and imported from there, a complex network of signaling mechanisms is required to regulate communication between the plastid and the nucleus (Pogson y Albrecht, 2011). These mechanisms have been classified into two categories: anterograde and retrograde regulation. The first of these involves nuclear-encoded proteins that regulate transcription, RNA metabolism, translation, and the assembly of chloroplast photosynthetic complexes in

response to environmental signals such as light (Hwang *et al.*, 2022; Rochaix y Ramundo, 2017). The second has as its main objective to notify the nucleus about the state of the plastid, which can trigger changes in the transcription of nuclear genes, particularly photosynthesis-associated nuclear genes (PhANGs) (Shimizu y Masuda, 2021).

To delve into the molecular details of chloroplast biogenesis, several omic studies have been conducted in plant models. One of the most interesting is the study conducted on wheat plants by Loudya *et al.* (2021). This study reports the first transcriptional map in a developmental gradient spanning from the meristem to the leaf tip. Their results propose two distinguishable phases. The first phase corresponds to the meristematic region and base leaf where proplastids abound. During this phase, all the cellular resources are dedicated to processes of division, transcription, translation, establishment of the protein import machinery and growth of the chloroplast and cell proliferation and expansion. In the second phase, the presence of chloroplasts dominates and is characterized by the build-up of the plastid photosynthetic machinery.

Albino plants, which exhibit partial or complete loss of photosynthetic pigments, have proven to be an invaluable and ignored model for the study of chloroplast biogenesis. These plants exhibit a range of unique features, including early blockage of proplastid differentiation and imperfect development of thylakoid membranes (Żur *et al.*, 2021; Kumari *et al.*, 2009). During *in vitro* micropropagation of green plantlets of *A. angustifolia* Haw., a spontaneous event with unknown causes gave rise to a completely albino somaclonal variant (Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016). Despite the progress made in the structural, cellular, genetic, epigenetic, molecular and physiological characterization of these albino plantlets, no particular attention has been paid to the study of the signaling mechanisms that govern chloroplast biogenesis and that could be severely compromised. In this study, we performed transcriptome analysis of albino *Agave* plantlets to gain insight into the molecular disturbances affecting chloroplast development. Our efforts are aimed at identifying whether the expression of certain key genes for plastid development and differentiation is compromised, and how this affects the chloroplast biogenesis. Our results reveal an intense transcriptional reprogramming that suggest the existence of a corrective retro-anterograde communication acting in this albino model.

2.2. MATERIALS AND METHODS

2.2.1. Plant materials

Two somaclonal lines with the same genetic background but different phenotypes were obtained by micropropagation (Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016). Plantlets from the green (G) and albino (A) phenotypes (**Figure 2.1)** were micropropagated and incubated as reported by Andrade-Marcial *et al.* (2023). Leaf and shoot meristem (referred to as "meristem" throughout the article) from ~2.5-cm tall plantlets of each phenotype were used in this study. This resulted in four study tissues: green leaf (GL), albino leaf (AL), green meristem (GM), and albino meristem (AM).

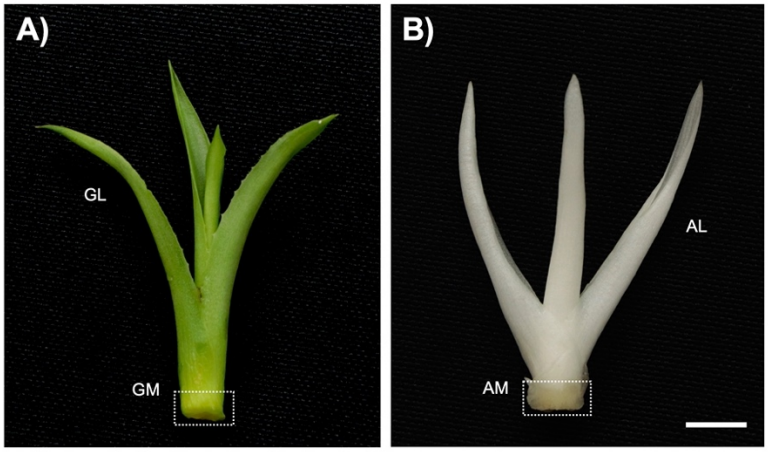


Figure 2.1. Somaclonal variants of *A. angustifolia*. Individual green **(A)** and albino **(B)** plantlets, their meristematic tissues (GM and AM) (indicated by the rectangle), and foliar (GL and AL) tissues are showed in both phenotypes. Scale bar = 0.5 cm

2.2.2. Extraction of nucleic acids

DNA was isolated from 500 mg of tissue by the protocol described by Echevarría-Machado *et al.* (2005). Total RNA was isolated from 100 mg of tissues using TRI Reagent® protocol (Sigma-Aldrich) and treated with DNase I, RNase-free (Thermo-Fisher Scientific). Integrity of nucleic acids was visualized on agarose gel electrophoresis and the concentration and purity were determined by Nanodrop™ One (Thermo-Fisher Scientific). RNA integrity was checked in the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

2.2.3. Library construction and sequencing

For library construction, only samples with an RNA integrity number over 6.3 were used. A total amount of 1 μg RNA per sample was used as input material. Library construction and sequencing was carried out by Novogene (Beijing, China, https://www.novogene.com/us-en/). Three libraries were made for each tissue of interest. A total of twelve libraries were sequenced via Illumina® HiSeq™ 4000 platform (Illumina, San Diego, CA, USA). A total of 60 million paired-end reads of 150 bp in length were obtained from each sample.

2.2.4. *De novo* transcriptome assembly and data analysis

In the absence of reference genome, *de novo* transcriptome assembly strategy was used in this study. Raw reads were processed using in-house perl scripts developed by Novogene (https://www.novogene.com/us-en/). Reads with adapter contamination, N bases in more than 10% of each read, and low quality (Q<20) that constitute more than 50% of the length were removed. The filtered reads were used for transcriptome assembly. The total paired-end reads obtained from GL, AL, GM and AM tissues were merged into a pair fastq files to obtain a single assembly. *De novo* transcriptome assembly was performed using Trinity software (vers. r20140413p1) (Grabherr *et al.*, 2011) with *min_kmer_cov* set to 2 and all other parameters set default. The redundancy of the assembled contigs was removed by clustering contigs using Corset software (vers. 1.05) (Davidson y Oshlack, 2014). The longest contigs of each cluster were selected as unigenes. This assembly was deposited at DDBJ/EMBL/GenBank under the accession (in progress).

The quality of the generated assembly was assessed using Fasta Statistics (https://github.com/galaxyproject/tools-iuc). Completeness of *de novo* aassembly was evaluated based on the content of single-copy orthologs conserved using BUSCO software (vers. 4.1.2) (Simão *et al.*, 2015). BUSCO analysis was performed using the *Embryophyta_odb9* database as reference.

Gene expression was estimated based on read counts by transcript with RSEM software (RNA-Seq by Expectation-Maximization) (vers.1.2.26) (Li y Dewey, 2011). By default, RSEM uses Bowtie software to align reads against the assembled transcripts and filtered by Corset. Expression counts were normalized using TPM (transcripts per million)

(Wagner *et al.*, 2012) values. The similarity of the replicates of tissues was confirmed by Pearson´s correlation analysis and Principal Component Analysis using the normalized expression matrix. Differential expression analysis was performed in the GL vs AL, GM vs AM, GL vs GM and AL vs AM tissue comparisons using edgeR package (vers. 3.30.3) (Robinson *et al.*, 2009) from Bioconductor. The differentially expressed genes (DEGs) were selected using a *p*-value ≤ 0.05 and a Log2 fold change (LFC) ≤ 1.0 or LFC ≥ 1.0 for genes under and over-expressed, respectively. Heatmaps representing the expression profiles of the unigenes were constructed using ComplexHeatmap software (vers. 2.4.3) (Gu *et al.*, 2016). Venn diagrams were constructed using the InteractiVenn digital tool (Heberle *et al.*, 2015).

2.2.5. Functional annotation and enrichment analysis

To achieve comprehensive gene functional annotation, seven databases were used. Unigenes and peptide sequences were aligned against the nucleotide database (Nt) and non-redundant protein sequences (Nr) databases of NCBI, respectively, using BLAST (vers. 2.2.28+) (Altschul *et al.*, 1990) with a cutoff *e*-value ≤ 1e-5. Prediction of Pfam domains were analyzed with HMMER software (vers.3.0) (Eddy, 2011) with a cutoff *E*-value ≤ 0.01. The BLAST and Pfam results were exported to Blast2GO software (vers. 2.5) (Götz *et al.*, 2008) for mapping and retrieving Gene Ontology (GO) terms with a cutoff *E*-value ≤ 0.01. KEGG pathways were assigned using the online KEGG Automatic Annotation Server (KAAS) (Moriya *et al.*, 2007) with a cutoff *E*-value ≤ 1e-10. Finally, unigenes were aligned to Swissprot-Uniprot and KOG (EuKaryotic Orthologous Groups) databases using Diamond software (v.0.822) (Buchnik *et al*., 2014) with a cutoff *E*-value ≤ 1e-5 and 1e-3, respectively. Additionally, nearest orthologs of all genes were searched by running a local BLASTp (vers. 2.10.1+) (Altschul *et al.*, 1990) against the TAIR11 proteome database of *A. thaliana* (downloaded on April 2023). The local BLASTp search was executed with default parameters, and sequence pairwise alignments with an identity percentage greater than 50% were selected. For sequence alignments with multiple hits, only the best match was selected and reported.

Enrichment analyzes were performed from the DEGs (up- and down-expressed genes) identified in the four tissue comparisons using the tools available in the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (2021 update) (Sherman *et al.*,

2022). For this analysis, the biological process (BP) and cellular component (CC) Gene Ontology (GO) terms were used, as well as KEGG pathways associated with the DEGs. Only terms/pathways with nominal *p*-values ≤ 0.1 were considered significantly enriched. Bubble plots were used to show the results of the enrichment analysis. The *ggplot2* package (vers. 3.4.0) available in RStudio software (vers. 2022.02.0+443) (RStudio, 2021) were used to design such graphs.

2.2.6. Selection of genes linked to chloroplast biogenesis

To evaluate the homeostatic and development state of plastids in the tissues, a set of nuclear-encoded chloroplast genes linked to plastid biogenesis were selected from Loudya *et al.* (2021). Nearest orthologs of these genes were also searched against the RefSeq protein database (directed to *Embryophyta* [*taxid:3193*]) (NCBI, February 2021) to confirm their previous annotation. The selected genes were grouped into categories based on their functions: I) plastid division-associated proteins (*PDV1*, *PDV2*, *FTsZ1*, *ARC5*, *ARC6* y *FZL*), II) TOC/TIC components and chaperones (*TOC34*, *TOC75*, *TOC132*, *TOC159*, *SP1*, *OEP80*, *TIC20-I*, *TIC40*, *TIC110*, *HSP21* y *cpHSC70*), III) plastid nucleoid-associated proteins (*PTAC2*, *PTAC10*, *PTAC12*, *PTAC14*, *MurE*, *WHY3* y *FLN1*), IV) transcription-related genes (*SIG1*, *SIG2*, *SIG3*, *SIG5*, *SIG6*, *RCB* y *Rpotp*), V) translation-related genes (*PRPL11*, *PRPL29*, *PRPS9*, *SCO1*, *SVR7*, *RABE1b* y *FUG1*), VI) plastid coverage (*REC* y *FMT*), VII) pigment and thylakoid biogenesis (*PDS, PSY, CURT1A, HEMA, HEMC, HEME1, HEMF, HEMG1, HEMG2, GSA1, ALADH1, UROS, FC1, FC2, CHLD, CHLG, CHLI, CHLM, GUN2, GUN3, GUN4, GUN5, CRD1, DVR, POR, CAO, HCAR, NOL, ERS y CHLI),* VIII) photophosphorylation, Calvin cycle and photorespiration (*LHCB, RBCS, PETE1, GLYK, PSAE, PSBO2, LHCA, SBPASE, ATPC, PGLP1),* IX) nuclear-encoded regulators of plastid development (*PIF1, PIF3, GLK1, COP1, DET1, HY5, GNC, CIA2, PHY y CRY1, CRY2),* and X) peroxidases/ reactive oxygen species (ROS)-associated proteins (*EX1, EX2, SAL1 and APX).*

2.2.7. Gene expression analysis by qRT-PCR

Relative quantitation of gene expression was performed by quantitative reverse transcription PCR (qRT-PCR). cDNA was obtained using the RevertAid First Strand cDNA Synthesis Kit (Thermo-Fisher Scientific). A total of eighteen nuclear-encoded chloroplast genes linked to plastid biogenesis were selected to validate their expression (*ARC5*, *TIC20*,

TOC34, *SIG3*, *SIG5*, *SIG6*, *FUG1*, *GSA*, *POR*, *CRD1*, *CAO*, *RBCS*, *PIF1*, *GLK*, *COP1*, *HY5*, *PHY* and *GUN1*). *Actin*, *tubulin* and *18S rRNA* were used as housekeeping genes. Additionally, fifteen genes encoded by the chloroplast genome were also selected to evaluate their expression levels (*rbcL*, *atpE*, *psaA*, *psaB*, *psbD*, *rps2*, *rps4*, *rps7*, *rps11*, *rpl14*, *rpl32*, *rpoA*, *rpoB*, *rpoC1* and *rpoC2).* The nucleotide sequences of these genes were taken from the chloroplast genome of *A. angustifolia* available under accession number MW540498 (GenBank-NCBI) (Qin *et al.*, 2021).The list of oligonucleotides synthesized for this study is presented in **Table 2.1**. qRT-PCR reactions were conducted using a StepOnePlus Real-Time PCR System and PowerUp™ SYBR™ Green Master Mix (2X) (Applied Biosystems, USA). A standard curve was generated of each pair of oligonucleotides, aiming for an efficiency between 90 and 110% as recommended. Each qRT-PCR reaction was prepared according to the supplier's instructions. The thermocycler program consisted of UDG activation at 50 °C for 2 min, an initial denaturation at 95 °C for 3 min, followed by 40 cycles each with 30 s denaturation at 95 °C, 30 s annealing at 58 °C, 60 s extension at 72 °C, and a final step of extension for 5 min at 72 °C. Three biological replicates were used for each tissue of interest. The relative expression levels of the genes were calculated using the 2–ΔΔCT method (Livak y Schmittgen, 2001).

Table 2.1. List of oligonucleotides used in the qPCR analysis

Gene ID	Oligonucleotide name	Sequence (5'→3')
AaCAO	*AaCAOf*	TTGGGACAGTGAATGAGG
	AaCAOr	GAGGTGGTTGAAAAGAAGG
AaPHY	*AaPHYf*	AGCCATAGGGAAGCATTTACTTACC
	AaPHYr	TCGCACAAGCATTCACAACCAAGAT
AaHY5	*AaHY5f*	AGCAGGCAAGAGAGAGGAAGAAG
	AaHY5r	GTCTCAGCATTTGGTTCTCATTCTG
AaCOP1	*AaCOP1f*	CATTTACTCGCTACAGTCGCTTGC
	AaCOP1r	TTCTCTTAGATACTCCAGCTGTAGC
AaSIG3	*AaSIG3f*	GATTCTAAAGTTTCTGCGCTCACTG
	AaSIG3r	GGTTTTCTCTAGGTCTGCAATCTCC
AaSIG6	*AaSIG6f*	AGCTGGCGAAACGTGTTGGCATC
	AaSIG6r	GTGAGAAGGCGTGAAGGTAAACAGG
AaTOC34	*AaTOC34f*	GCCCATAATGTGTTCGCGGACA
	AaTOC34r	CAGAAGCACATCAATGGTCTTGTCT
AaTIC20-I	*AaTIC20-If*	GCCTACCTCTTCACCGTGCTTGA
	AaTIC20-Ir	GAGCAATGGCATCCCTTCAAACC
AaGLK1	*AaGLK1f*	GTGGAACAGCTGGGAGTGGACAAG
	AaGLK1r	ATGGCTGGCGATGTTGTGGCGAGT
AaPIF1	*AaPIF1f*	GATGTGGAAGTGGAGCATAATGATG
	AaPIF1r	GCCATCCTTCCATCCAGCAAATG
AaGSA1	*AaGSA1f*	GGTGAATGCTGCCTTGATTGAGAC
	AaGSA1r	CAAACCGGACCATTTCGATGCAG
AaCRD1	*AaCRD1f*	AGCTGTCCTCTGGCAATCATTAAG
	AaCRD1r	GGCATGGCGACTTCTTCTCTG
AaPOR	*AaPORf*	GAAGTGGAACGTGATCATGGCC
	AaPORr	TGTCAAGGGAGGCGAGATCGAG

AaARC5	*AaARC5f*	CGGAGAATAGTTATGCAGATTGACC
	AaARC5r	GCAACCATCCCATGTACAAGCAG
AaSIG5	*AaSIG5f*	CTGCGGTGAATATGACTGTGCCA
	AaSIG5r	GTCTTGGAACTTGTCACCACTTGC
AaFUG1	*AaFUG1f*	GCTTGATGAGGAGGACTTGGATATG
	AaFUG1r	CGCTCCAATACCTTGTGTAATGCC
AaRBCS	*AaRBCSf*	TTACCTCCCTCCCTTGTC
	AaRBCSr	GCTCCTTCACAACCTGG
AaGUN1	AaGUN1f	CGTCAAGTTTGGGAGAATG
	AaGUN1r	CCTTCGGAGACAATAGATCG
AaatpE	*AaatpEf*	CAAACCACGCCCCTATTG
	AaatpEr	CATCTCCGCATCATTTCC
AapsbD	*AapsbDf*	CCAACTCAAGCCGAAGAGAC
	AapsbDr	CTACCCCAAGAGCACTCATC
AapsaB	*AapsaBf*	TCTTGCTCTAGCCTCTTTAGGG
	AapsaBr	CTGCGATGTATTGGTGATGAG
AapsaA	*AapsaAf*	GAGTGGCACGTACTTTTATGGC
	AapsaAr	GCGGCTTTGTGATAATGGAAC
AarbcL	AarbcLf	TGATTGCGCCTGTTTCG
	AarbcLr	TATGGTCGTCCCCTATTG
AarpoA	*AarpoAf*	GAAGGTGCCGCATGAATATTCTAC
	AarpoAr	ATCTAGGACCTCTGACACAAATGGA
AarpoB	*AarpoBf*	GGAATGGGTACACTTGACGATATGA
	AarpoBr	CTCCACATATAGTTCCTCGAATCGC
AarpoC1	*AarpoC1f*	CCGATCTATTAGCAACAAGTAGGTC
	AarpoC1r	CCTCATTGGTTGTCCGCGTATC
AarpoC2	*AarpoC2f*	GTACTATCCGAGGTATTTCCGTGAG
	AarpoC2r	GAGTGGCAATGCATCGTATCCC
Aarps2	*Aarps2f*	TGCAAACATATCTGGGTGGGATCAA
	Aarps2r	GAGATCTGGGTCACAATTTGTATCG
Aarps4	*Aarps4f*	GATCGCAATATCGTATTCGTCTAGA
	Aarps4r	CCAAACGCATCTCAAGTAGTTGTAG
Aarps7	*Aarps7f*	GCCATTCGTTGGTTATTAGGG
	Aarps7r	TGAGTCTCTTCCTTTTTGCG
Aarps11	*Aarps11f*	GATCAGGGTATGCAACGAGCAG
	Aarps11r	TTAGGAGGTCGACATCCATTATGTG
Aarpl14	*Aarpl14f*	CAATATGCCTCTAGAAAGATCA
	Aarpl14r	CAACCGCTGCATTGTCATCATATCG
Aarpl32	*Aarpl32f*	CGTATTCGTAGAAATATTTGGAAG
	Aarpl32r	TCTTACCTGTTGTTTGTTGCAC
AaGUN5	*AaGUN5f*	GAGCAGTGAAGATGGTGGCGGA
	AaGUN5r	GTTGACGTTGGAGGAGTAGGAGC
Aa16S rRNA (Loudya *et al*., 2021)	*Aa16SrRNAf*	CATCGGCTAACTCTGTGCCA
	Aa16SrRNAr	GGTTGAGCCCTGGGATTTGA
Aa18S and Aa16S rRNA (Loudya et al., 2021)	*Aa16_18SrRNA_I*	ACCTTGTTACGACTTC
Actin	*AaACTf*	GTCGTACAACTGGTATTGTGCTGGA
	AaACTr	GTAACCACGCTCAGTCAGGATCTTC
β-tubulin	*AaTUBf*	CTCACCTTCTCTGTATTCCCATCC
	AaTUBr	CCTCATTGTCAAGAACCATACACTC
Aa18S rRNA (Suárez-González et al. 2014)	*Aa18SrRNAf*	GCTACCACATCCAAGGAAG
	Aa18SrRNAr	TCGTTAAGGGATTTAGATTGT

Asterisks (*) indicate the unigenes selected as reference genes.

2.2.8. Differential accumulation of chloroplast biogenesis-related proteins

To understand the relationship between the gene expression and protein accumulation in chloroplast biogenesis, a comparative analysis of transcriptomic and proteomic data from two somaclonal variants was conducted. The data on the differential accumulation of these proteins in the two somaclonal variants of interest were taken from

(Andrade-Marcial *et al.*, 2023). Differentially accumulated proteins (DAPs) were selected using a *p*-value < 0.05 and a Log2 fold change (LFC) ≥ 0.58 and ≤ -0.58 for the up- and down-accumulated proteins, respectively.

2.2.9. Quantitation of the chloroplastic genome and accumulation of plastid ribosomes.

Quantitation of plastid genome copy number was performed using quantitative PCR (qPCR). Five genes were selected for this analysis: two nuclear single-copy genes (*GUN5* and *SIG3*) and three genes encoded by the plastid genome (*rbcL*, *psbD* and *rps7*). The protocol used is described in Loudya *et al.* (2021). The sequences of oligonucleotides used are listed in **Table 2.1**.

2.2.10. Statistical analysis

The relative expression of selected genes, plastid genome copy numbers and plastid rRNA quantitation were subjected to one-way analysis of variance (ANOVA) and the statistical differences were obtained by Tukey post-hoc test ($p \leq 0.05$). When the assumption of data normality was not met, the Kruskal–Wallis test were performed followed by Wilcoxon Rank Sum test and Benjamini & Hochberg (BH) correction ($p \leq 0.05$). These analyses were performed using RStudio software (vers. 1.4.1106) (RStudio, 2021). The Graph Pad Prism (vers. 9.2.0) (Graph Pad software, www.graphpad.com) and *ggplot* (vers. 3.3.5) were used to design graphs.

2.3. RESULTS

2.3.1. *De novo* transcriptome assembly of three phenotypes of *A. angustifolia*

Twelve cDNA libraries were constructed from GL, AL, GM and AM tissues and sequenced using Illumina technology. A total of 1,546,301,528 raw reads were obtained. The raw reads were processed to remove low quality reads. A total of 1,468,610,446 clean reads were conserved and used for the *de novo* transcriptome assembly (**Table 2.2**).

Table 2.2. General summary of RNA-seq libraries obtained from GL, AL, GM and AM tissues of somaclonal variants of *A. angustifolia.*

Sample	Raw reads[1]	Clean reads (%)[2]	Q20(%)[3]	Q30(%)[4]	GC (%)[5]	No. of mapped reads (%)[6]
GL1	141,465,176	134,188,024 (94.9)	95.8	90.0	50.5	90,166,738 (67.2)
GL2	121,772,882	117,110,812 (96.2)	95.5	89.3	50.7	75,782,910 (64.7)
GL3	125,622,510	118,810,690 (94.6)	96.3	90.9	50.9	77,814,306 (65.5)

GM1	123,467,388	118,712,094 (96.2)	95.7	89.7	49.3	79,098,238 (66.6)
GM2	120,084,778	114,167,888 (95.1)	96.3	91.1	49.2	76,481,712 (67.0)
GM3	126,004,116	116,736,828 (92.6)	96.3	91.1	49.2	77,005,654 (66.0)
AL1	125,742,402	119,338,710 (94.9)	96.4	91.3	49.7	81,442,100 (68.2)
AL2	126,425,306	119,993,540 (94.9)	96.5	91.4	49.9	82,268,812 (68.6)
AL3	159,324,536	149,002,438 (93.5)	96.6	91.6	51.4	101,259,812 (68.0)
AM1	128,986,022	122,690,770 (95.1)	96.2	90.9	49.8	82,302,200 (67.1)
AM2	121,870,576	117,514,624 (96.4)	97.0	92.3	49.5	77,603,270 (66.0)
AM3	125,535,836	120,344,028 (95.9)	96.9	92.2	50.0	80,975,736 (67.3)
Total	1,546,301,528	1,468,610,446				

[1] Number of reads obtained from sequencing by Illumina platform, [2] number of conserved reads after removing low quality and adapter-contaminated sequences, [3,4] percentages of bases whose correct base recognition rates are greater than 99 y 99.9% in total bases, respectively, [5] percentages of G and C in total bases, and [6] total number of reads that can be mapped to the assembled transcriptome filtered by Bowtie/RSEM software.

The assembly generated by Trinity software consisted of 474,349 contigs. After redundancy removal by Corset software, the number of sequences was reduced to 270,163. A summary of the assembly metrics obtained before and after redundancy removal is presented in **Table 2.3**. The completeness of the transcriptome assembly was determined by BUSCO software. The analysis revealed that 82% of conserved orthologous genes were complete, indicating a high level of completeness. Of this percentage, 52% were duplicated and only 29.9% were single-copy genes. The rest of the genes were fragmented (12%) and missing (6%) (**Figure 2.2B**).

Table 2.3. Summary of statistics of the *A. angustifolia de novo* transcriptome assembly

Statistic	Original transcriptome assembly (Trinity)	Final transcriptome assembly (after Corset)
Total contigs/unigenes	474,349	270,163
Total assembled bases (pb)	300,192,062	244,269,151
Average length of contigs (pb)	632	904
Median contig length (pb)	357	616
Maximum contig length (pb)	15,535	15,535
Minimum contig length (pb)	201	201
GC%	43.4	43.2
N50 (pb)	978	1,282

The clean reads were aligned to the reference transcripts using RSEM software. The overall alignment rate per library was between 64-68% (**Table 2.2**). Based on the read counts of each unigene, an expression matrix was generated and normalized by TPMs values. As expected, biological replicates of the same tissue had the highest correlation coefficients ($r > 0.89$), which confirmed the high similarity between them (**Figure 2.2A**).

2.3.2. Functional annotation

The assembled unigenes were subjected to functional annotation using seven databases (*Nt*, *Nr*, *Pfam*, *SwissProt-Uniprot*, *KOG*, *KO*, and *GO*). A total of 75,885 (28.1%) and 117,256 (43.4%) unigenes were annotated against *Nt* and *Nr* database, respectively. The annotation against *Pfam* database showed 90,297 (33.4%) significant hits. Moreover, 91,669 (33.9%), 38,204 (14.1%), and 44,428 (16.4%) unigenes were annotated against *SwissProt-Uniprot*, *KOG* and *KO* databases, respectively. Finally, 90,821 (33.6%) unigenes were annotated using Blast2GO suite against *GO* terms. The majority of the unigenes annotated were assigned to molecular function (*MF*) (75,530, 83.2 %), followed by biological process (*BP*) (67,306, 74.1%), and cellular component (*CC*) (43,263, 47.6%). In summary, 141,541 unigenes (52.4%) were annotated in at least one database (**Table 2.4**).

Table 2.4. Successfully annotated unigenes of *A. angustifolia* transcriptome using seven databases.

Database	Number of unigenes	Percentage (%)
Annotated in NR	117,256	43.40
Annotated in NT	75,885	28.08
Annotated in KO	44,428	26.44
Annotated in SwissProt-Uniprot	91,669	33.93
Annotated in Pfam	90,297	33.42
Annotated in GO	90,821	33.61
Annotated in KOG	38,204	14.14
Annotated in all databases	19,262	7.12
Annotated in at least one database	141,541	52.39

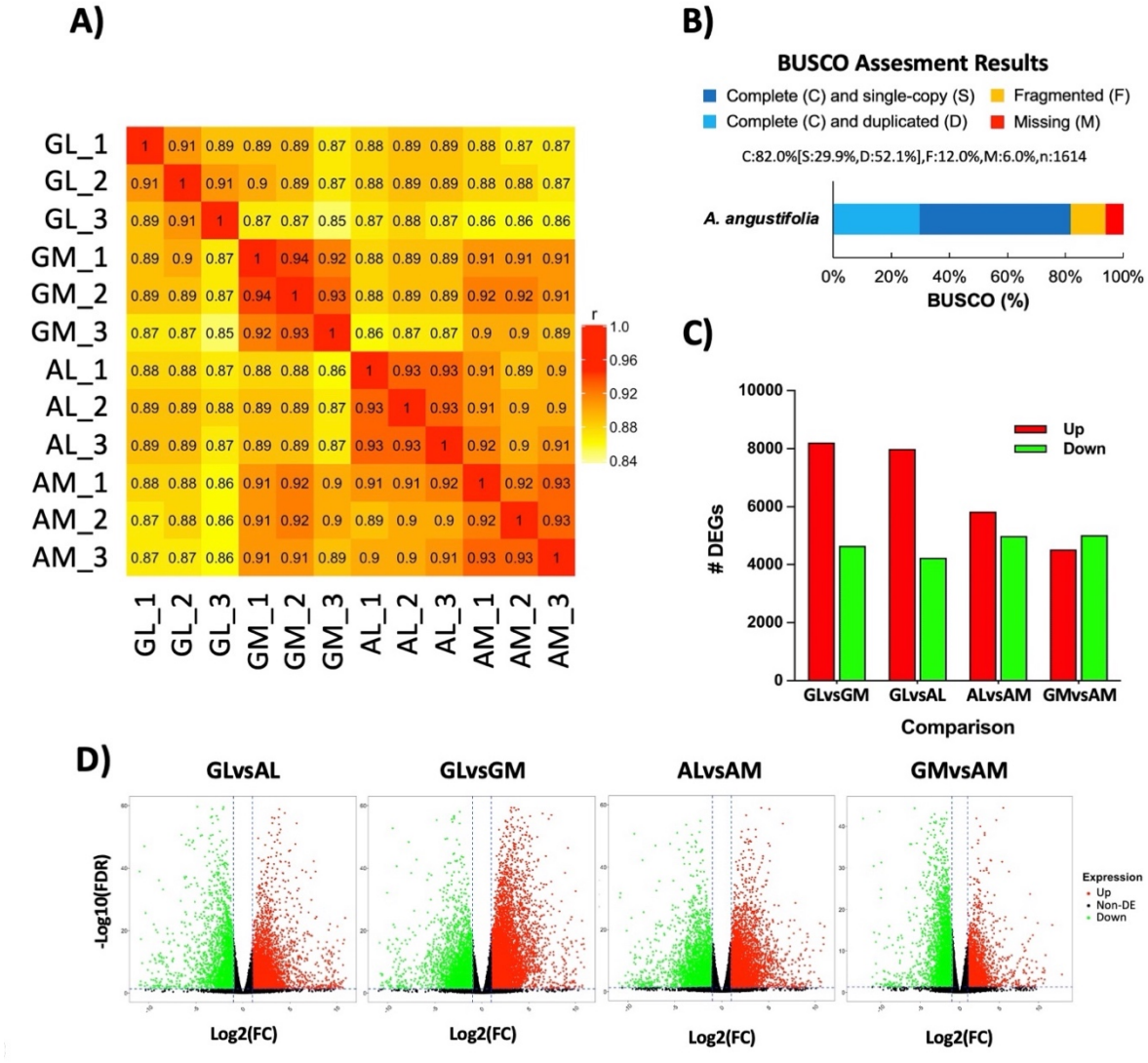


Figure 2.2. Similarity of the replicates, completeness of the assembly, and differentially expressed genes identified in the somaclonal variants of *A. angustifolia*. A) Heatmap shows the correlation between the expression profiles of tissues from somaclonal variants of *A. angustifolia*. The scale colors are assigned based on the calculation of the determination coefficient (r). The numbers after each sample (e.g., HG_1) indicate the number of replicates. **B)** Completeness of the *de novo* transcriptome assembly of *A. angustifolia* using BUSCO. The x-axis (%BUSCO) indicates the percentage of assembled transcripts for each category (complete and single-copy, complete and duplicated, fragmented, and missing). The reference database used was "Embrophyta_odb9" (vers. 4.1.2) with 1,614 core genes in total. The y-axis shows the sample analyzed. **C)** Number of DEGs is shown for each of the examined comparisons. In each comparison, the initial phenotype is used as a reference to determine whether the proteins are expressed more (represented by red bars) or less

(represented by green bars). **D)** Volcano plots of DEGs in tissue comparisons of green and albino somaclonal variants of *A. angustifolia*. The *x*-axis represents the Log2 of the fold-change and the *y*-axis shows the -Log10 of the false discovery rate. DEGs marked with green dots indicate under-expressed genes while those with red dots indicate over-expressed genes in each comparison. GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

2.3.3. Differentially expressed genes in green and albino tissues

Differential expression analysis identified 9,538 DEGs in GM vs AM comparison. Of these, 4,520 and 5,018 genes were over and under-expressed in GM compared to AM, respectively. On the other hand, 10,820 DEGs were identified in AL vs AM comparison. A total of 5,831 and 4,989 transcripts were over and under-expressed in AL compared to AM, respectively. In the GL vs GM comparison was identified 12,851 DEGs. A total of 8,210 and 4,641 transcripts were over- and under-expressed in GL compared to GM, respectively. Finally, 12,222 DEGs were identified in the GL vs AL comparison. Of this number, 7,990 and 4232 transcripts were over- and under-expressed in GL compared to AL, respectively (**Figure 2.2C and D**).

2.3.4. Enrichment analyzes suggest a dysregulation in the ultrastructure and function of the chloroplast in albino plantlets

Enrichment analyses was performed to determine the main BP, CC and KEGG pathways involved in the foliar and meristematic tissues of green and albino plantlets. We focused on results related to changes in chloroplasts and their functions. Analysis of enriched CC terms revealed that chloroplast and its ultrastructure (thylakoid, photosystems, outer and inner membrane, nucleoid, stroma, grana, etc.) were strongly enriched in GL and AL tissues compared to GM and AM tissues, respectively. Notably, even in the absence of mature chloroplasts in albino plantlets, there was a marked enrichment of these terms in AL compared to GL. AM tissue also showed a strong enrichment of these terms compared to GM. Additionally, the terms plastid-encoded plastid RNA polymerase complex, plastid large ribosomal subunit, and those linked to mitochondria showed similar enrichment to that previously described. The chloroplastic endopeptidase *Clp* complex was also enriched in the leaf tissues of both phenotypes. Finally, proteasome-linked terms were strongly enriched only in AL tissue (**Figure 2.3**).

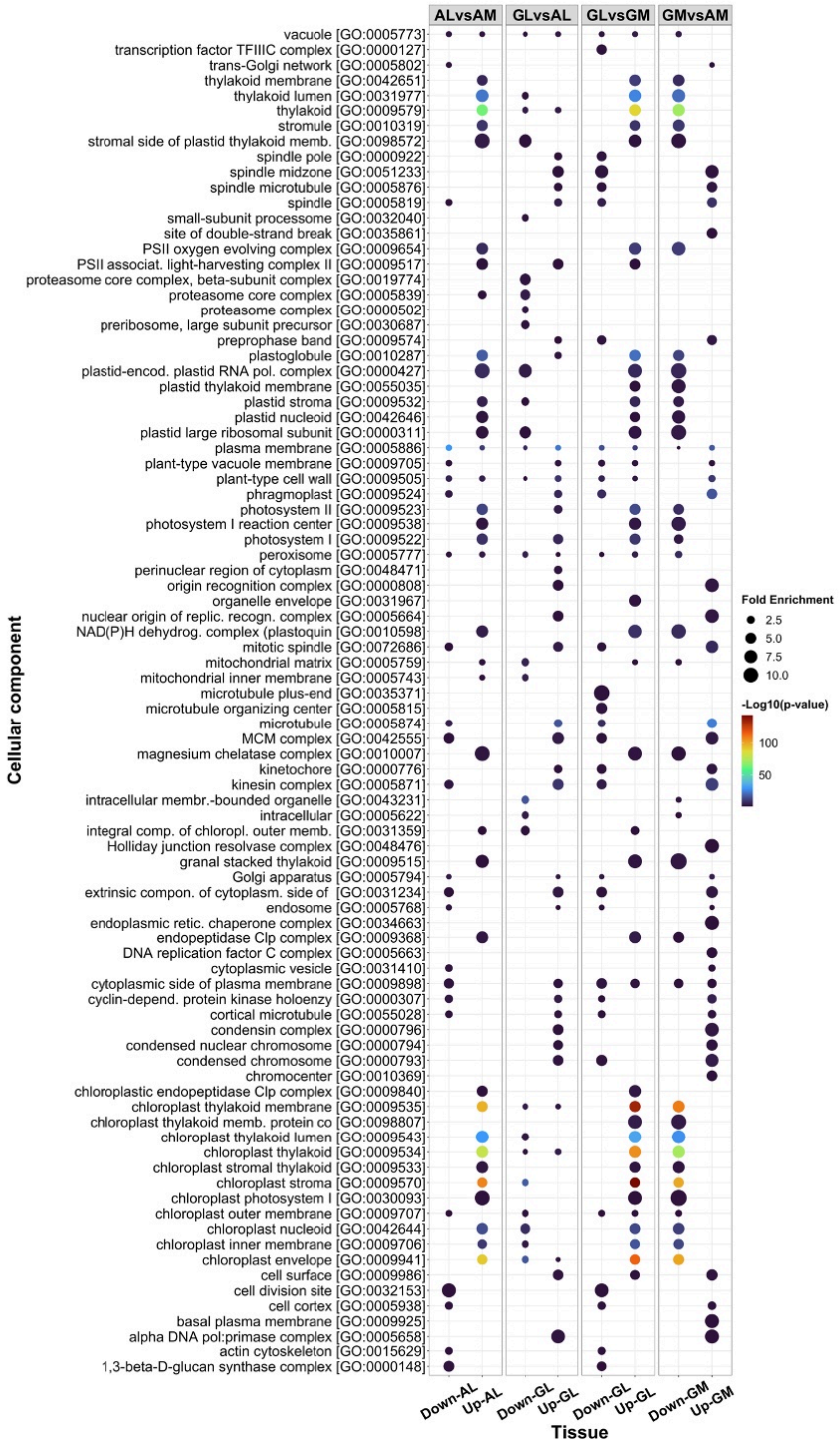


Figure 2.3. Bubble plot of GO cellular component enrichment analysis between tissues of somaclonal variants of *A. angustifolia*. The *y*-axis represents the cellular component terms, while the *x*-axis is organized into four columns based on the four comparisons between tissues: AL vs AM, GL vs AL, GL vs GM, and GM vs AM. The terms enriched down or up relative to the first tissue

indicated in each comparisons are indicated. The size of the bubble represents the fold enrichment, and a color gradient illustrates the -Log10 (*p*-value).

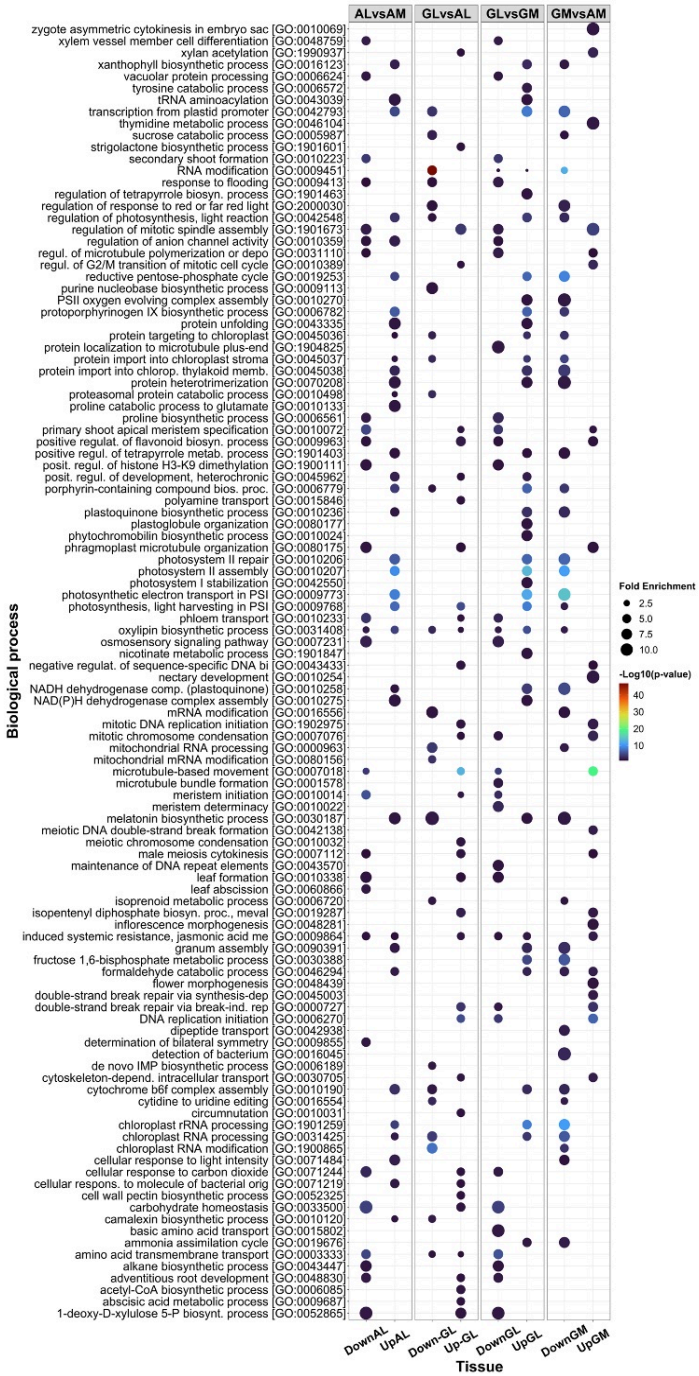


Figure 2.4. Bubble plot of GO biological process enrichment analysis between tissues of somaclonal variants of *A. angustifolia*. The *y*-axis represents the cellular component terms, while the *x*-axis is organized into four columns based on the four comparisons between tissues: AL vs AM, GL vs AL, GL vs GM, and GM vs AM. The terms enriched down or up relative to the first tissue

indicated in each comparisons are indicated. The size of the bubble represents the fold enrichment, and a color gradient illustrates the -Log10 (*p*-value).

Something similar was found in the enrichment analysis of terms linked to BP. The terms related to photosynthesis, photosystems, and chlorophyll biosynthesis were enriched in GL and AL tissues when compared to GM and AM tissues, respectively. Interestingly, the GL vs AL comparison did not reveal any noticeable differences in the enrichment of these terms. AM tissue also showed an enrichment of this group of terms. A similar pattern was found in terms related to protein import into chloroplast and melatonin biosynthesis. However, when comparing the enrichment in the GL vs AL comparison, an enrichment in AL was observed in both cases. Another group of terms that stands out for its enrichment in albino tissues (AL and AM) is linked to RNA metabolism. This process seems to be active both at the mitochondrial and chloroplast level, with the plastid exhibiting the highest enrichment. Similar to CC, the proteasome also showed strong enrichment in AL tissue (**Figure 2.4).**

KEGG pathway enrichment analysis revealed that nitrogen metabolism, proteasome, taurine and hypotaurine metabolism, ubiquinone and other therpenoid-quinone, glutathione metabolism and valine, leucine and isoleucine degradation were enriched in AL tissue. In the last four pathways, enrichment was also observed in AM tissue. Pathways associated with photosynthesis and carbon fixation showed strong enrichment in GL and AL tissues when compared to GM and AM tissues, respectively. AM tissue also showed a strong enrichment of these pathways compared to GM tissue. Finally, the carotenoid biosynthesis pathway was mainly enriched in GL (**Figure 2.5**). These results suggest that the enrichment of pathways related to the ultrastructure and metabolism of the albino plantlet plastids could be compromising the hemostasis and biogenesis of this organelle.

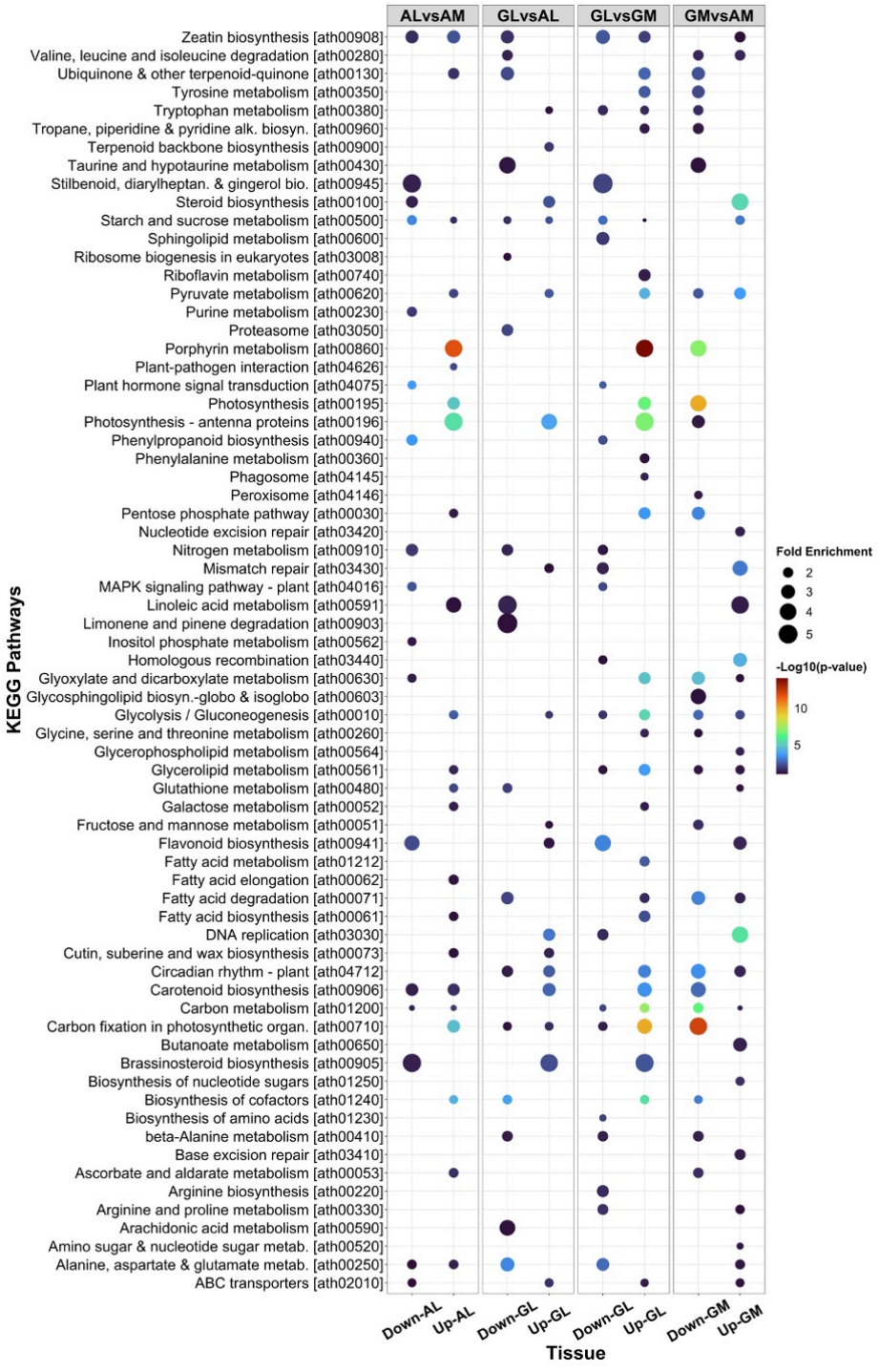


Figure 2.5. Bubble plot of KEGG pathways enrichment analysis between tissues of somaclonal variants of *A. angustifolia*. The *y*-axis represents the cellular component terms, while the *x*-axis is organized into four columns based on the four comparisons between tissues: AL vs AM, GL vs AL, GL vs GM, and GM vs AM. The terms enriched down or up relative to the first tissue indicated in each

comparisons are indicated. The size of the bubble represents the fold enrichment, and a color gradient illustrates the -Log10 (*p*-value).

2.3.5. Expression profiles of genes related with chloroplast biogenesis in tissues of the somaclonal variants of *A. angustifolia*

Albino plantlets may have deregulation in plastid biogenesis, which could explain the absence of mature and functional chloroplasts. To investigate this further, we decided to analyze the expression profiles of genes linked to this process.

2.3.5.1. The albino meristem exhibits earlier and more intense transcriptional activation than the green meristem

A comparison of the transcriptional activity of genes involved in chloroplast biogenesis between the meristems of the green and albino phenotypes (GM vs AM) revealed that AM had a higher transcriptional activity than GM. This higher transcription in AM included 63 genes distributed in all categories of interest except "*peroxidases/reactive oxygen species (ROS)-associated proteins*" (**Figure 2.6**). The genes with the highest expression levels in AM tissue were *FUG1_4*, *RBCS1A_1*, *PSY_2*, *SIG3*, *SBPASE*, *ATPC*, *SIG6*, *REC_2*, *SCO1_2*, *PSBO2_3,* and *PSAE2*. In contrast, the *LHCB1.3_1 and _2, GLK1*, *PHY_4,* and *PDS_1* genes were overexpressed in GM (**Figure 2.7**). In summary, these results demonstrate that AM exhibits a strong transcriptional activity that is turned on earlier than GM.

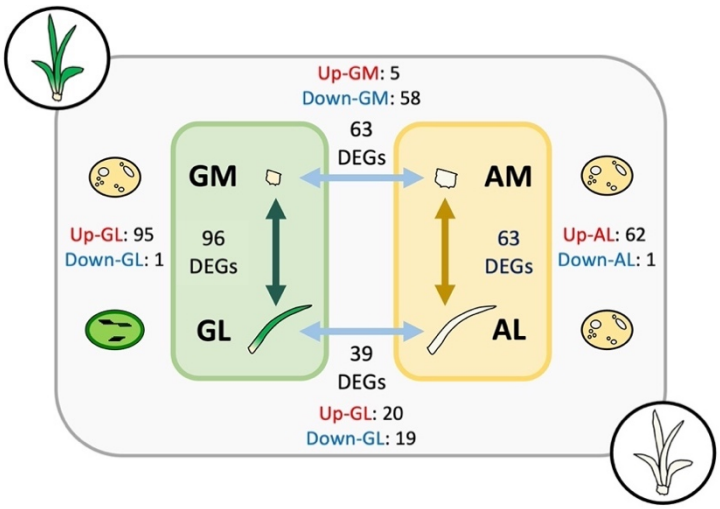


Figure 2.6. General scheme of DEGs identified in different tissues derived from the G and A plantlets of *A. angustifolia*. GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet, Up: up-expressed genes, Down: down-expressed genes.

2.3.5.2. Chloroplast differentiation in the green plantlets demands a strong transcriptional activity

To evaluate genes that are differentially expressed and actively participate in the differentiation of proplastids to chloroplasts in green plantlets, the comparison GL vs GM was taken as reference. A total of 96 DEGs were identified, and all were strongly over-expressed in GL tissue, except the *SIG2* gene, which was down-expressed (**Figure 2.6**). The genes with the highest expression levels in GL tissue were *FUG1_4*, *POR*, *LHCB1.3_1* and *2*, *PSAE2*, *LHCA1_1* and *2*, *RBCS1A_1*, *2*, and *3*, *PSBO2_1*, *2*, and *3*, and *PSY_1* (**Figure 2.7**). These results suggest that the increase in transcriptional activity in GL tissue is due to a high rate of differentiation of proplastids into functional chloroplasts, which has not yet started in the meristem.

2.3.5.3. Blockage in proplastid differentiation in the albino plantlet is associated with reduced transcriptional activity

Similar results to those described in the GL vs GM comparison were found in the albino plantlet tissues. The comparison of albino foliar and meristematic (AL vs AM) tissues identified 63 DEGs distributed in eight of the eleven categories (**Figure 2.6**). The three missing categories were plastid division, PPR proteins, and nuclear regulator of plastid development. Despite the clear increase in the expression of 62 genes in AL tissue, this was not as intense and numerous as that recorded in GL tissue in the comparison GL vs GM, where 95 genes were over-expressed. Only the *SIG5* gene was down-expressed in AL compared to AM tissue (**Figure 2.7**). These data reveal that chloroplast biogenesis could be severely compromised in the albino variant, as a consequence of alterations at the transcriptional level of a group of genes during proplastid to chloroplast differentiation.

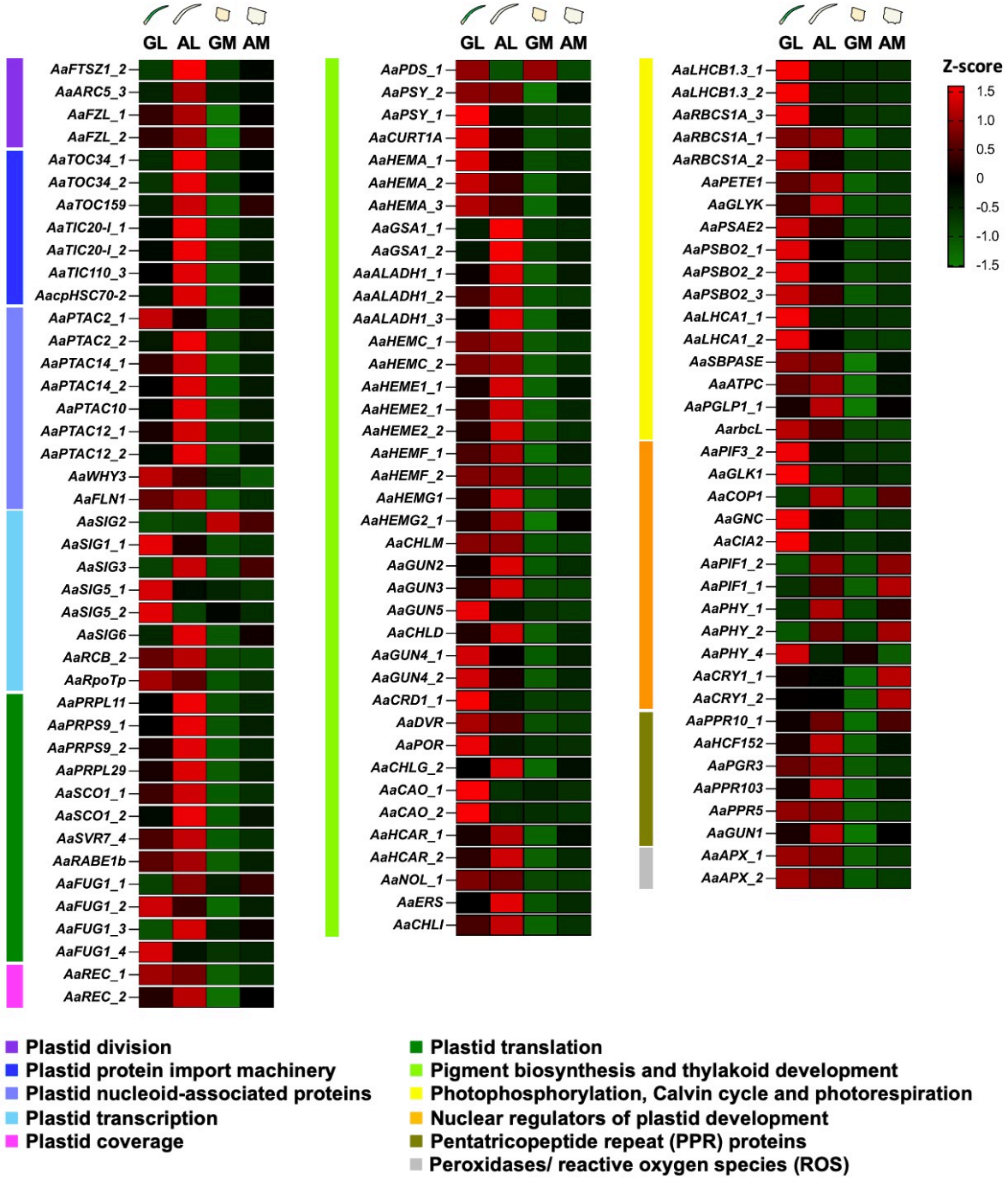


Figure 2.7. Heatmap representing the expression profiles of DEGs involved in chloroplast biogenesis obtained from RNA-seq data. The tissues are shown at the top of the figure, and the transcript identifies are presented on the left side. The color bar indicates the function of the genes. GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

To identify key genes in chloroplast biogenesis in the developmental gradient from meristem to leaf of both somaclonal variants, unique DEGs were selected from the GL vs GM and AL vs AM comparisons. In the GL vs GM comparison, 36 unique DEGs were identified, which were over-expressed in GL tissue. These genes were involved in plastid division (*FZL_1* and *_2*), protein import machinery (*TIC110_3*), plastid nucleoid-associated proteins (*PTAC2_1*), transcription (*RpoTp*, *SIG5_1* and *SIG6*), translation (*FUG1_2* and *_4*), pigment biosynthesis (*PSY_2*, *HEMA_1-3*, *ALADH1_3*, *HEMF_1*, *HEMG2_1*, *GUN4_1-2*, *CAO_1-2*, *CRD1_1* and *HCAR_1*), plastid coverage (*REC_2*), photophosphorylation (*ATPC)*, Calvin cycle (*SBPASE* and *PGLP1_1*), nuclear regulators of plastid development (*PIF3_2*, *GLK1*, *GNC* and *CIA2*), PPR proteins (*PPR10_1*, *HCF152*, *PGR3*, *PPR103* and *GUN1*) and peroxidases/ROS (*APX_2*). In the AL vs AM comparison, three unique DEGs were identified, which were over-expressed in AL tissue: *GSA1_1*, *PTAC2_2* and *ERS*. On the other hand, the unique down-expressed genes identified in both GL vs GM and AL vs AM were SIG2 and SIG5_2, respectively (**Figure 2.8**). These results suggest that albino plantlets may have a transcriptional dysregulation of a group of genes during the differentiation process from meristem to leaf. This dysregulation could compromise the correct proplastid to chloroplast biogenesis.

2.3.5.4. Transcriptional reprogramming in albino plant leaves disconnects chlorophyll biosynthesis and photosynthesis

In the comparison between leaf tissues of both somaclonal variants (GL vs AL), 39 DEGs were identified: 20 and 19 genes were over-expressed in GL and AL, respectively (**Figure 2.6**). The over-expressed genes in GL tissue were involved with plastid transcription (*SIG5_2*), translation (*FUG1_4*), pigment biosynthesis (*PDS_1*, *PSY_1*, *GUN5*, *CRD1_1*, *POR* and *CAO_1-2*), photophosphorylation (*LHCB1.3_1-2*, *PSBO2_1-2*, and *LHCA1_1-2*), Calvin cycle (*RBCS1A_3)* and nuclear regulators of plastid development (*GLK1*, *GNC*, *CIA2* and *PHY_4*). On the other hand, the over-expressed genes in AL tissue were involved in plastid division (*FTSZ1_2* and *ARC5_3*), protein import machinery (*TOC34_1-2*, *TIC20-I_1-2*), plastid nucleoid-associated proteins (*PTAC2_2*), transcription (*SIG3* and *SIG6*), translation (*PRPL11*, *PRPS9*, *SCO1_2*, *FUG1_1* and *FUG1_3*), pigment biosynthesis (*GSA1_1-2*) and nuclear regulators of plastid development (*PIF1_2* and *PHY_1-2*) (**Figure 2.7**). These results reveal that the leaf tissue of the albino plantlet exhibits a reduction in the

expression of genes associated with pigment biosynthesis, photosynthesis as well as nuclear regulators of plastid development. In turn, this tissue exhibits an over-accumulation of genes involved in protein import machinery, transcription and translation.

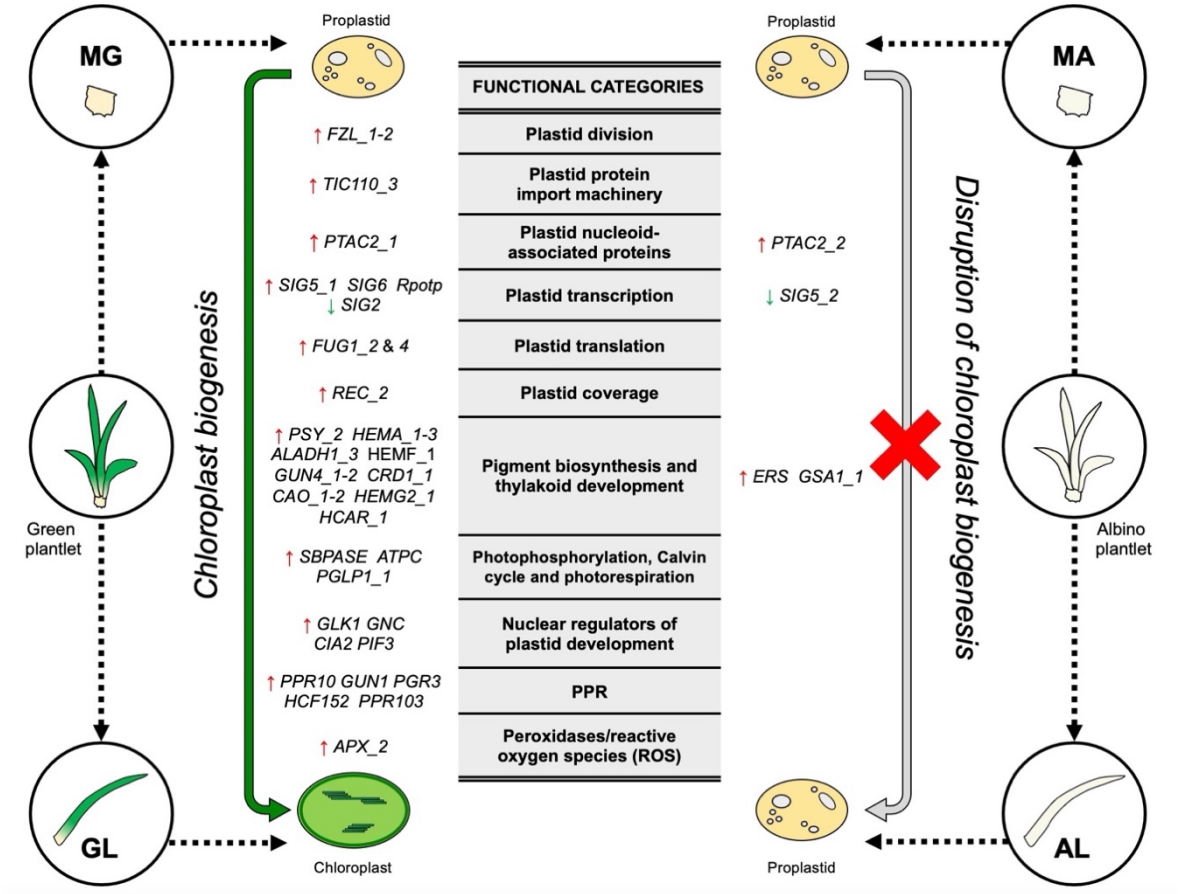


Figure 2.8. General scheme of DEGs identified in comparisons between leaf tissues from green and albino plantlets of *A. angustifolia*. The scheme suggests that the transcriptional differences between GL and AL tissues could play a key role in the disruption of chloroplast biogenesis in the *Agave* albino. GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

2.3.6. qRT-PCR analysis corroborates transcriptomic data generated from the somaclonal variants of *A. angustifolia*

A total of eighteen nuclear-encoded chloroplast genes involved in chloroplast biogenesis were chosen to measure expression by qRT-PCR. The results showed that seventeen genes appeared to be overexpressed in AL tissue compared to the rest of the tissues. The only gene that did not show high expression in AL was *AaGLK* (**Figure 2.9**). This tendency for genes to be overexpressed in AL was similar to that previously described in the RNA-seq data, where thirteen genes were up-expressed in AL. The *AaGLK* gene was also over-expressed in GL as reported in qRT-PCR analysis (**Figure 2.7**). However, *AaSIG5*, *AaPIF1*, and *AaCAO* showed contrasting expression profiles in both analyses. In summary, the qRT-PCR analysis corroborates the reliability of the expression profiles generated in the transcriptomic analysis.

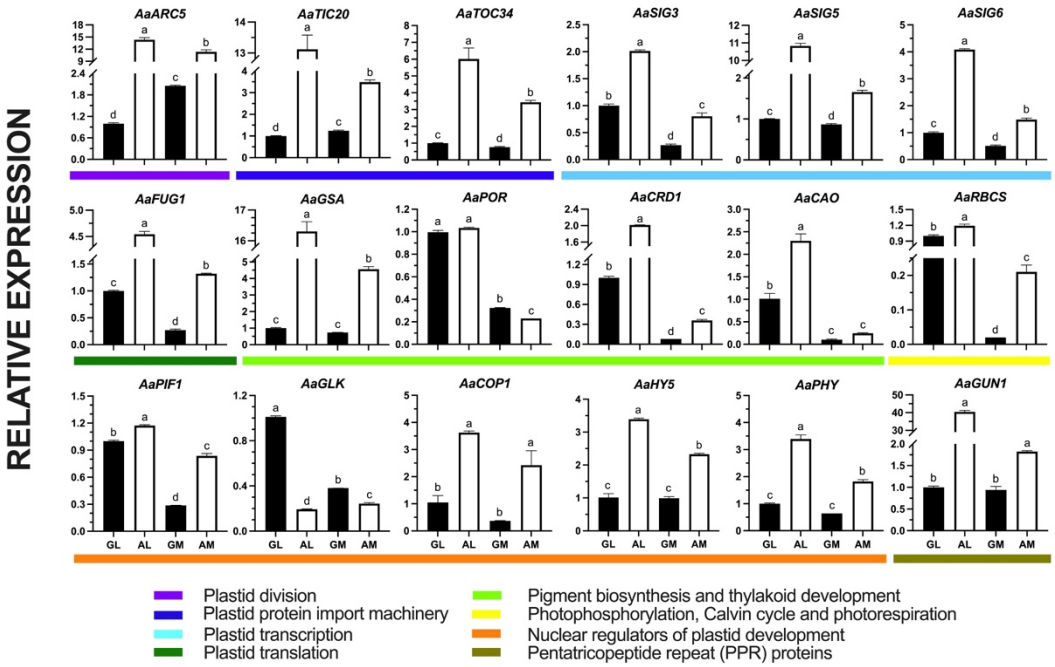


Figure 2.9. Validation of nuclear-encoded chloroplast genes evaluated in green and albino tissues of plantlets of *A. angustifolia* by qRT-PCR. The expression levels for each mRNA were normalized using 2–ΔΔCT, with the GL tissue as the reference expression level. The means and the standard error are shown in each graph. Bars with different letters indicate differences between tissues ($p < 0.05$). GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

2.3.7. Albino plantlets exhibit increased expression of genes linked to plastid ribosomes and plastid-encoded RNA polymerase

A relative expression analysis was performed to assess changes in the expression levels of chloroplast-encoded genes. Genes encoding the large subunit of rubisco (*rbcL*), a subunit E of the plastid ATP synthase (*atpE*), and a structural subunit of the 50S subunit of the ribosomal plastid (*rpl32*) were over-expressed in foliar tissues of both phenotypes, with higher expression in GL tissue. In the meristems of both phenotypes, there was low expression of these genes, with a slight increase in AM tissue. This was similar to the expression of the *psaA* gene of photosystem I, which had a high expression in GL but very low expression in AL tissue. The *psaB* and *psbD* genes, which encode proteins that make up photosystem I and II, respectively, were over-expressed in foliar tissues of both phenotypes, with higher expression in AL tissue. The expression levels in meristems revealed that AM had higher expression compared to GM.

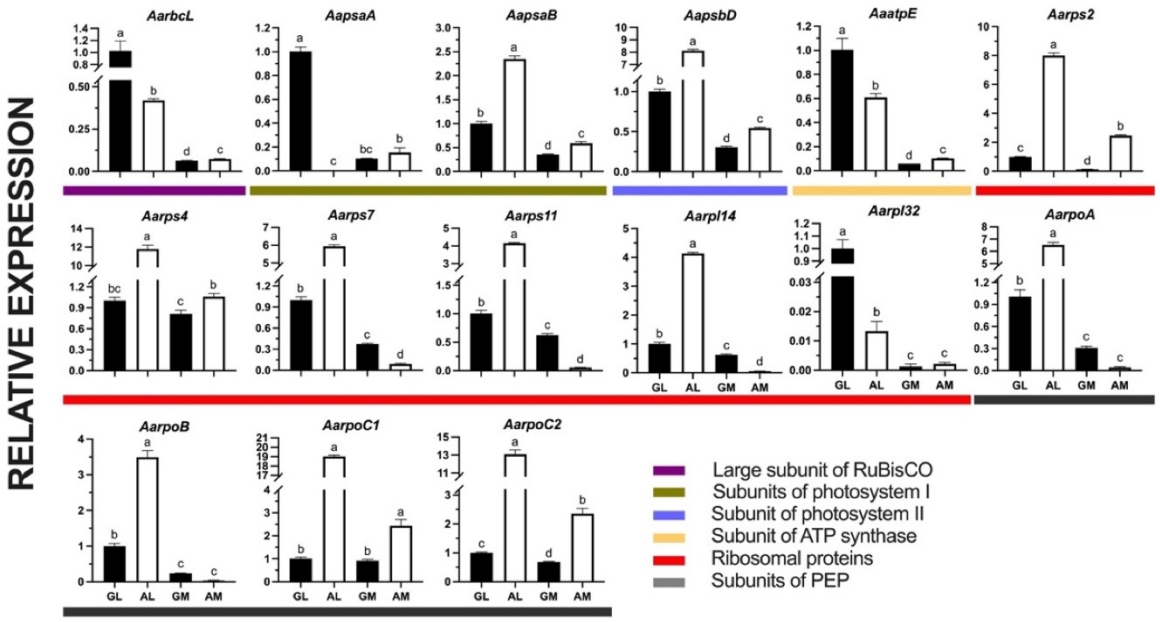


Figure 2.10. Validation of plastoma-encoded chloroplast genes evaluated in green and albino tissues of plantlets of *A. angustifolia* by qRT-PCR. The expression levels for each mRNA were normalized using 2–ΔΔCT, with the GL tissue as the reference expression level. The means and the standard error are shown in each graph. Bars with different letters indicate differences between

tissues ($p < 0.05$). GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

On the other hand, the four genes encoding structural subunits of the 30S subunit (*rps2*, *rps4*, *rps7* and *rps11*) and the 50S subunit (*rpl14*) were over-expressed in AL tissue. Similarly, all four PEP subunits (*rpoA*, *rpoB*, *rpoC1* and *rpoC2*) also showed a strong increase in AL tissue. Only *rpoC1* and *rpoC2* also showed a strong increase in AM tissue, while *rpoA* and *rpoB* did so in GL tissue (**Figure 2.10**). In summary, these results reveal that the leaves of the albino plant have high transcriptional activity of genes linked to plastid ribosomes and PEP, as well as a reduced level for genes linked to carbon fixation, plastid ATP-producing machinery and one subunit of the photosystem I.

2.3.8. Integration of transcriptomic and proteomic data reveals low levels of genes/proteins linked to photosynthesis and carbon fixation in albino plantlets.

An integrative analysis between the transcriptomic and proteomic data of the somaclonal variants was performed to integrate gene expression data with protein accumulation. From the list of genes linked to chloroplast biogenesis, a total of 24 proteins were identified in the proteome of the green and albino phenotypes (**Figure 2.11**). Of this number, fifteen proteins were differentially accumulated, with twelve over-accumulated in green plantlets and three over-accumulated in albino plantlets. The over-accumulated proteins in green plantlets were involved in carotenoid biosynthesis (PDS_2), photophosphorylation (PSBO2_1 and 3, LHCA1_2, LHCB1.3_1, ATPC and PSAE2) and the Calvin cycle (PGLP1_1, SBPASE, RBCL, RBCS1A_1 and 3). In albino plantlets, the over-accumulated proteins were involved in plastid protein import machinery (HSC70-2 and TOC34_2) and peroxidases/ROS (APX_1). Integration of transcriptomic and proteomic data revealed that five genes (*TOC34_2*, *PSBO2_1*, *LHCA1_2*, *RBCS1A_3* and *LHCB1.3_1*) showed correlation between mRNA and protein levels. The remaining nine proteins that showed differences in their accumulation (APX_1, HSC70-2, PDS_2, PSBO2_ 3, ATPC, PSAE2, PGLP1_1, SBPASE, RBCL and RBCS1A_ 1), did not show a correlation with their mRNA levels.

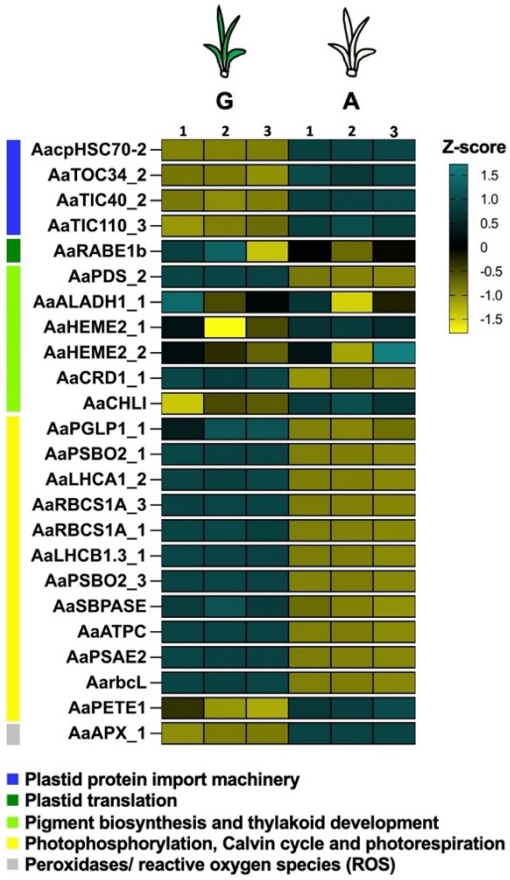


Figure 2.11. Heatmap representing the accumulation of DAPs involved in chloroplast biogenesis obtained from proteomic analysis. The tissues are shown at the top of the figure, and the protein identifies are presented on the left side. The color bar indicates the function of the genes. GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

2.3.9. The meristem and leaf of the albino plant exhibit a high number of the plastid genome

Previous research has shown that plastid multiplication is a highly active process in the meristem of green plants. To meet the high demand for photosynthetic proteins in later stages of development, it is necessary to increase the rate of copies of the plastid genome (Loudya *et al.*, 2021). To determine the changes in the number of chloroplast DNA (cpDNA) copies relative to the nuclear genome (gDNA), a qPCR analysis was performed using specific primers. The results showed that the albino plantlets had a higher number of cpDNA

copies than the green plantlet at the meristematic level. When comparing the leaf tissue (GL and AL) to the meristematic tissues (GM and AM), an increase in the number of copies was registered in leaf tissues (**Figure. 2.12**). However, the leaf tissue from the albino plantlet (AL) had a significantly higher number of cpDNA copies than the leaf tissue from the green plantlet (GL). These results suggest that there is a marked increase in the number of cDNA copies in the tissues of the albino plantlets compared to its green counterpart.

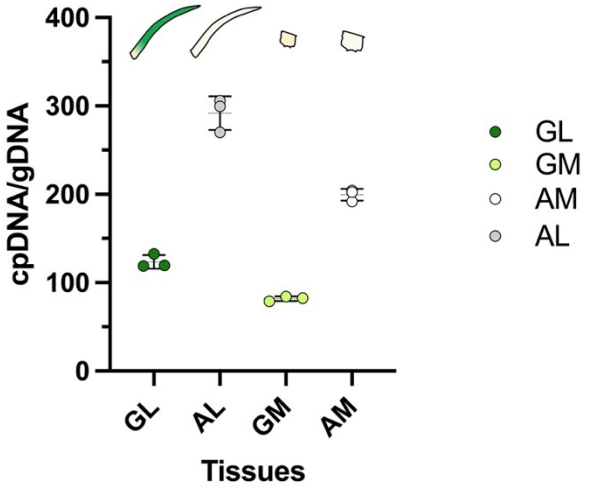


Figure 2.12. Total number of plastome copies per cell in green and albino tissues of *A. angustifolia*. The means and the standard error are shown in each graph. Bars with different letters indicate differences between tissues ($p < 0.05$). GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

2.4. DISCUSSION

Albino plants, which exhibit partial or complete loss of photosynthetic pigments, are an invaluable but underappreciated model for studying chloroplast biogenesis. Transcriptome analysis of albino models has revealed important clues about the metabolic and physiological alterations that these plants undergo in the context of the chloroplast (Yan *et al.*, 2022; Wang *et al.*, 2020a; Li *et al.*, 2018b; Li *et al.*, 2017; Shi *et al.*, 2017; Satou *et al.*, 2014). These studies have identified several plastome and nuclear-encoded genes that are involved in the regulation of chloroplast differentiation, including photoreceptors, transcription factors, genes encoding positive and negative regulators of photomorphogenesis, photosynthetic-associated genes, among others. Our Agave albino,

which was obtained spontaneously and without mutagenesis through several micropropagation cycles, is positioned as a key model for the study of chloroplast biogenesis in monocots (**Figure 2.1**).

Previous transcriptomic studies have been carried out on species of the *Agave* genus, each focusing on different aspects related mainly to their ecological and economic importance (Huang *et al.*, 2022; Morreeuw *et al.*, 2021; Wang *et al.*, 2020a; Huang *et al.*, 2019; Sarwar *et al.*, 2019). However, our study stands out from others because it pays special attention to aspects related to chloroplast in a unique *in vitro* albino model. Throughout the discussion, we will take as reference the transcriptomic map of the developmental gradient spanning from the meristem to the leaf tip of a green wheat plant published by Loudya *et al.* (2021). This map presents the expression profiles of a set of genes grouped based on key functions during chloroplast biogenesis in a monocotyledonous plant, which were also analyzed here. Additionally, due to the similarity of the transcriptomic and proteomic profiles of the *apg2* and *apg3* mutants of *A. thaliana* with our albino *Agave*, we will also mention these mutants throughout this section (Satou *et al.*, 2014; Motohashi *et al.*, 2012). We will discuss our results of chloroplast biogenesis in the albino *Agave* from the two main actors of the process: the nucleus and the plastid.

2.4.1. Plastids are actively multiplying in *Agave* albino plantlets

The gene expression profiles of three nuclear genes involved in plastid division are of particular interest in albino *Agave* tissues. The *FZL* gene is most highly expressed in AM tissue, while the *FtsZ1* and *ARC5* genes are more prominent in AL tissue (**Figure 2.7)**. The *FZL* gene is involved in the biogenesis of thylakoid networks and determines the abundance and morphology of the plastid (Liang *et al.*, 2018; Gao *et al.*, 2006). The reason for its accumulation in AM is unknown, but it could indicate an early cellular effort to assemble thylakoid networks and suggest high plastid multiplication. The high transcription of the *FtsZ1* gene, which together with *FtsZ2*, participates in the formation of the Z-ring for chloroplast division (Liu *et al.*, 2021), could suggest that albino plastids are found in a very active division process. Something similar happens with the *ARC5* gene, which also regulates plastid division in the cytosolic surface of the outer chloroplast envelope membrane (Gao *et al.*, 2003).

The expression profiles of *FtsZ1* and *ARC5* in albino *Agave* tissues are contrary to those reported by Loudya *et al.* (2021), who found that these genes are expressed more highly in the proplastid phase in the meristem and the leaf base, and that their expression decreases near the zone of transition from proplastid to chloroplast. On the other hand, the *apg2* and *apg3* mutants of *A. thaliana* confirm our results by showing a higher expression of the *FtsZ1* and *FZL* genes (Satou *et al.*, 2014). The fact that AL tissue contains numerous plastids that are very similar to proplastids suggests that there is a high plastid proliferation (Hernández-Castellano *et al.*, 2020), which could start from the meristem and continue active at the foliar level. Our transcriptomic data could indicate that the albino plant is pigeonholed in a stage of early development characterized by high plastid multiplication. This could be similar to the idea of the "juvenile" plastid stage proposed for the virescent mutant *cue8* by Loudya *et al.* (2020).

2.4.2. *Agave* albino plastids may accumulate higher levels of the TOC/TIC complex to facilitate greater plastid protein import.

Of the nuclear genes encoding components of the TOC/TIC complex for plastid protein import, *TOC34* and *TIC20-I* stand out. *TOC34* acts as an early receptor of the TOC complex for the import of preproteins, while *TIC20-1* is a subunit of the TIC complex that participates in the formation of the channel through the inner membrane of the chloroplast (Nakai, 2018; Sveshnikova *et al.*, 2000). In our albino model, both genes are overexpressed in AL tissue. Additionally, other genes of the TOC/TIC complex such as *TOC159*, *TOC110*, and *cpHSC70-2*, also showed a tendency to accumulate in albino tissues (**Figure 2.7)**. Proteomic evidence validates the expression profiles of some of these genes, as four proteins of the TOC/TIC complex, cpHSC70-2, TOC34, TIC40, and TIC110, had a tendency to accumulate in *Agave* albino (**Figure 2.11)**. Our results contradict those of Loudya *et al.* (2021), who reported that strong gene expression of the TOC/TIC complex appears in the proplastid phase and decreases shortly before the proplastid-chloroplast transition, reaching its lowest levels in differentiated plastids. On the other hand, similar transcriptomic and proteomic responses have been reported for these and other genes of the TOC/TIC complex in albino *apg2* and *apg3* mutants of *A. thaliana* (Satou *et al.*, 2014; Motohashi *et al.*, 2012). Therefore, the strong transcriptional activity of these genes could indicate that, as a result of the block in the transition from proplastid to chloroplast in AL tissue and coupled with the

high plastid proliferation that maintains this organelle in a "juvenile" state, there is a production of the plastid protein import apparatus similar or greater than that which occurs in the proplastid.

2.4.3. Plastid transcription in albino plantlets is strongly regulated by nuclear genes

One of the functions that deserves special attention in albino plantlets is the transcriptional and post-transcriptional regulation in the chloroplast by nuclear-encoded genes. Of the nuclear genes encoding plastid nucleoid-associated proteins, *pTAC2*, *-10*, *-12*, *-14* and *FLN1* were found to be over-expressed in albino tissues, particularly in AM (**Figure 2.7**). Although the pTACs proteins perform various functions in the nucleoid, those analyzed here participate in the assembly and regulation of the activity of the PEP complex (Chang *et al.*, 2017; Pfalz y Pfannschmidt, 2013; Gao *et al.*, 2011; Pfalz *et al.*, 2005). The expression profiles for the *MurE-like*, *FLN1*, *pTAC2*, *-12*, *-10*, and *-14* genes reported by Loudya *et al.* (2021), revealed a higher expression in the proplastid phase in the leaf base and a gradual reduction shortly before the proplastids/chloroplast stage transition. However, in albino *Agave*, the marked reduction at the foliar level does not occur, but rather the opposite (**Figure 2.7**). Similarly, in other albino plants, such as the *apg2* and *apg3* mutants of *A. thaliana*, high levels of mRNA and protein from *pTAC*s genes have also been reported (Satou *et al.*, 2014; Motohashi *et al.*, 2012).

The over-expression of *pTAC* genes in albino *Agave* tissues could be indirect evidence that PEP is highly active. However, the expression profiles of PEP-dependent genes in our albino model showed both up- and down-expression, which does not allow us to conclude about the PEP status (see below). On the other hand, although the expression of the *pTAC* genes analyzed did not reveal cases of transcriptional repression in the albino plant, this does not rule out the possibility that other *pTAC* genes may be repressed. This should be considered in further studies, as most *ptac* knockout mutants exhibit an albino/chlorotic phenotype (Pfalz y Pfannschmidt, 2013). For example, knockout mutants *ptac2*, *-6*, and *-12* showed reduced expression of PEP-dependent genes, while NEP-dependent genes are either unchanged or over-expressed (Pfalz *et al.*, 2005). In general terms, this behavior seems to be similar to that described in our model (see below). Therefore, we cannot rule out the idea that some *pTAC* genes could be transcriptionally repressed and cause the *Agave* albino phenotype.

Another group of nuclear-encoded genes that regulate plastid transcriptional activity are sigma factors (SIGs). SIG factors are dissociable subunits of PEP that confer specific promoter binding and ensure the transcription of plastome genes (Ortelt y Link, 2021; Macadlo *et al.*, 2019) In albino plantlets, four putative SIG factors stand out: *SIG1*, *SIG3*, *SIG5* and *SIG6*. *SIG3* and *SIG6* accumulated at higher levels in albino tissues. On the other hand, *SIG1* and *SIG5* showed a reduction in their expression, mainly in AL. Only *SIG1* showed over-expression in AM (**Figure 2.7**). However, when comparing the expression profiles of these genes in *Agave* phenotypes with those found in wheat (Loudya *et al.*, 2021), they reveal the same trend except for *SIG3*, where the roles are reversed.

The target genes of these four sigma factors have previously been described in orthologs of other species. For example, SIG1 participates in the transcription of the *psaA*, *psbB* and *psbE* operons and in the abundance of the PSI core protein complex in rice (Tozawa *et al.*, 2007) while SIG3 regulates the transcription of the *psbN* and *atpH* genes in A. *thaliana* (Zghidi et al., 2006). On the other hand, SIG5 regulates the transcription of the *psbD* and *psbA* genes that encode PSII subunits (Tsunoyama *et al.*, 2004; Tsunoyama *et al.*, 2002) and SIG6 various PEP-dependent genes such as *psaA*, *rbcL*, *psbA*, *psbB*, *psbC*, *psbD*, *psbH*, *psbN*, *psbT*, *rrn16*, some *atp* genes and tRNAs (Schweer *et al.*, 2006; Ishizaki *et al.*, 2005). Despite the efforts made in various study models to characterize the targets of SIG factors, a general conclusion is that these genes may have more targets than those reported so far (Chi *et al.*, 2015b). Our results suggest that SIG factors could play a central role in the transcriptional regulation of plastids in albino agave plantlets. Therefore, it is important to further investigate the regulatory mechanisms of each sigma factor, as well as its relevance in chloroplast biogenesis in our model.

Another group of genes studied related to transcriptional aspects were the genes that encode PPR proteins. The main function of PPR proteins is to participate in post-transcriptional regulation mechanisms in the chloroplast (Wang *et al.*, 2023). We found that of the six PPR genes analyzed (*PPR10*, *HCF152*, *PGR3*, *PPR103*, *PPR5* and *GUN1*), three showed a modest increase in their expression in albino tissues, specifically *PPR10_1* and *GUN1* in AM and *PPR5* in AL (**Figure 2.7**). This increase in the expression of PPR genes has also been reported in other plant species with albino phenotypes, such as etiolated and albino cultivars of *Camellia sinensis* (Zhang *et al.*, 2022) and the *apg2*, *cla1* and *apg3*

mutants of *A. thaliana*, (Satou *et al.*, 2014). Previously, in our *Agave* albino model, we have found that this family of genes exhibits increased expression that is dependent on both the state of cellular and plastid differentiation. We have proposed that this overexpression of *PPR* genes could be part of a retro-anterograde compensatory response that attempts to correct the blockage in plastid biogenesis. However, this compensatory response is not sufficient to remove the plastid from is not sufficient to remove the plastid from its undifferentiated state and reverse albinism (Andrade-Marcial *et al.*, 2022).

2.4.4. Dysregulation of plastome expression genes can negatively affect chloroplast biogenesis

The chloroplast is another major player in these signaling processes. The *A. angustifolia plastome* was recently assembled by Qin *et al.* (2021). A total of 132 genes were identified, including 86 are protein coding genes, 38 tRNAs and 8 rRNAS. To identify possible transcriptional alterations linked to genes encoded by the plastome, the expression of a set of these genes was evaluated. Our results revealed that several groups of key plastid-encoded genes, including subunits of the PEP complex, ribosomes, PSI, PSII, ATP synthase, and RuBisCO, are transcriptionally dysregulated in albino *Agave* plantlets.

2.4.5. Albino Plantlet Plastid Transcription: NEP-Dependent or PEP-Dependent?

The plastid transcriptional machinery in angiosperms is a complex system that relies on two types of RNA polymerase: the nuclear-encoded RNA polymerase (NEP) and the plastid-encoded RNA polymerase (PEP). NEP is a monomeric T3/T7 phage-type enzyme that is encoded by a nuclear gene known as *RpoTp*. PEP is a bacterial-type whose core enzyme is composed of four subunits: α, β, β′ and β″. These subunits are encoded by plastome genes called *rpoA*, *rpoB*, *rpoC1* and *rpoC2*, respectively (Park *et al.*, 2022). In both green and albino *Agave* plantlets, the *rpo* genes showed a strong increase in their expression during the transition from meristem to leaf tissue. The greatest increase in expression was observed in the AL tissue (**Figure 2.10**). The *RpoTp* gene also showed a tendency to be more highly expressed in leaf than in meristematic tissue in both plantlets. However, no differences in expression were found at the leaf level (**Figure 2.7**).

The transition from NEP-dependent transcription to PEP-dependent transcription is a critical checkpoint in chloroplast development (Shimizu y Masuda, 2021). NEP regulates the transcription of housekeeping genes, including the *rpo* complex subunits, during the early stages of plastid development. Subsequently, PEP assumes the leading role in transcribing photosynthetic genes (Park *et al.*, 2022; Qiu *et al.*, 2022). This transcription pattern is well-documented in other plant species (Loudya *et al.*, 2021), but it is not clear in *Agave* somaclonal variants, where NEP and PEP seem to be more expressed in the leaf than in the meristem.

The wide variety of regulatory mechanisms acting in our albino model makes it difficult to select a single explanation for what happens at the level of the plastid transcriptional machinery. Of particular interest are the high levels of expression of the *rpo* complex subunits, which could indicate a strong transcriptional activity of NEP in albino plantlets (**Figure 2.7**). On the PEP side, it is noteworthy that despite the strong accumulation of *rpo* subunit mRNAs, low expression of the *psaA* and *rbcL* genes, which are PEP-dependent, has been found (**Figure 2.10**). This could be an indirect sign of post-transcriptional, translational or post-translational regulation mechanisms that are negatively affecting the *rpo* genes and consequently the assembly and activity of PEP. On the other hand, it is also possible that other players associated with PEP activity, such as pTACs and sigma factors, are absent or repressed in our albino model (see above). The fact that the expression of the *psaB* and *psbD* genes, which are also PEP-dependent, do not show the same expression profiles as *psaA* and *rbcL* could support this hypothesis. Finally, it has also been reported that NEP is capable of transcribing PEP-dependent genes, which could indicate that NEP is acting as part of a compensatory response that balances the low activity of PEP (Qiu *et al.*, 2022).

Surprisingly, our results related to expression of plastome-encoded genes are very similar to the proteomic results of the *apg* (Motohashi *et al.*, 2012). In this study, a high abundance of PEP subunits as well as some pTACs was found in albino/chlorotic plants. On the other hand, a low accumulation of photosynthetic proteins, among which RBCL, ATPE, PsaA, PsaB, and PsbD stand out, were also described. These results demonstrate that despite the high accumulation of *rpo* subunits, this does not translate into increased PEP activity. Some of the hypotheses that the authors put forward to explain this behavior

include post-transcriptional regulation, a high protein turnover of photosynthetic proteins, or the regulation of PEP by nucleus-encoded sigma factors.

2.4.6. A retroanterograde response may be acting in the albino *Agave*?

The transcriptional reprogramming observed in the tissues of the albino plant seems to be very similar to that found in the *cue8* mutant of *A. thaliana* reported by Loudya *et al.* (2020). Our results reveal that a retroanterograde correction mechanism could be acting in the albino *Agave* to compensate for defects during early plastid differentiation. This could be very similar to the mechanism proposed for the *cue8* mutant. As in this mutant, the plastids of the albino *Agave* with problems to differentiate send retrograde signals, still of unknown origin, to the nucleus. These signals cause a reprogramming of nuclear gene expression characterized by a reduction in the expression of PhANGs and GLK, an imbalance in the expression of key sigma factors, as well as increased transcription of the gene that encodes NEP. However, unlike the characteristics of the corrective mechanism of the *cue8* mutant, an increase in cpDNA copies was found in tissues of the albino *Agave* plant. On the other hand, although the authors propose that GUN1 probably participates as part of the retrograde response, more studies on our model are necessary to determine the actors responsible for transmitting the signal to the nucleus. Finally, the state of “juvenility” that in the plastids of the *cue8* mutant is temporary, in our albino model seems to be permanent, keeping the albino phenotype unchanged, even in leaf tissues that should contain functional chloroplasts. This in turn could be strongly linked to the high replication of cpDNAs.

CAPÍTULO III

Chloroplastic pentatricopeptide repeat proteins (PPR) in albino plantlets of *Agave angustifolia* Haw. reveal unexpected behavior

Artículo publicado:

Andrade-Marcial, M., Pacheco-Arjona, R., Góngora-Castillo, E., De-la-Peña, C. (2022). Chloroplastic pentatricopeptide repeat proteins (PPR) in albino plantlets of *Agave angustifolia* Haw. reveal unexpected behavior. *BMC Plant Biology*, 22, 352.

3.1. INTRODUCTION

Photosynthetic organisms are essential to support life on our planet for their ability to capture energy from sunlight providing most of the atmospheric oxygen and key chemical compounds (Kaiser *et al.*, 2019; Füssy y Oborník, 2018). In the context of these processes, the main eukaryotic subcellular component involved is the chloroplast, an organelle of endosymbiotic origin (Armbruster y Strand, 2020). The chloroplast in higher plants is part of a diverse group of interconvertible organelles known as plastids. Even though each plastid type exhibits specific metabolic functions in the cell, all of them derive from proplastids, which are characterized by being undifferentiated colorless plastids in cell meristems (Rolland *et al.*, 2018; Solymosi *et al.*, 2018). Although the chloroplast is a semi-autonomous organelle, the majority of the genes necessary to ensure its complete biogenesis do not reside in its plastome (Dobrogojski *et al.*, 2020; Jarvis y López-Juez, 2013). Around 95% of chloroplastic proteins are encoded from the cell nucleus (Shi y Theg, 2013). Within the wide range of proteins encoded by nuclear genes, those that carry out functions related to regulating RNA metabolism in the chloroplast are considered crucial elements for its biogenesis and development (Lee y Kang, 2020). Within this group of proteins known as nucleus-encoded RNA-binding proteins (RBPs) (Lee y Kang, 2020) are the mitochondrial transcription termination factor (mTERF) proteins (Robles y Quesada, 2021), DEAD-Box RNA Helicases (RHs) (Nawaz y Kang, 2017), chloroplast ribonucleoproteins (cpRNPs) (Wu *et al.*, 2021), pentatricopeptide repeat (PPR) proteins (Rovira y Smith, 2019), and others. Within the variety of RBPs, PPR proteins are considered one of the most important players in post-transcriptional processes (Wang *et al.*, 2021; Hicks *et al.*, 2019).

PPR proteins constitute one of the most numerous eukaryotic gene families in plants, with between 400-600 members in terrestrial plant genomes (Fujii y Small, 2011; Lurin *et al.*, 2004). PPR proteins are structurally characterized by presenting tandem arrays of a degenerate motif of ~ 35 AAs (Small y Peeters, 2000). This tandem array is made up of 2 to 26 copies of the PPR motifs reported (P, P1, L1, S1, P2, L2, S2, SS, E1 and E2) as well as other domains such as DYW or the SMR (Cheng *et al.*, 2016; Zoschke *et al.*, 2012; Lurin *et al.*, 2004).

The members of the PPR family have been classified into two subfamilies based on the type of motifs they have. Proteins of the P subfamily have copies only of the canonical P motif. Members of the PLS subfamily exhibit an array of tandem motifs represented by the triad of motifs, P1L1S1 (Barkan y Small, 2014; Rivals *et al.*, 2006; Lurin *et al.*, 2004). In addition, members of the PLS subfamily have at their C terminal a combination of motifs that groups them into classes: PLS, E1, E2, E+ and DYW (Cheng *et al.*, 2016). At a functional level, members of the P subfamily perform functions related to RNA metabolism. These include RNA endonuclease activity, transcript stability, splicing, and translation regulation (Higashi *et al.*, 2021; Wang *et al.*, 2020b; Lee *et al.*, 2019b; Zhou *et al.*, 2017). On the other hand, the subfamily PLS has been mainly connected to editing mechanisms in RNA (Gutmann *et al.*, 2020; Huang *et al.*, 2020).

PPR genes have central roles in organelle biogenesis and development. For instance, *ppr* knock-out mutant plants frequently have photosynthetic dysfunctions such as low levels of chlorophyll and carotenoids (Lv *et al.*, 2020), alterations in the conformation of photosystem I and II (PSI and PSII) (Wang *et al.*, 2020b), partial or total decrease in photosynthetic activity (Lee *et al.*, 2019b), increase in the accumulation of reactive oxygen species (Tan *et al.*, 2014), damage in the chloroplast ribosome biogenesis (Hammani *et al.*, 2016), delayed embryo development (Zhang *et al.*, 2017), lethality of seedlings in early stages of development and abnormal responses to stress (Wang *et al.*, 2016; Tan *et al.*, 2014). These biochemical and structural disruptions have a direct impact on plant phenotype, as in most cases an albino or pale-green phenotype appears (Lee y Kang, 2020).

In recent decades, an attempt to elucidate the factors that could determine the appearance and maintenance of albino phenotypes has been made; however, there is a gap in knowledge about the role of *PPR* genes in a plant without functional chloroplasts (albino). Factors such as the cultivar (Caredda *et al.*, 2000), environmental conditions (Jiang *et al.*, 2020; Yan *et al.*, 2020), growth regulators (Salazar-Iribe y De-la-Peña, 2020; Dewir *et al.*, 2018), incompatibility between the nuclear and plastid genomes (Yao y Cohen, 2000), alterations in plastid DNA (Mozgova *et*

al., 2012) or alterations in chlorophyll biosynthesis pathways (Shi *et al.*, 2017) have been proposed as causes of the emergence of albino phenotypes (Kumari *et al.*, 2009). However, the current approach to studying *PPR* genes has been directed toward the structural and functional description of individual genes in non-albino model plants without a deep understanding of PPR functions. In this study, we present an integrated and novel strategy designed to find and identify chloroplastic *PPR* genes from transcriptome data. Taking into consideration that *PPR* genes are key regulators in chloroplast biogenesis and that the albino somaclonal variant lacks this organelle, knowing the transcriptional behavior of chloroplastic *PPR* genes in the albino plantlet would provide the first clues about their role and possible targets in plants.

3.2. MATERIALS AND METHODS

3.2.1. Plant materials

In this study, two *A. angustifolia* Haw. somaclonal lines that differed phenotypically from each other were used, obtained by micropropagation from plants with the same genetic background (Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016). Plantlets from the phenotypes green (G) and albino (A) (**Figure 3.1**) were cultured in Magenta boxes containing 50 ml of modified Murashige and Skoog (MS) medium (Murashige y Skoog, 1962) supplemented with 2,4-D (0.11 µM) and 6 BA (22.2 µM) and solidified with agar (0.175 %) and gel-rite (0.175 %) (Robert *et al.*, 2006). The plantlets of each phenotype (G and A) were incubated in a growth chamber at 27 ± 2 °C under a 12-h photoperiod (40 µmol/m^{-2}/s^{-1}). Both leaf and shoot apical meristem (referred to as "meristem" throughout the article) tissues were excised using a scalpel blade from 2.5-cm tall plantlets of each phenotype. This resulted in four study conditions: green leaf (GL), albino leaf (AL), green meristem (GM), and albino meristem (AM). The authors have complied with all relevant institutional and national guidelines and legislation in experimental research and field studies on plants.

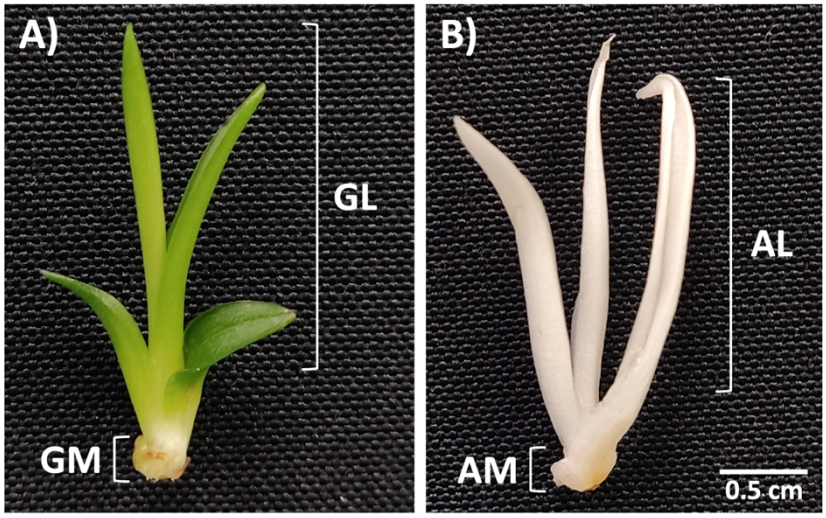


Figure 3.1. The green and albino phenotypes of *A. angustifolia* plantlets. Individual green **(A)** and albino **(B)** plantlets. The meristematic (GM and AM) and foliar (GL and AL) tissues used in this study are indicated in both plantlets.

3.2.2. Searching the *A. angustifolia* transcriptome for PPR sequences

The search for PPR sequences was carried out in the *A. angustifolia* transcriptome results (unpublished data) of GL, GM, AL and AM tissues. The nucleotide sequence data of the *PPR*s for this study were deposited at NCBI in the nucleotide database under accession numbers OM156485 - OM158065. To identify the candidate coding regions and the peptide sequences in the assembled transcripts, TransDecoder software (vers. 5.5.0) (https://github.com/TransDecoder/TransDecoder/wiki) was used, set at a minimum ORF ≥31 AAs long (Gutmann *et al.*, 2020).

The putative PPR peptide sequences were searched using the software HMMER (vers. 3.3.2) (Eddy, 2011) and PPRFinder, Pfam and CDD's profiles. The tool PPRFinder (vers. 1.0) (Gutmann *et al.*, 2020; Cheng *et al.*, 2016) and the *all_PPR.hmm* profile were used to identify PPR motifs in transcriptomic data using *hmmsearch* option from HMMER with the default parameters. The criteria to identify the sequences with motifs of interest was a cutoff score of 0, for SS motifs a score > 10 and for the DYW motifs a score > 30 (Gutmann *et al.*, 2020; Cheng *et al.*, 2016). All the sequences with a single motif were discarded excepted those that contain a complete single DYW functional domain (Gutmann *et al.*, 2020; Boussardon *et al.*, 2012). The other two resources used for the identification of sequences with PPR motifs were Pfam (Mistry *et al.*, 2020) and Conserved Domains Database (CDD, NCBI, July 2020). Six full domain

alignments related to PPR sequences were downloaded from the Pfam database [PPR (PF01535), PPR_1 (PF12854), PPR_2 (PF13041), PPR_3 (PF13812), DYW_deaminase (PF14432) and PPR_long (PF17177)], and seven alignments from CDD [one from TIGRFAM (TIGR00756), three from Protein Clusters (PLN03081, PLN03218, PLN03077) and the accession sd00004]. An HMM profile was built for each alignment using HMMER. These profiles were used to search for PPR sequences with an *E*-value cutoff of ≤1e-10. Venn diagrams were constructed to represent the number of ORFs identified using the InteractiVenn digital tool (Heberle *et al.*, 2015).

3.2.3. Filtering PPR sequences

The recovered sequences were filtered using a sum score cutoff of >40 with PPRFinder. Only the sequences that contained a consecutive array of PPR motifs in the same strand, regardless of the reading frame (RFs), were conserved. The sequences were joined by adding "X" residues to maintain approximate length and indicate the binding site if structural continuity was observed between two ORFs of different RFs (Gutmann *et al.*, 2020). Although these joined sequences are part of the number of sequences reported and analyzed in this study, due to their hypothetical nature they were not considered in final analyses. Additionally, the criteria reported by (Cheng *et al.*, 2016) were taken into account, which allow the analysis to be more rigorous when identifying PPR motifs.

To contrast and improve the prediction and functional annotation of PPR motifs made by PPRFinder, an analysis was carried out with TPRpred software (vers. 1.0) (Karpenahalli *et al.*, 2007) using the default parameters and a score cutoff greater than or equal to 12, and InterProScan 5 (version 5.45-80.0) (Jones *et al.*, 2014) using InterPro's signatures (Pfam, TIGRFAM and PrositeProfiles) to detect probable PPR motifs/domains.

3.2.4. Sequence analysis with DYW and E + domains

Sequences with full DYW domain were analyzed using the MEME suite (vers. 5.3.3) (Bailey *et al.*, 2015). The matrices obtained with MEME were used as a reference to search for conserved regions (PG box, active site, and C-terminal) in the sequences with truncated DYW domain of the E+ class using the FIMO tool with a *p*-value cutoff of ≤1e-5. To visualize the conserved regions in the DYW domain, multiple sequence alignments were performed with the MAFFT tool (vers. 7.471) (Katoh *et al.*, 2017).

3.2.5. Functional annotation

Ortholog genes of the putative PPR sequences were searched by running a local BLASTp (vers. 2.10.1+) (Altschul *et al.*, 1990), using a filtered file of characterized and high-quality plant PPR sequences downloaded from the RefSeq database (NCBI, March 2021) (Pruitt *et al.*, 2011). The local BLASTp search was executed with the default parameters, and sequence pairwise alignments that showed an identity percentage equal or greater than 50% were selected. For sequence alignments with multiple hits, only the best match was selected and reported. The subcellular localization of PPR sequences was predicted using Predotar (vers. 1.04) (Small *et al.*, 2004) and TargetP (vers. 2.0) software (Almagro-Armenteros *et al.*, 2019). In parallel, for sequences that was not possible to determine the subcellular localization using the prediction strategy, resulting Blast hits of their orthologs were filtered for the keyword “chloroplast” to identify hypothetical plastid sequences.

3.2.6. RNA extraction and cDNA synthesis

Expression analyses were performed by quantitative real-time PCR (qRT-PCR). Total RNA from AL, GL, AM and GM tissues was extracted with TRI Reagent® (Sigma-Aldrich) according to the manufacturer’s instructions. After ethanol precipitation, the RNA was resuspended in 30 µL RNA-free water and treated with RNase-free DNase I. The quality of extracted RNA was visualized on native agarose gel at 1.0%. The cDNA was synthesized using oligo (dT)18 with SuperScriptTM IV Reverse Transcriptase (Thermo Fisher Scientific).

3.2.7. Relativequantification by qRT-PCR

All of the PPR sequences identified were filtered and grouped according to the following criteria: 1) complete structure (exhibited a start codon and a stop codon), 2) a hypothetical chloroplast localization inferred from orthologs of other plant species, and 3) differential expression detected in transcriptomic analysis. Twelve putative *PPR* sequences were chosen to validate their expression by qRT-PCR with members of two subfamilies of the PPR family covering the three criteria above described. *Actin*, *tubulin* and *18S* rRNA were used as reference genes. The oligonucleotide pairs used are listed in **Table 3.1** (Suárez-González *et al.*, 2014). Heatmaps were constructed using ComplexHeatmap software (vers. 2.4.3) (Gu *et al.*, 2016) to represent the expression profiles of the PPR sequences.

Table 3.1. Oligonucleotides designed from the twelve PPR transcripts selected for their validation by qRT-PCR.

Unigene	Transcript ID	Oligonucleotide name	Sequence	Amplicon length (bp)
Cluster-46512.134236	*AaPPR1*	AaPPR1F	5'-GTCGCAGGACAACAATGCAGC-3'	167
		AaPPR1R	5'-AGCCTGACACCTTCATCTGGAATG-3'	
Cluster-46512.110771	*AaPPR2*	AaPPR2F	5'-CTCTCCTTCCAGGTGTTTGATGG-3'	172
		AaPPR2R	5'-CAGAAAGGACGCTCACCATGG-3'	
Cluster-46512.98910	*AaPPR3*	AaPPR3F	5'-AGGCTACGCAAGTGTTCGAGAG-3'	132
		AaPPR3R	5'-AGCCAGAGGATTCGAGACTTGAG-3'	
Cluster-46512.116645	*AaPPR4*	AaPPR4F	5'-CTTGAGCAGCACCAGATGACATGA-3'	146
		AaPPR4R	5'-TGCTCTCACGGATATGTTATGGC-3'	
Cluster-46512.94898	*AaPPR5*	AaPPR5F	5'-GTTGGACGACGATGATCACCTC-3'	150
		AaPPR5R	5'-TCACTGCCCTCTCTTCACCACA-3'	
Cluster-46512.118600	*AaPPR6*	AaPPR6F	5'-GCAGCATAAGTGATCCTATTCGGC-3'	187
		AaPPR6R	5'-CAGGTAACTTTGATGGCGCACT-3'	
Cluster-46512.64555	*AaPPR10*	AaPPR10F	5'-GGATCGCAACAAGCTCTTAAACAC-3'	188
		AaPPR10R	5'-CCTACAATGGAGTGATGGATGCC-3'	
Cluster-46512.92555	*AaPPR11*	AaPPR11F	5'-AGGAAGCTTGTTGGTTGGGGTG-3'	172
		AaPPR11R	5'-TGATGCATTGTACCTTGAAGCTCC-3'	
Cluster-46512.113807	*AaPPR13*	AaPPR13F	5'-GGAGAAGCTGTGGAAGGAGATG-3'	143
		AaPPR13R	5'-TCCCTCTCTTCGCCTTGAACTC-3'	
Cluster-46512.145584	*AaPPR15*	AaPPR15F	5'- GAGGGAGAAGGCCTGGAATG-3'	222
		AaPPR15R	5'- CACCTCCAGCAACTCCATGC-3'	
Cluster-46512.166473	*AaPPR18*	AaPPR18F	5'-GAGGAATGTGGTGACGTGGAAC-3'	142
		AaPPR18R	5'-GTTCACCGCATACCCAGACACC-3'	
Cluster-46512.132512	*AaPPR20*	AaPPR20F	5'-CAAGACGTAGCTCAGATCATGAGC-3'	223
		AaPPR20R	5'-GACATGGCTTCAGCTATGAGAGC-3'	
Cluster-46512.119684*	*Actin*	AaACTF	5'-GTCGTACAACTGGTATTGTGCTGGA-3'	158
		AaACTR	5'-GTAACCACGCTCAGTCAGGATCTTC-3'	
Cluster-46512.109058*	*β-tubulin*	AaTUBF	5'-CTCACCTTCTCTGTATTCCCATCC-3'	124
		AaTUBR	5'-CCTCATTGTCAAGAACCATACACTC-3'	
Cluster-46512.119422* Suárez-González et al. 2014	*18S ribosomal RNA (18S rRNA)*	Aa18SrRNAF	5'-GCTACCACATCCAAGGAAG-3'	135
		Aa18SrRNAR	5'-TCGTTAAGGGATTTAGATTGT-3'	

Asterisks (*) indicate the unigenes selected as reference genes.

Table 3.2. Hypothetical RNA targets of chloroplastic PPR proteins of *A. angustifolia*.

Transcript ID	Regular expression	RNA targets	RNA target sequence
AaPPR1	U[AU]UUU[AU]U[CAU][AC]U[AU]UU[ACGU][ACGU]	*ycf1* (CDS)	5'-UAUUUUUCAUUUUCA-3'
AaPPR2	[ACGU][ACGU][UCG][AU][UCG][UCG]A[UCG]G[UCG][AU]GAU[ACGU]	*rpoC1* (intron)	5'-UCUUUUAGGGAGAUA-3'
		*rps12** (intron)	5'-GCUUGGAGGGAGAUC-3'
AaPPR5	[ACGU][CG]A[AU]GAAAA[AU]G[ACGU][AU]A[ACGU][GU]	*trnK-UUU* (intron)	5'-AGAAGAAAAUGGAAUU-3'
AaPPR6	[UC][GC][GC]A[CU][CA]UAG[GU]	*rps16* (intron)	5'-UCGACAUAGU-3'
		petN (CDS)	5'-UGGAUAUAGU-3'
		atpE (CDS)	5'-UGGAUAUAGG-3'
AaPPR10	[ACGU]C[ACGU]UUCC[ACGU]CC	*rpoC1* (CDS)	5'-UCCUUCCUCC-3'
		psbC (CDS)	5'-GCGUUCCCCC-3'
		petA (CDS)	5'-UCUUUCCCCC-3'
		*rrn23** (exon)	5'-GCGUUCCGCC-3'
AaPPR11	[UC][UC][ACGU]A[ACGU]U[ACGU][ACGU][ACGU][ACGU][ACGU][ACGU]C[ACGU][UC]UCUU[CU]U	*ycf2** (CDS)	5'-UUCAAUCCUUUUCCUUCUUCU-3'
AaPPR13	[GU][ACGU][ACGU]C[ACGU][ACGU]G[ACGU][ACGU][ACGU]AUGU	*atpA* (CDS)	5'-UAUCCAGGAGAUGU-3'
		rps14 (CDS)	5'-UGGCAAGAAAAUGU-3'
		*trnI-GAU** (intron)	5'-UGACCCGGAGAUGU-3'
		ycf1 (CDS)	5'-UCUCAAGCAUAUGU-3'
AaPPR15	UUG[CU]CGC	*rbcL* (CDS)	5'-UUGCCGC-3'
		rpl33 (CDS)	5'-UUGUCGC-3'
AaPPR18	UC[GC]UAU[AU][CU][ACGU]G	*ycf2** (CDS)	5'-UCCUAUACGG-3'
		psbD (CDS)	5'-UCCUAUUUGG-3'
		psbC (CDS)	5'-UCGUAUUCUG-3'
		psaB (CDS)	5'-UCGUAUUUGG-3'
		ndhG (CDS)	5'-UCCUAUUUUG-3'

[ ]: character group, e.g. [AGCU] match the characters A,G,C and U; asterisk (*): RNAs generated from duplicated genes present in the inverted repeats (IR) regions of the *A. angustifolia* chloroplast genome; CDS, coding sequence.

The analysis by qRT-PCR was performed using a Rotor-Gene Q (Qiagen). Three ten-fold serial dilutions (10^{-1}, 10^{-2} y 10^{-3}) of cDNA from GL tissue were quantified to generate standard curves for each primer pair. The reaction efficiency was calculated based on the slopes of each standard curve. The efficiency of the oligonucleotide pairs was between 90-110% as recommended. Each qRT-PCR reaction was performed in a final volume of 20 µL using: 0.25 µL of each primer at 10 µM, 10 µL of PowerUp™ SYBR™ Green Master Mix (2X) (Applied

Biosystems), 100 ng of cDNA, and nuclease-free water. The thermocycler program consisted of UDG activation at 50 ºC for 2 min, an initial denaturation at 95 ºC for 3 min, followed by 35 cycles each with 30 s denaturation at 95 ºC, 30 s annealing at 60 ºC, 60 s extension at 72 ºC, and a final step of extension for 5 min at 72 ºC. To analyze, relative expression data was used to perform the 2–ΔΔCT method (Livak y Schmittgen, 2001) using three technical replicates. The relative expression data generated by qRT-PCR were subjected to one-way analysis of variance (ANOVA) ($P \leq 0.05$), and the statistical differences between tissues were obtained by a Tukey post-hoc test ($P \leq 0.05$) using RStudio software (vers. 1.4.1106) (RStudio, 2021). The Graph Pad Prism (vers. 9.2.0) (Graph Pad software, www.graphpad.com) was used to design graphs.

3.2.8. RNA target prediction

To identify the amino acid residues at the 5th and last position for each motif in a PPR protein, PPRFinder (Gutmann *et al.*, 2020; Cheng *et al.*, 2016) and PPRCODE prediction server (vers. 1.6.11) (Yan *et al.*, 2019) were used in twelve validated chloroplastic *PPR* genes. Regular expressions (RegExp) were constructed from the hypothetical sequences of the RNA targets identified by PPRfinder and PPRCODE and searches were carried out in the genes encoded by the chloroplast genome of *A. angustifolia* (data were downloaded from GenBank-NCBI, accession number MW540498) (Qin *et al.*, 2021). The constructed RegExp are described in (**Table 3.2**).

3.3. RESULTS

3.3.1. Putative PPR sequences in the *A. angustifolia* transcriptome

Triplicate RNA samples from green leaf (GL), albino leaf (AL), green meristem (GM), and albino meristem (AM) of *A. angustifolia* (**Figure 3.1**) were subjected to Illumina-based sequencing. *De novo* transcriptome assembly generated 270,163 unigenes, of which 18,829 were differentially expressed (data not shown). The massive identification of sequences harboring PPR motifs was performed in the 462,910 open reading frames (ORFs) predicted from the *A. angustifolia* transcriptome using TransDecoder. The strategy followed here for the search and identification of PPR motifs using PPRFinder, Pfam and CDD's profiles is summarized in **Figure 3.2**.

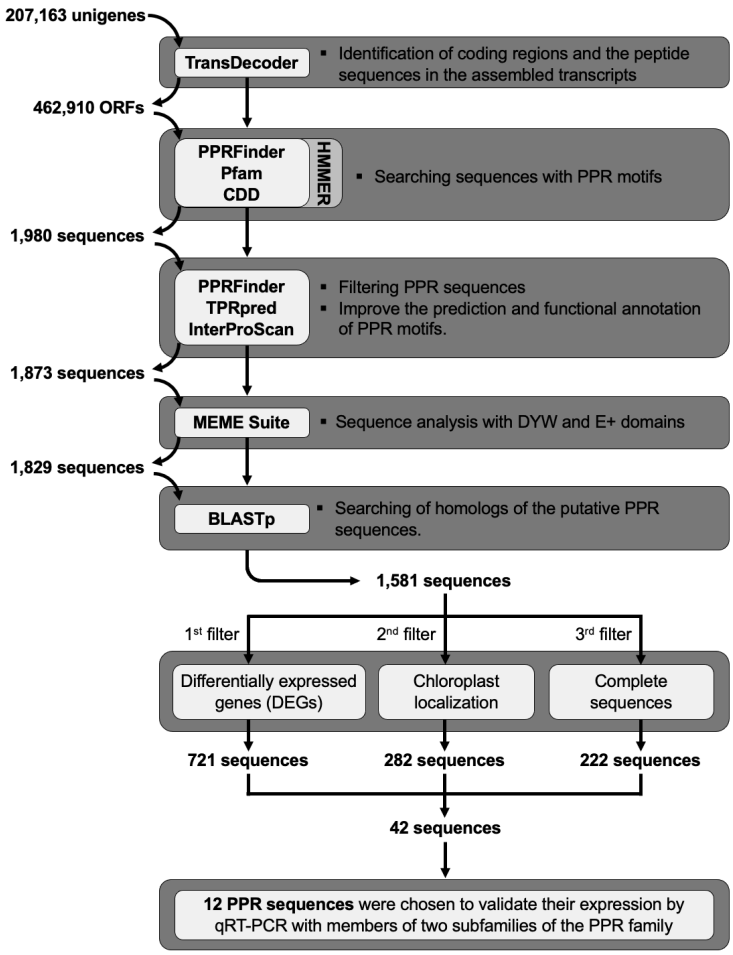


Figure 3.2. Schematic representation of the bioinformatic pipeline followed in the search for the putative PPR sequences.

The analysis with PPRFinder identified PPR motifs distributed in 3,089 sequences. On the other hand, the analysis performed with Pfam and CDD profiles identified PPR motifs in 2,794 and 2,920 sequences, respectively. Thus, the three different strategies helped to identify PPR motifs in a total of 3,232 sequences (**Figure 3.3A**). However, 142 sequences identified with Pfam and CDD profiles were discarded because they did not show a classic array of PPR motifs. This step reduced the number of putative PPR sequences from 3,232 to 3,090 for downstream analyses (**Figure 3.2**).

The 3,090 PPR sequences were analyzed to filter out redundant motifs, and 2,095 sequences had an array of motifs similar to that of a classic PPR protein. Of these, 222 sequences were candidates to be joined due to a structural continuity in their tandem array of PPR motifs,

given a total of 1,980 PPR putative sequences. Additionally, these sequences were analyzed with TPRpred and InterProScan to improve the prediction of PPR motifs (**Figure 3.2**). TPRpred identified PPR motifs in 1,904 sequences that do not belong to other protein families with solenoid-type repeats. InterProScan identified 1,782, 1,754 and 1,930 sequences with PPR motifs using the TIGRFAM, PROSITE and Pfam's profiles, respectively (**Figure 3.3B**). TPRpred and InterProScan reported a small group of sequences without PPR motifs. However, PPRFinder previously classified these sequences as part of the PLS subfamily, specifically of the E2, E+ and DYW classes. In summary, TPRpred and InterProScan appear to have a low capacity to detect sequences with DYW domain and variants of the classic PPR motif. Finally, although a 222 ORFs merge is a suggestion of the designers of the PPRFinder code, these sequences were discarded due to their hypothetical structure. Therefore, only 1873 sequences that contain a single ORF were retained for downstream analyses (**Figure 3.2**).

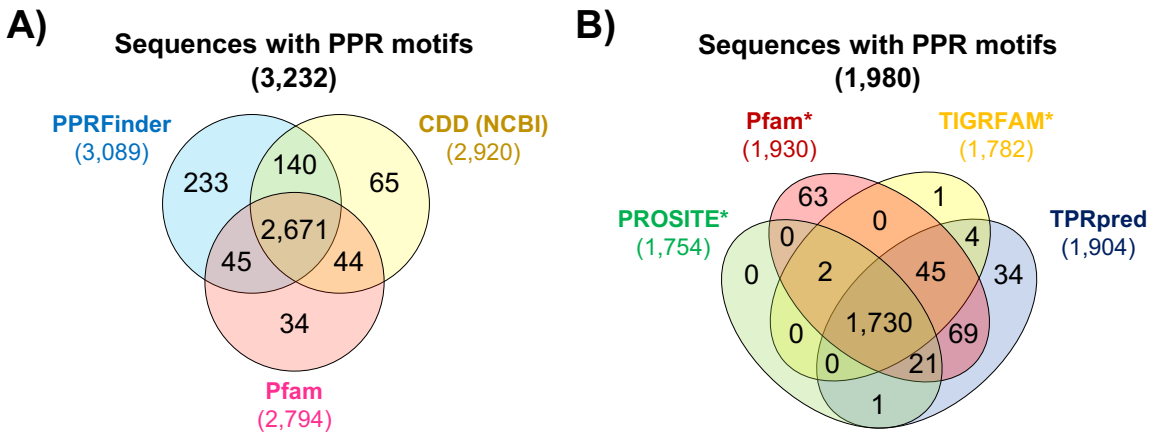


Figure 3.3. Venn diagram of the putative PPR sequences identified in the *A. angustifolia* transcriptome. The diagram **(A)** summarizes the 3,232 sequences that presented PPR motifs and that were identified using PPRFinder, Pfam and CDD profiles. The diagram **(B)** shows the number of sequences that presented PPR motifs using TIGRFAM, PROSITE and Pfam profiles and TPRpred software in the 1,980 previously filtered PPR sequences. The asterisk (*) indicates the databases that were used as part of the analysis in InterProScan 5.

3.3.2. Sequence analysis with DYW and E + domains

An in-depth analysis was performed with the MEME suite to analyze the structure of the DYW domains in the sequences of the DYW and E+ classes (**Figure 3.2**). Out of 1,873 sequences, 231 had DYW domains and 86 had truncated DYW domains. The analysis of the 231 sequences with the DYW domain allowed the identification of three conserved regions: the PG

box, the active site and the C-terminal (**Figure 3.4**). The analysis of each conserved region, along with the alignment of these sequences (**Figure 3.5**), showed that 23 sequences were individual DYW domains lacking the PG box region and other PPR motifs; these were, therefore, discarded from downstream analyses. Sequences of the E+ class showed an incomplete DYW domain; therefore, only the sequences with an arrangement of motifs at the *N*-terminal and an E+ domain with at least the PG box region were retained for downstream analysis. The alignments of the 86 sequences of the E+ class (**Figure 3.6**) along with FIMO analysis showed that 65 of the sequences contain a PG region. The remaining 21 sequences were discarded due to lacking motifs at the *N*-terminal and PG conserved region, reducing the total number of sequences from 1,873 to 1,829 (**Figure 3.2**).

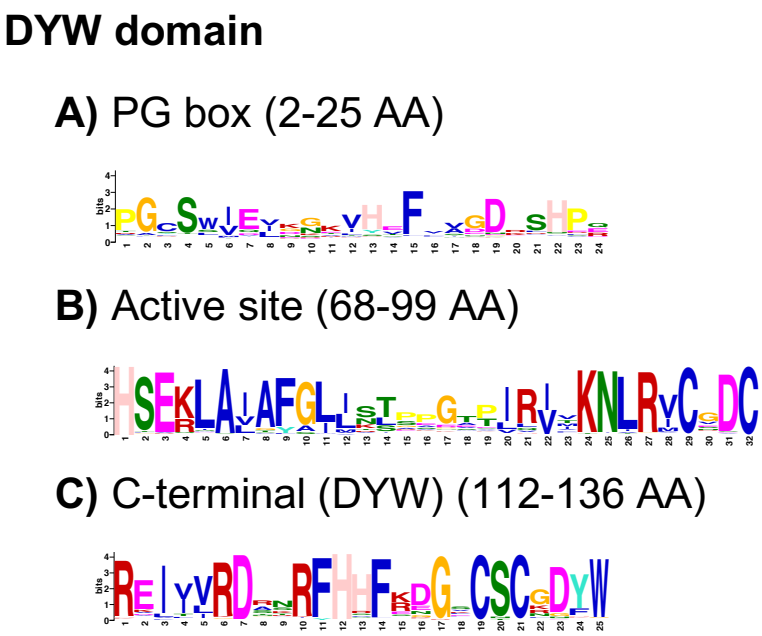


Figure 3.4. Sequence logos for the three regions of the DYW domain identified with MEME. The identification of these regions was carried out using the 231 putative PPR proteins of the DYW class identified in *A. angustifolia*. This domain has a length of ~136 amino acid residues. **(A)** Logo of the PG box region with a length of 24 residues that is located between residues 1-26 of the DYW domain. **(B)** Logo of the region of the active site with a length of 32 amino acids that is located between residues 68-99 of the DYW domain. **(C)** Logo of the C-terminal region with a length of 25 amino acids that is located between residues 112-126 of the DYW domain.

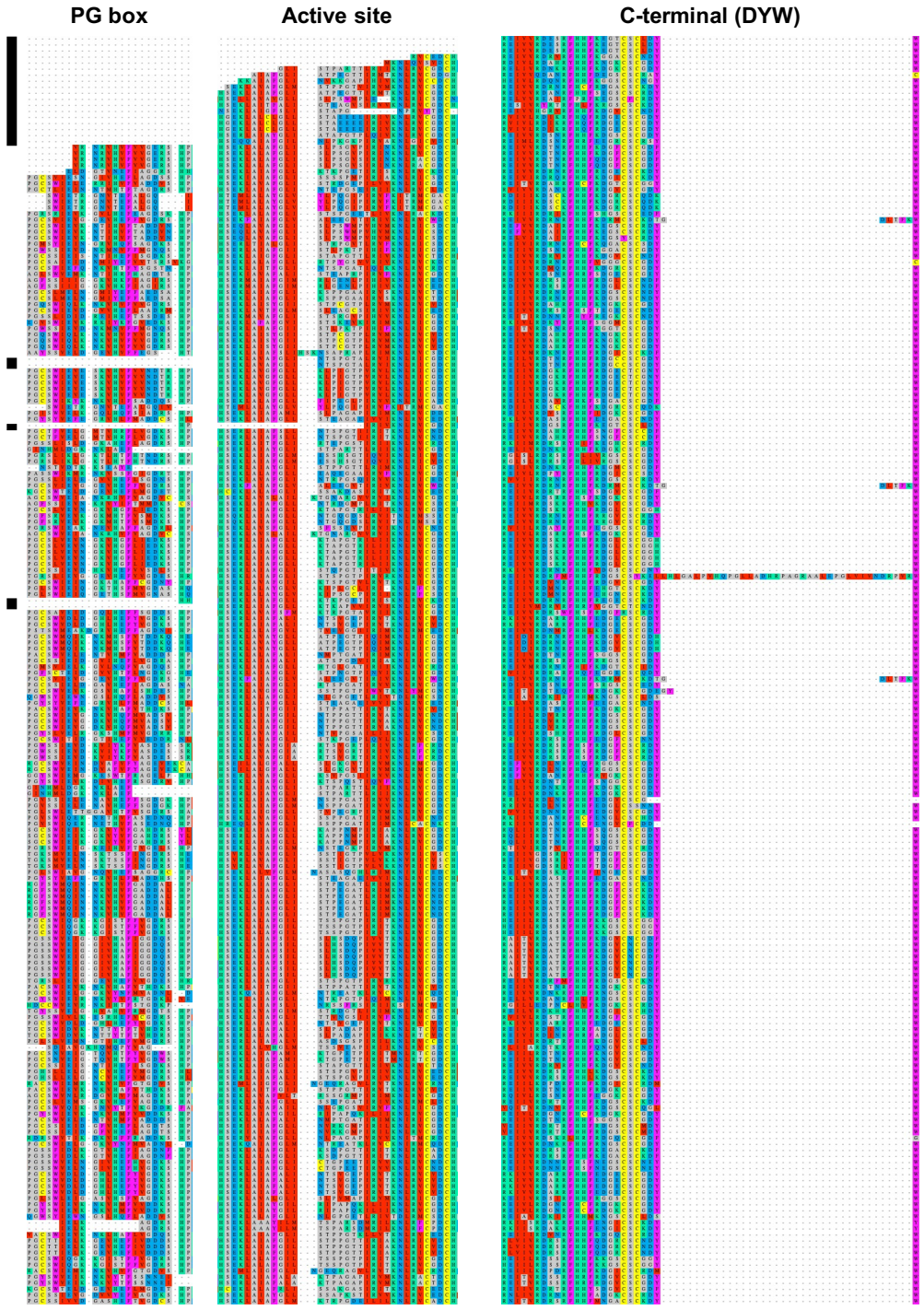


Figure 3.5. Multiple alignment of 232 sequences of the DYW class. Only the three conserved regions of the DYW domain (PG box, active site and C-terminal) are shown in the alignment. The black bars indicate the 23 sequences discarded as exhibiting individual incomplete DYW domains and lacking PPR motifs at the N-terminus.

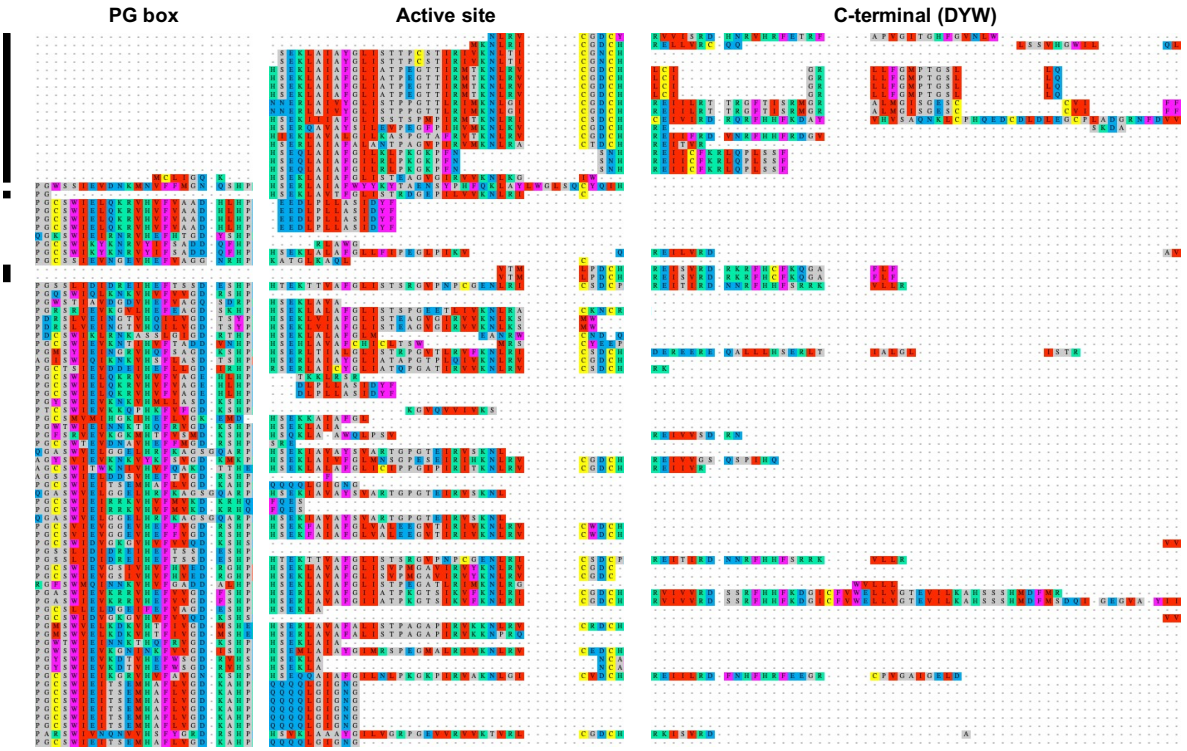


Figure 3.6. Multiple alignment of the 86 sequences of class E+. Only the three conserved regions of the classic DYW domain (PG box, active site and C-terminal) are shown in the alignment. The black bars indicate the 21 sequences that were discarded due to lacking motifs at the N-terminal and the PG box region in the DYW domain.

3.3.3. Functional annotation and structural classification of PPR sequences

In order to identify the closest homolog of the 1,829 putative PPR sequences, a local search was performed against a filtered file of plant PPR sequences downloaded from RefSeq (**Figure 3.2**). The results revealed that 75% of the sequences (1,389) had the best hits with other PPR sequences from monocot species such as *Asparagus officinalis* (39.41%), *Elaeis guineensis* (26.44%), *Phoenix dactylifera* (12.46%) and *Dendrobium catenatum* (6.45%) (**Figure 3.7**). A total of 1,581 sequences that showed structural PPR motifs and an identity percentage equal to or greater than 50% were retained and deposited at NCBI in the nucleotide database under accession numbers OM156485 - OM158065.

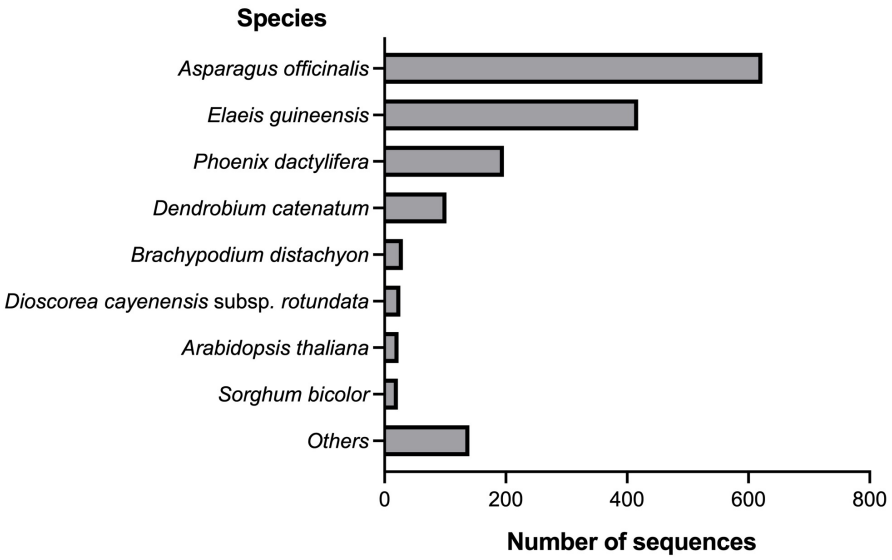


Figure 3.7. Distribution of the number of PPR sequences of *A. angustifolia* with homologues in other species.

The 1,581 putative PPR sequences ranged in length from 58 to 1,370 amino acid (AA) residues, with an average length of 427 AA (**Figure 3.8A**). Furthermore, 758 (47.94%) and 823 (52.06%) sequences were grouped within the P and PLS subfamily, respectively. Within this last subfamily, E2 and PLS classes were the ones that hosted the highest number of sequences with 344 (21.76%) and 227 (14.36%), respectively, followed by DYW with 176 (11.13%), E+ with 63 (3.98%) and E1 with 13 (0.82%) (**Figure 3.8B**). The number of PPR motifs per sequence ranged from 2 to 30 (an exception is the case of sequences with a single DYW domain) with an average of 10 motifs per sequence. In the P subfamily it was common to observe sequences with 3–12 PPR motifs and 3–14 motifs from the PLS subfamily (**Figure 3.8C**).

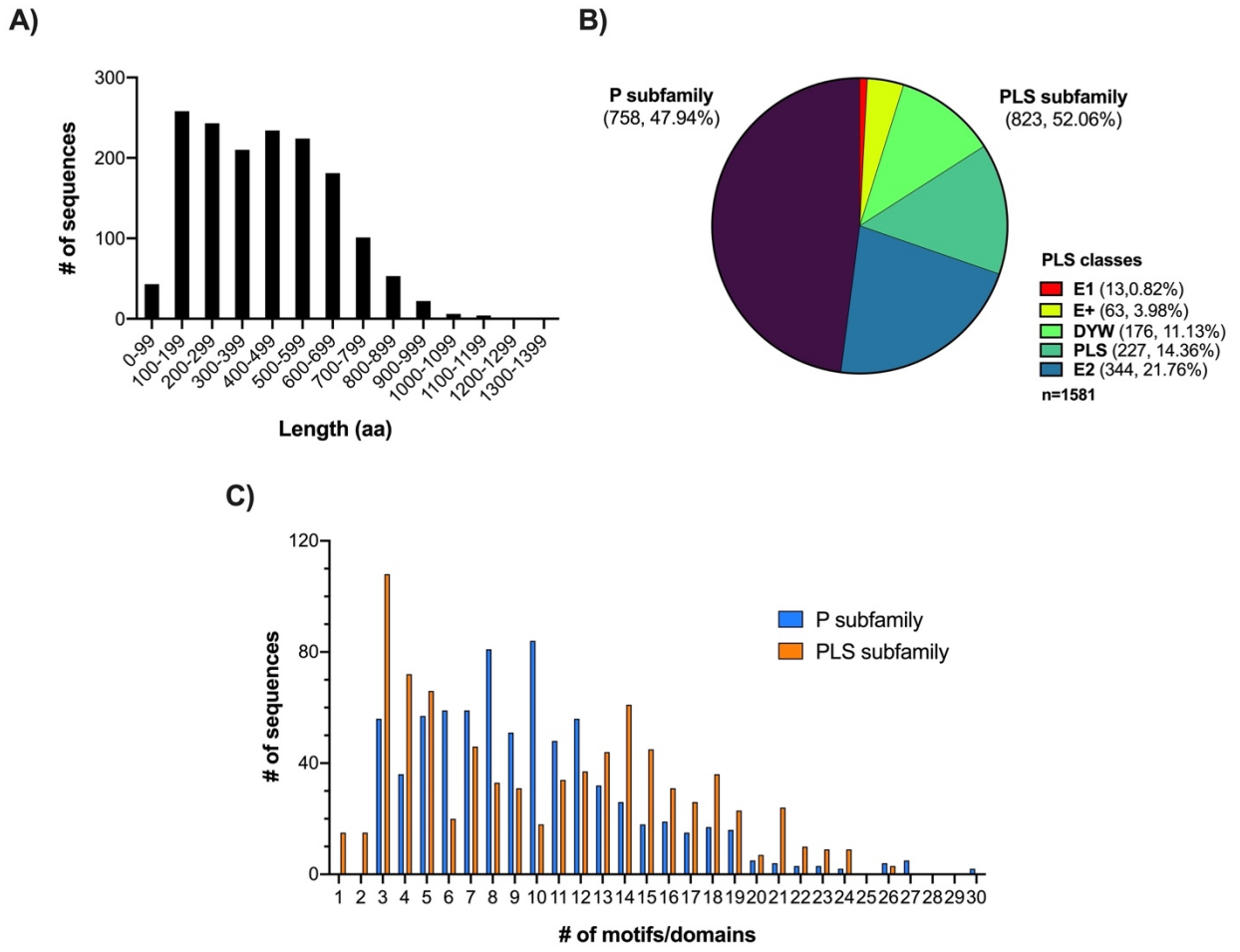


Figure 3.8. Structural characteristics of the 1,581 putative PPR sequences identified in *A. angustifolia.* A) Frequency of the length in amino acid residues of the putative PPR sequences. **B)** Pie chart showing the different subfamilies and classes into which the PPR family is subdivided as well as the number of putative PPR sequences that each one houses. **C)** Frequency in the number of PPR motifs detected per sequence and separated according to the P and PLS subfamily. The nucleotide sequence data of the 1,581 *PPR*s for this study are deposited at NCBI in the nucleotide database under accession numbers OM156485 - OM158065.

3.3.4. *PPR* transcripts differentially expressed in the transcriptome of *A. angustifolia*

From the 1,581 putative PPR sequences found in the transcriptome (GenBank-NCBI, accession number OM156485 - OM158065), a total of 222 were identified as differentially expressed transcripts in the six comparisons between pairs of tissues (GL vs AL, GL vs GM, GL vs AM, GM vs AL, GM vs AM, and AL vs AM) (**Figure 3.2**). The expression profiles of these 222 *PPR* transcripts in the four tissues evaluated (GL, AL, GM and AL) were plotted in a heatmap

(**Figure 3.9A**). The expression profiles revealed that in AL tissue and to a lesser extent in AM, a high percentage of *PPR* transcripts is overexpressed with respect to GL and GM tissues, respectively. The expression levels of most *PPR* in GM tissue are the lowest compared to the rest of the tissues (**Figure 3.9A**).

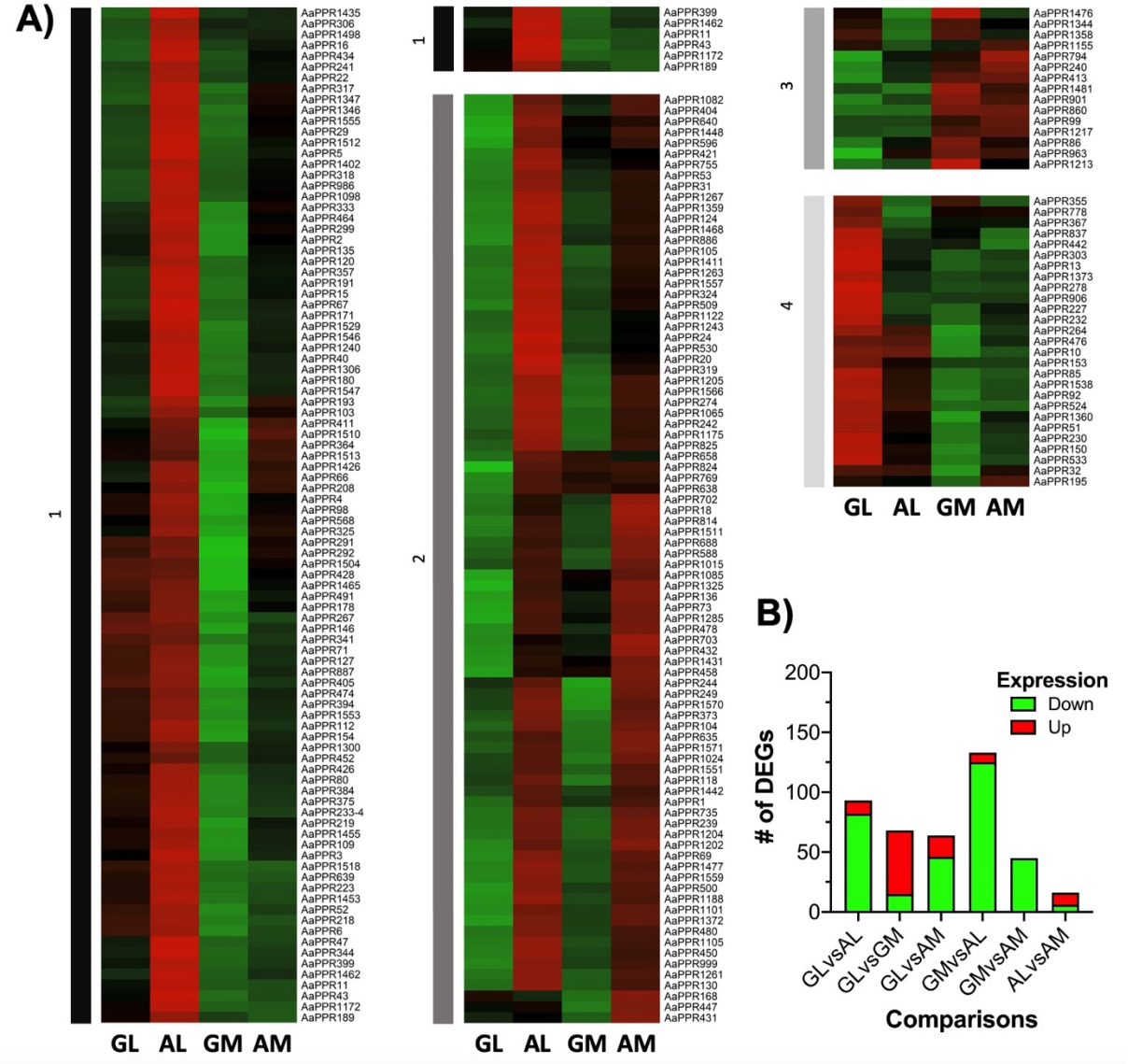


Figure 3.9. Differentially expressed *PPR* transcripts identified in the transcriptomic analysis of *A. angustifolia*. A) Heatmap representing the expression profiles of the 222 *PPR* transcripts differentially expressed in the GL, GM, AL and AM tissues of *A. angustifolia*. At the bottom of the figure the four tissues studied are shown, and the transcript identifiers are on the right. Clustering was applied in the heatmap in order to group the transcripts according to their expression levels. The clusters generated were indicated with bars and numbers on the left (1-4). **B)** Number of differentially expressed *PPR* transcripts in each of the six tissue pair comparisons studied (GL vs GM, GL vs AL, GL vs AM, GM vs AL, GM vs AM and AL vs

AM). In each comparison, the first tissue is the reference to indicate whether the genes are up-regulated (indicated by red bars) or down-regulated (indicated by green bars). The numbers within each bar indicate the number of up- or down-regulated transcripts. GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet. Up: up-regulated genes, Down: down-regulated genes.

Comparisons between tissues from different phenotypes (GL vs AL, GL vs AM, GM vs AL, and GM vs AM) showed that 88, 71, 93, and 100% of the *PPR* transcripts were down-regulated in green tissues, respectively. The number of *PPR* transcripts up- and down-regulated in each tissue pair comparison is summarized in (**Figure 3.9B**). In the comparisons between tissues of the same phenotype, GL vs GM and AL vs AM, the comparison between GL and GM had the highest percentage of overexpressed *PPR* transcripts in leaf tissue (around 78%). On the other hand, AL vs AM was the comparison that showed the lowest number of differentially expressed transcripts, with only 16 (**Figure 3.9B**). Therefore, these results obtained from the transcriptomic analysis (GenBank-NCBI, accession number OM156485 - OM158065) suggest that there is a greater number of *PPR* transcripts overexpressed in tissues of plantlets of *A. angustifolia* with an albino phenotype compared to those with a green phenotype.

3.3.5. Selection of *PPR* transcripts for validation by qRT-PCR

The criteria established for the selection of a set of sequences for its validation by qRT-PCR revealed that of the 1,581 PPR sequences, a total of 721 sequences (45.60%) were found to have a complete ORF (**Figure 3.10A**), and 282 sequences (17.84%) had an orthologue with a chloroplastic site of action (**Figure 3.10B**). Additionally, the 222 differentially expressed transcripts (14.04%) according to the RNA-seq data were also considered. Only 42 PPR sequences fulfilled the three previous criteria (**Figure 3.2**). Of these 42 sequences, a group of twelve were selected to be validated by qRT-PCR. This selection was made considering a balanced representation of the two subfamilies into which the PPR family is subdivided. *AaPPR1*, *AaPPR2*, *AaPPR3*, *AaPPR5*, *AaPPR15* and *AaPPR18* were the selected transcripts of the PLS subfamily. From the P subfamily, *AaPPR4*, *AaPPR6*, *AaPPR10*, *AaPPR11*, *AaPPR13* and *AaPPR20* transcripts were chosen. **Figure 3.11** shows a structural scale representation of the twelve putative *PPR* transcripts grouped according to the subfamilies/classes to which they were classified.

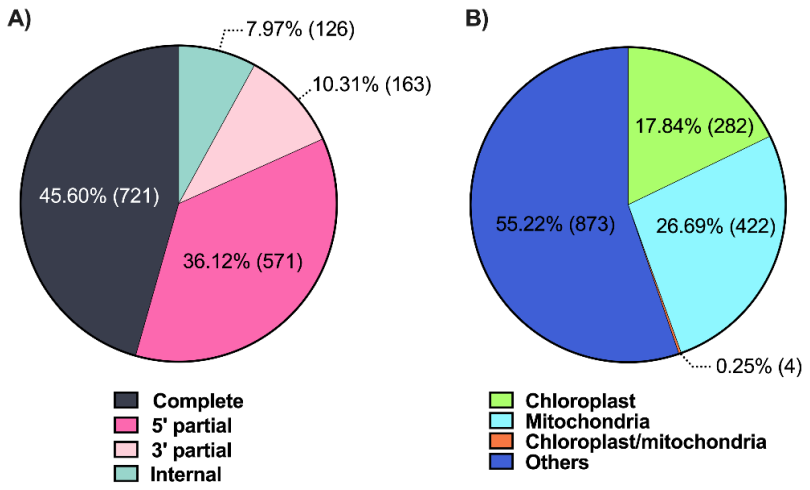


Figure 3.10. Criteria of selection of PPR sequences for their validation by qRT-PCR. A) Pie chart showing the status of the identified ORFs based on the results of the Transdecoder software. **B)** Pie chart showing the hypothetical subcellular location of putative PPR sequences based on information from their RefSeq orthologs identified by BLASTp analysis. The number of sequences in each category is indicated within the parentheses.

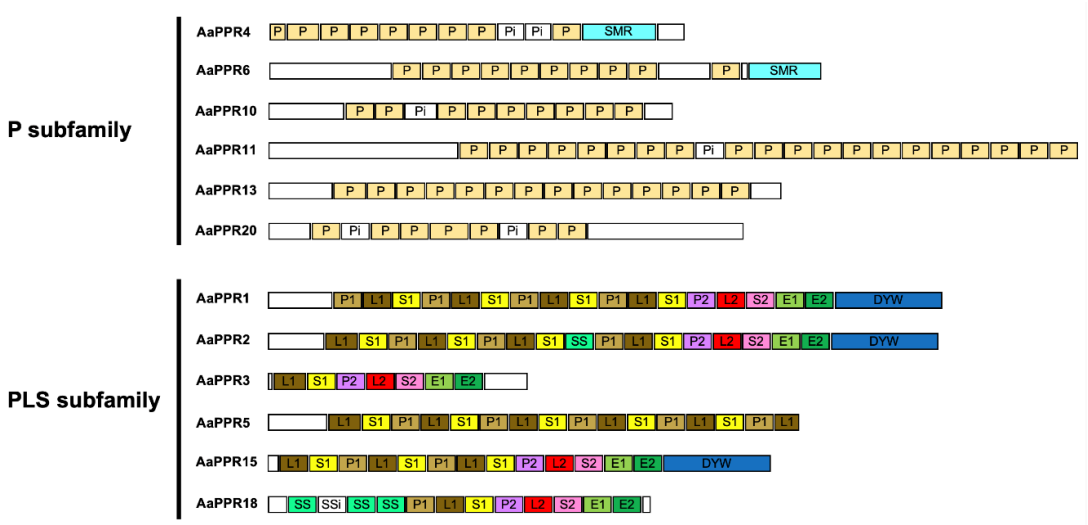


Figure 3.11. Hypothetical structure of twelve chloroplastic *PPR* transcripts selected for validation by qRT-PCR. Diagram showing the distribution of the PPR motifs/domains identified in each of the proteins encoded by the 12 selected *PPR* transcripts. The PPR motifs identified are represented by boxes of different colors with their respective names. The inferred motifs were indicated with an "i" after the corresponding motif name.

The expression profiles of the 12 selected *PPR* transcripts obtained from the RNA-seq data were represented in a heatmap (**Figure 3.12**). Moreover, a summary of the Log2 Fold

change of these genes in each tissue comparison is presented in **Table 3.3**. Differential expression analysis performed on RNA-seq data revealed the following results: in the comparison of GL vs AL, the transcripts *AaPPR1*, *AaPPR5*, *AaPPR15*, *AaPPR18* and *AaPPR20* were over-expressed in AL. In GL vs GM, the transcripts *AaPPR2*, *AaPPR3*, *AaPPR4*, *AaPPR6*, *AaPPR10* and *AaPPR13* were over-expressed in GL while only *AaPPR1* was over-expressed in GM. In GL vs AM, the transcripts *AaPPR1*, *AaPPR18* and *AaPPR20* were over-expressed in AM, and *AaPPR13* was over-expressed in GL. The comparison between meristems (GM vs AM) showed that *AaPPR2*, *AaPPR3*, *AaPPR4*, *AaPPR18* and *AaPPR20* were over-expressed in AM. In the comparison of GM vs AL, the transcripts *AaPPR1*, *AaPPR2*, *AaPPR3*, *AaPPR4*, *AaPPR5*, *AaPPR6*, *AaPPR10*, *AaPPR11* and *AaPPR20* were over-expressed in AL. Finally, in the case of the comparison between AL and AM, there was no difference in the expression of *PPR* transcripts. In summary, RNA-seq data shows that there is a tendency for these 12 *PPR* transcripts (*AaPPR1*, *AaPPR2*, *AaPPR3*, *AaPPR4*, *AaPPR5*, *AaPPR6*, *AaPPR10*, *AaPPR11*, *AaPPR13*, *AaPPR15*, *AaPPR18* and *AaPPR20*) to be more expressed in leaf tissue than in meristematic tissue, and this is much more evident in AL.

Table 3.3. Log2 Fold change values of the twelve chloroplastic PPR sequences selected from RNA-seq data of *A. angustifolia*.

Transcript ID	AL vs. AM		GL vs. AL		GM vs. AM		GL vs. GM		GL vs. AM		GM vs. AL	
	LFC	FDR	LFC	FDR	LFC	FDR	LFC	FDR	LFC	FDR	LFC	FDR
AaPPR1	0.11	9.50E-01	-5.05*	6.50E-06	-1.52	1.80E-01	-3.42*	1.40E-03	-4.94*	8.00E-06	-1.63	1.10E-01
AaPPR2	0.85	1.40E-04	-1.02	3.60E-06	-1.72*	2.70E-14	1.55**	4.60E-12	-0.17	5.40E-01	-2.57*	2.20E-31
AaPPR3	1.39	1.80E-06	-1.05	3.70E-04	-2.29*	1.50E-12	2.64**	6.90E-17	0.34	3.30E-01	-3.69*	5.40E-32
AaPPR4	0.74	5.40E-06	-0.59	3.90E-04	-1.82*	2.40E-30	1.98**	8.30E-36	0.15	4.40E-01	-2.56*	4.10E-58
AaPPR5	1.05	9.20E-07	-1.52*	1.40E-12	-0.77	1.30E-03	0.3	2.70E-01	-0.47	5.60E-02	-1.82*	7.80E-18
AaPPR6	1.27	1.20E-09	-0.42	7.10E-02	-0.77	7.30E-04	1.63**	8.50E-15	0.86	5.60E-05	-2.04*	8.10E-23
AaPPR10	1.07	3.70E-09	0.05	8.40E-01	-0.54	7.10E-03	1.66**	2.10E-20	1.12	4.00E-10	-1.61*	1.50E-19
AaPPR11	1.46	4.80E-04	-0.92	3.10E-02	-0.14	8.40E-01	0.68	1.40E-01	0.54	2.60E-01	-1.6*	5.80E-05
AaPPR13	1.09	7.50E-04	1.48	1.70E-06	-0.87	1.30E-02	3.44**	7.20E-27	2.57**	9.80E-17	-1.96*	3.00E-10
AaPPR15	1.09	3.20E-03	-1.51*	3.20E-05	-1.39	4.20E-04	0.97	1.80E-02	-0.42	3.30E-01	-2.48*	7.10E-12
AaPPR18	-0.68	3.60E-01	-1.72*	9.40E-03	-1.63*	1.40E-02	-0.78	3.00E-01	-2.4*	1.80E-04	-0.95	1.50E-01
AaPPR20	1.24	8.80E-04	-3.71*	5.40E-22	-2.03*	6.80E-08	-0.45	3.40E-01	-2.47*	3.90E-11	-3.26*	1.20E-18

The under and over-expressed PPR transcripts identified in each comparison between tissue pairs are indicated with asterisks (*) and double asterisks (**), respectively. LFC, Log2 Fold change; FDR, False Discovery Rate; GL, green leaf; AL, albino leaf; GM, green meristem; AM, albino meristem.

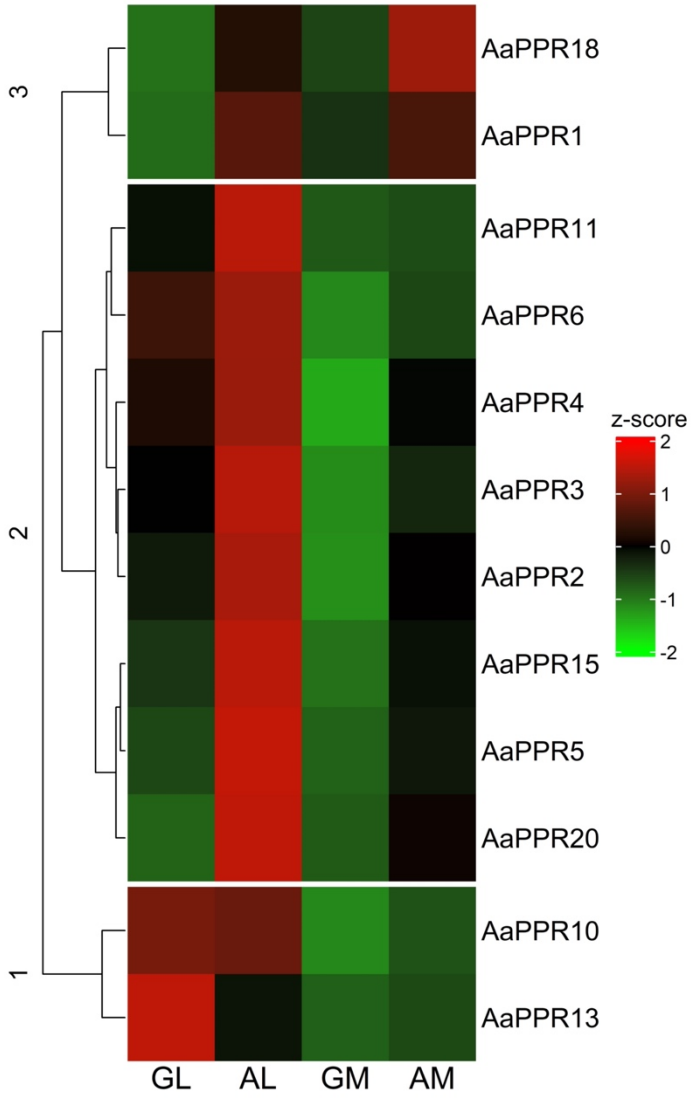


Figure 3.12. Expression profiles of twelve chloroplastic *PPR* transcripts in *A. angustifolia* obtained from RNA-seq data. At the bottom of the figure are shown the four tissues studied, with the gene identifiers on the right. Clustering was applied in the heatmap in order to group the transcripts according to their expression levels. The clusters generated were indicated with the numbers on the left (1-3). GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

3.3.6. Expression analysis of *PPR* transcripts by qRT-PCR

The results obtained from the expression analysis by qRT-PCR revealed that all of the evaluated genes appeared to be overexpressed in AL compared to the rest of the tissues evaluated (**Figure 3.13**). This tendency for *PPR* transcripts to be overexpressed in AL is very similar to that previously described in the RNA-seq data (**Figure 3.12**). AM tissue was the only

one that showed a slight reduction in the expression of five transcripts (*AaPPR1*, *AaPPR2*, *AaPPR10*, *AaPPR13* and *AaPPR15*) in qRT-PCR results compared to the transcriptomic analysis. In the same context, only the *AaPPR10* and *AaPPR13* showed an overexpression in AL tissue, which had not been identified in RNA-seq data.

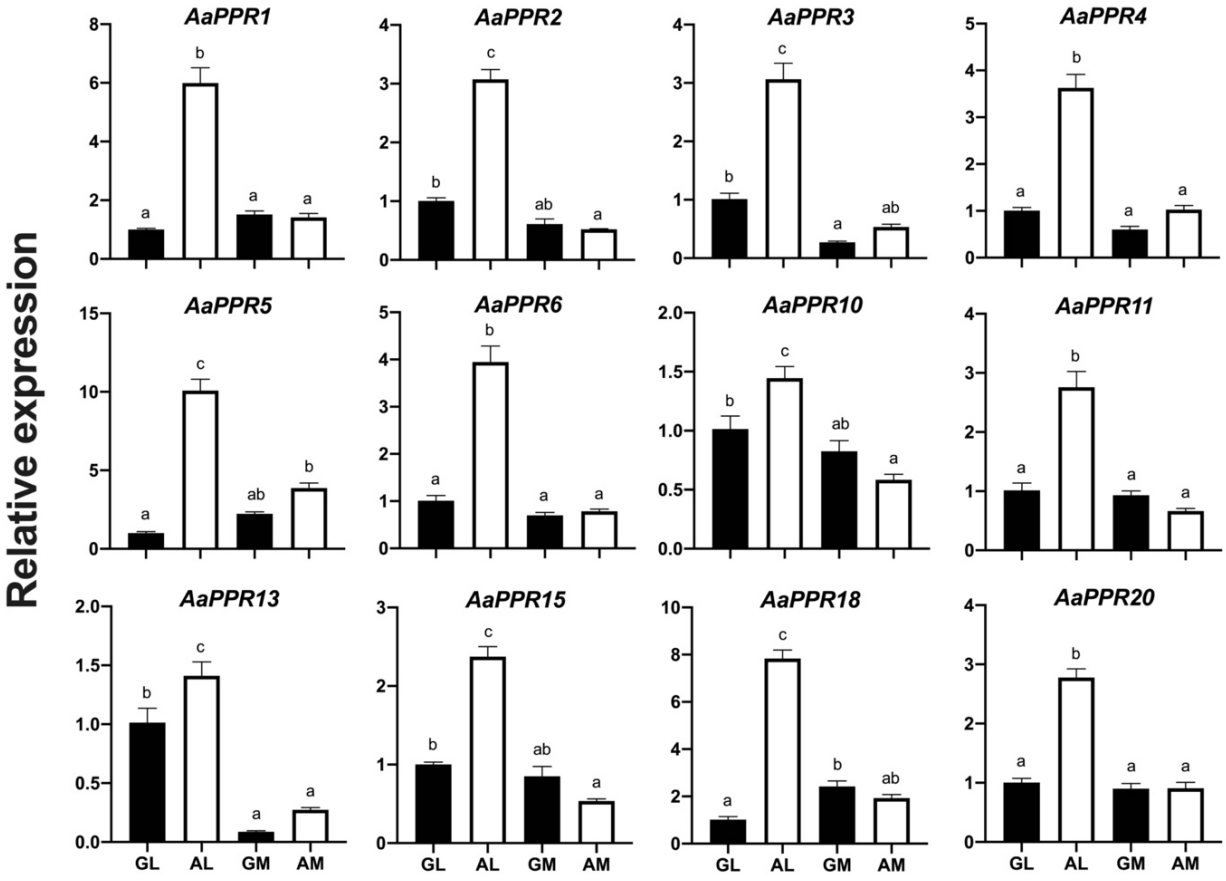


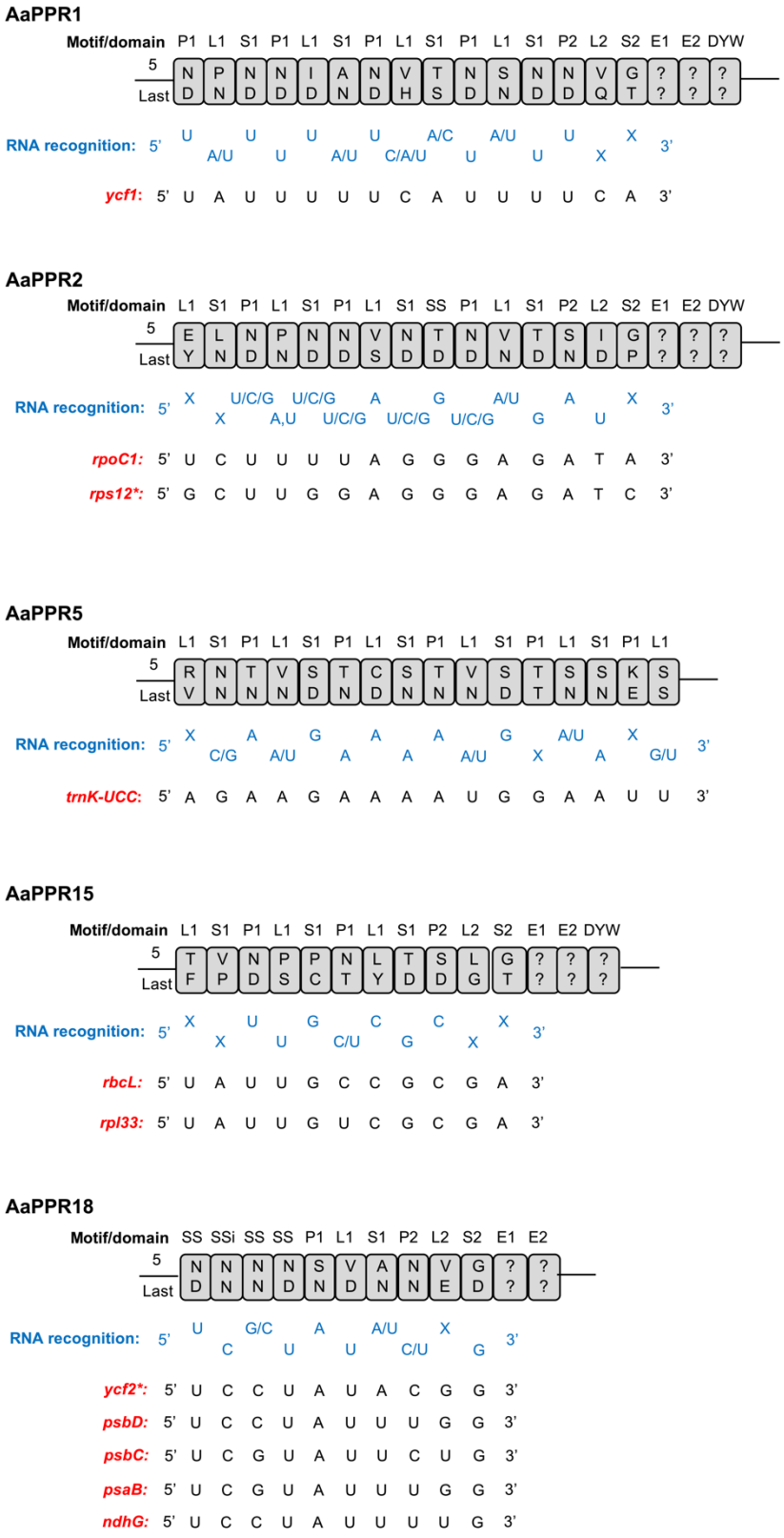
Figure 3.13. Validation of twelve chloroplastic *PPR* transcripts evaluated in *A. angustifolia* by qRT-PCR. Expression levels for each transcript were normalized using 2–ΔΔCT, taking the GL tissue as the reference expression level. The values of the means and the standard error are represented in each graph. The bars with different letters indicate significant differences between tissues according to Tukey's test ($p<0.05$). GL: leaf from green plantlet, GM: meristem from green plantlet, AL: leaf from albino plantlet, AM: meristem from albino plantlet.

3.3.7. Potential RNA targets for PPR proteins

The bioinformatics strategy based on the construction of RegExp allowed for the identification of hypothetical RNA targets for nine of the twelve chloroplastic PPR proteins in the chloroplast genome of *A. angustifolia* (GenBank-NCBI, accession number MW540498) (**Table**

3.2). The RNA targets identified were: *ycf1* for AaPPR1, *rpoC1* and *rps12* for AaPPR2, *trnK-UUU* for AaPPR5, *rps16*, *petN* y *atpE* for AaPPR6, *rpoC1, psbC*, *petA* and *rrn23* for AaPPR10, *ycf2* for AaPPR11, *atpA*, *rps14*, *trnI-GAU* and *ycf1* for AaPPR13, *rbcL* and *rpl33* for AaPPR15 and *ycf2*, *psbD*, *psbC*, *psaB* and *ndhG* for AaPPR18 (**Figure 3.14**). Nine of the RNA targets encode photosynthesis-related proteins: three are subunits of photosystem I (psaB) and II (psbC and psbD), two are components of the cytochrome b/f complex (petA and petN), two are subunits of ATP synthase (atpA and atpE), one is a subunit of NADPH dehydrogenase (ndhG) and the large subunit of RuBisCO (rbcL). Four of the targets encode ribosomal proteins (rps12, rps14, rps16 and rpl33), two are transfer RNAs (*trnI-GAU* and *trnK-UUU*) and one is a ribosomal RNA (*rrn23*). One target is a subunit of plastid encoded RNA polymerase (PEP) (*rpoC1*). Two of the targets have unknown functions (*ycf1* and *ycf2*). The *rps12*, *rrn23*, *ycf2* and *trnI-GAU* genes are duplicated in the chloroplast genome of *A. angustifolia* and both copies encode the same RNA. For the AaPPR2, AaPPR10, AaPPR11, AaPPR13 and AaPPR18 proteins that have these two RNAs as targets, both were identified in this study. The hypothetical target sequences to which AaPPR2, AaPPR6 and AaPPR13 bind are located in introns of the *rpoC1* and *rps12*, *rps16* and *trnI-GAU* RNAs, respectively. The rest of the evaluated proteins have their targets in exons. For the case of the APPR3, APPR4 and AaPPR20, the number of identified targets was very high (with more than 60 RNA targets per protein, data not shown).

A)


AaPPR1
Motif/domain P1 L1 S1 P1 L1 S1 P1 L1 S1 P1 L1 S1 P2 L2 S2 E1 E2 DYW
5 Last
ND PN ND ND ID AN ND VH TS ND SN ND ND VQ GT ?? ?? ??
RNA recognition: 5' U A/U U U U A/U U A/U C/A/U A/C U A/U U X 3'
ycf1: 5' U A U U U U U C A U U U U C A 3'
AaPPR2
Motif/domain L1 S1 P1 L1 S1 P1 L1 S1 SS P1 L1 S1 P2 L2 S2 E1 E2 DYW
5 Last
EY LN ND PN ND ND VS ND TD ND VN TD SN ID GP ?? ?? ??
RNA recognition: 5' X X U/C/G A,U U/C/G U/C/G A U/C/G U/C/G G A/U U A U X 3'
rpoC1: 5' U C U U U U A G G G A G A T A 3'
rps12*: 5' G C U U G G A G G G A G A T C 3'
AaPPR5
Motif/domain L1 S1 P1 L1 S1 P1 L1 S1 P1 L1 S1 P1 L1 S1 P1 L1
5 Last
RV NN TN VN SD TN CD SN TN VN SD TT SN SN KE SS
RNA recognition: 5' X C/G A A/U G A A A/U G X A/U A X G/U 3'
trnK-UCC: 5' A G A A G A A A A U G G A A U U 3'
AaPPR15
Motif/domain L1 S1 P1 L1 S1 P1 L1 S1 P2 L2 S2 E1 E2 DYW
5 Last
TF VP ND PS PC NT LY TD SD LG GT ?? ?? ??
RNA recognition: 5' X X U U G C/U C G C X X 3'
rbcL: 5' U A U U G C C G C G A 3'
rpl33: 5' U A U U G U C G C G A 3'
AaPPR18
Motif/domain SS SSi SS SS P1 L1 S1 P2 L2 S2 E1 E2
5 Last
ND NN NN ND SN VD AN NN VE GD ?? ??
RNA recognition: 5' U C G/C U A U A/U C/U X G 3'
ycf2*: 5' U C C U A U A C G G 3'
psbD: 5' U C C U A U U U G G 3'
psbC: 5' U C G U A U U C U G 3'
psaB: 5' U C G U A U U U G G 3'
ndhG: 5' U C C U A U U U U G 3'

B)

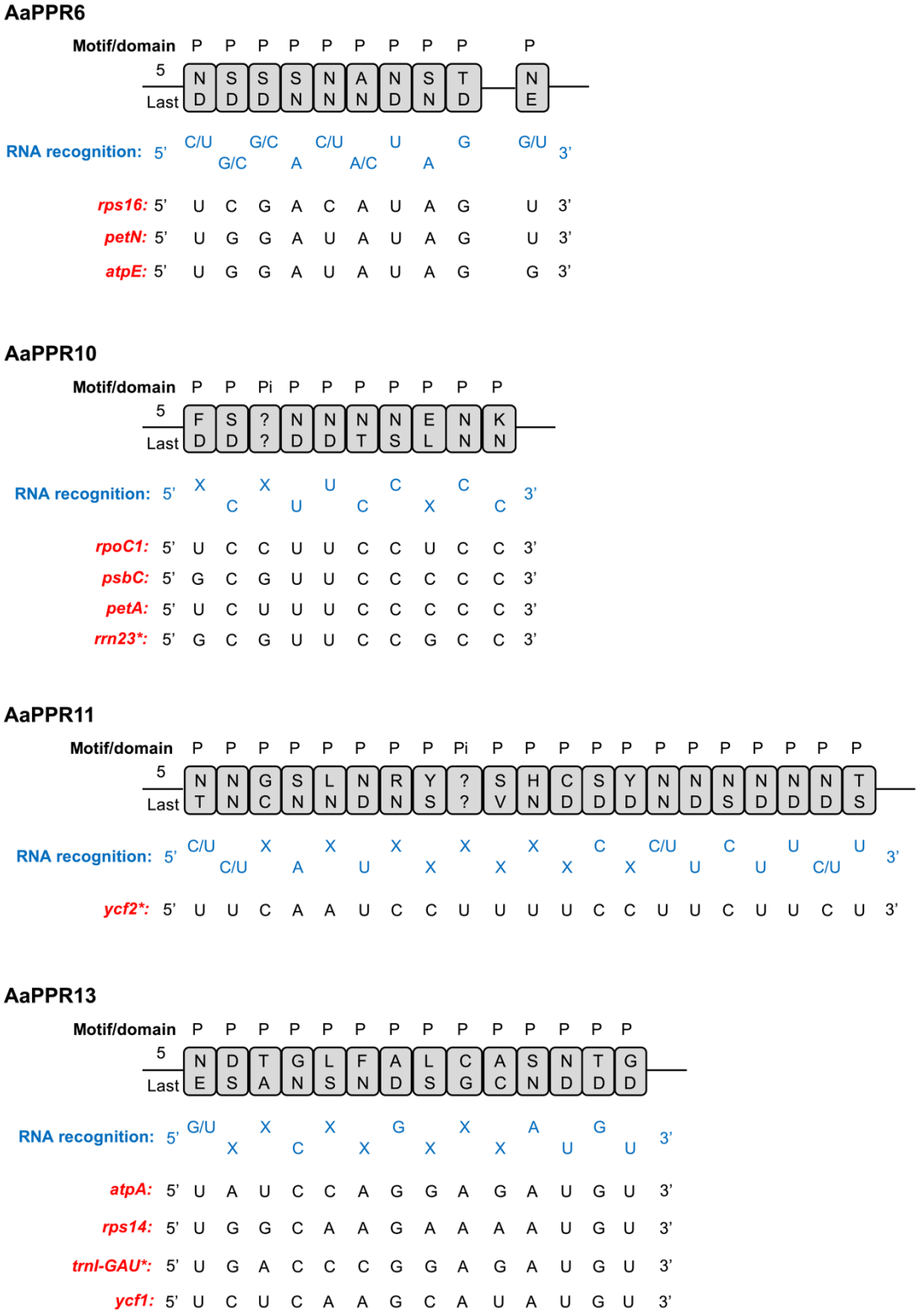

AaPPR6
Motif/domain P P P P P P P P P P
5
Last
ND SD SD SN NN AN ND SN TD NE
RNA recognition: 5' C/U G/C C/U U G G/U 3'
G/C A A/C A
rps16: 5' U C G A C A U A G U 3'
petN: 5' U G G A U A U A G U 3'
atpE: 5' U G G A U A U A G G 3'
AaPPR10
Motif/domain P P Pi P P P P P P P
5
Last
FD SD ?? ND ND NT NS EL NN KN
RNA recognition: 5' X X U C C 3'
C U C X C
rpoC1: 5' U C C U U C C U C C 3'
psbC: 5' G C G U U C C C C C 3'
petA: 5' U C U U U C C C C C 3'
rrn23*: 5' G C G U U C C G C C 3'
AaPPR11
Motif/domain P P P P P P P P Pi P P P P P P P P P P P P
5
Last
NT NN GC SN LN ND RN YS ?? SV HN CD SD YD NN ND NS ND ND ND TS
RNA recognition: 5' C/U X X X X X C C/U C U U 3'
C/U A U X X X X U U C/U
ycf2*: 5' U U C A A U C C U U U U C C U U C U U C U 3'
AaPPR13
Motif/domain P P P P P P P P P P P P P P P
5
Last
NE DS TA GN LS FN AD LS CG AC SN ND TD GD
RNA recognition: 5' G/U X X G X A G 3'
X C X X X U U
atpA: 5' U A U C C A G G A G A U G U 3'
rps14: 5' U G G C A A G A A A A U G U 3'
trnI-GAU*: 5' U G A C C C G G A G A U G U 3'
ycf1: 5' U C U C A A G C A U A U G U 3'

Figure 3.14. Prediction of the potential RNA targets for nine chloroplastic PPR proteins. Each diagram represents an individual PPR sequence. The sequences were ordered by subfamily: PLS subfamily **(A)** and P subfamily **(B)**. The motifs identified in each PPR protein are represented by gray rectangles in tandem. The type of motif is indicated at the top of each rectangle. The inferred motifs were indicated with an "i" after the corresponding motif name. The residues at the 5th and last position that determine nucleotide-binding specificity are shown with capital letters. The most probable combinations of nucleotides recognized by each PPR motif are marked in blue letters. Together these combinations represent the hypothetical sequence of the RNA target, and were considered for the design of the RegExp. The potential RNA targets (marked in red letters) as well as its complete nucleotide sequence identified after the search with RegExp are presented at the bottom of the scheme. Question marks (?) indicate there is no information available to identify the PPR code, “X” indicates any RNA nucleotide and asterisks (*) indicate that the RegExp was identified in two copies of the same gene.

3.4. DISCUSSION

The PPR protein family is characterized by presenting a structure made up of a tandem array of PPR motifs (Lurin *et al.*, 2004). In the present research, we used a wide array of bioinformatic tools such as PPRFinder, HMMER, InterProScan, TPRpred, MEME, and BLASTp to identify putative PPR sequences in somaclonal variants of the non-model plant *A. angustifolia* (**Figure 3.2**). We identified 1,581 putative PPR sequences in the green (G) and albino (A) phenotypes, from which 282 were chloroplastic *PPRs* (**Figure 3.10B**). We also evaluated the quantitative expression of 12 *PPRs* (**Figure 3.13**) and their possible targets (**Figure 3.14**).

PPR proteins are considered central players in plastid RNA metabolism. This is mainly due to the fact that they are very active in the early stages of chloroplast biogenesis, where they play roles as post-transcriptional regulators (Wang *et al.*, 2021; Lurin *et al.*, 2004). We found that chloroplastic *PPR* transcripts in the albino somaclonal variant presented a higher expression in the albino leaf (AL) in comparison with the green leaf (GL) or the meristem tissues (GM and AM) (**Figure 3.12** and **3.13**). This could suggest a nucleus-chloroplast miscommunication during chloroplast biogenesis (Zhao *et al.*, 2020; Jarvis y López-Juez, 2013).

Since PPR proteins are encoded by nuclear genes, their transcription, processing and translation are carried out by the nuclear machinery and finally exported to their organelles of action, whether is the mitochondria or the chloroplast (Saha *et al.*, 2007). The absence of functional chloroplasts in mesophyll cells of the leaf in the *A. angustifolia* albino variant (Hernández-Castellano *et al.*, 2020) opens new research avenues about the functional role of chloroplastic PPRs in the retrograde signaling. *GUN1* (*GENOMES UNCOUPLED 1*), a central

regulator of plastid to nucleus retrograde signaling, is a *PPR* highly expressed in young and expanding leaves of *A. thaliana*, while in mature leaves, stem, and roots its expression is strongly reduced (Wu *et al.*, 2018). This overexpression in young *Arabidopsis* seedlings has been described in other chloroplastic *PPR* genes such as *PDM3* and *AtDPG1* (Zhang *et al.*, 2017; Liu *et al.*, 2016). Therefore, it suggests that the expression of *PPRs* in plants is conditioned by two factors: the state of cellular differentiation and the state of differentiation of the plastid.

Most of the available information on chloroplastic *PPRs* and their functions is from knockout or knockdown mutants in plants (Rovira y Smith, 2019). The partial or total reduction in the expression of chloroplastic *PPR* in these mutants has been associated with the emergence of phenotypes with alterations in their pigmentation, dominated by those with albino and pale-green phenotypes (Lee y Kang, 2020). On the other hand, the alterations have also been associated with other changes. For instance, *osppr16* and *ossla4* mutants in *O. sativa* showed damage to the structure of thylakoid membranes, low accumulation of photosynthetic pigments and disruption of photosynthetic capacity and stomatal variables (Huang *et al.*, 2020; Wang *et al.*, 2018b). In *A. thaliana*, the *atppr4* mutant exhibited seedling lethality under autotrophic growth conditions, alterations in key embryo morphogenetic events and defects in plastid protein synthesis (Tadini *et al.*, 2018). PPR mutants in *Z. mays* such as *emb-7I* showed reduction of plastid-encoded RNA polymerase (PEP) and increased expression of plastid-encoded RNA polymerase (NEP)-dependent chloroplastic genes, respectively (Yuan *et al.*, 2019). We found that the expression profiles of chloroplastic *PPRs* in the albino plantlets of *A. angustifolia* were overexpressed (**Figure 3.12** and **3.13**), unlike the reported pigment-impaired PPR mutants in maize, *Arabidopsis* and rice, in which the expression of chloroplastic *PPR*s is very low or absent (Huang *et al.*, 2020; Yuan *et al.*, 2019; Tadini *et al.*, 2018; Wang *et al.*, 2018b).

Our results indicate that the increase in the expression of chloroplastic *PPRs* in phenotypes with alterations in their pigmentation (**Figure 3.12** and **3.13**) and with numerous undifferentiated plastids, such as the albino plantlets of *A. angustifolia* (Hernández-Castellano *et al.*, 2020), could be closely related to the blockage of chloroplast biogenesis. It was recently found that in samples from the base of the leaf in wheat, where proplastids are very numerous, the *PPR* transcripts present their highest peak of activity; this result demonstrates their central role in the early biogenesis of the chloroplast. This role was confirmed when the plastid begins to differentiate and mature, which leads to a decrease in the expression of these transcripts (Loudya *et al.*, 2021). Furthermore, the transcriptomic analysis of four mutants of *A. thaliana* with different degrees of alteration in their pigmentation, such as *apg2*, *cla1*, *apg3*, and *ch42*, support the

relationship between the state of chloroplast biogenesis and the expression of *PPR*s. The *apg2*, *cla1* and *apg3* mutants, which showed a strong reduction in photosynthetic pigments and damage to the chloroplast ultrastructure, overexpressed ten chloroplastic *PPR*s compared to the *ch42* mutant, which showed a less severe phenotype and had only two overexpressed chloroplastic *PPR*s (Satou *et al.*, 2014).

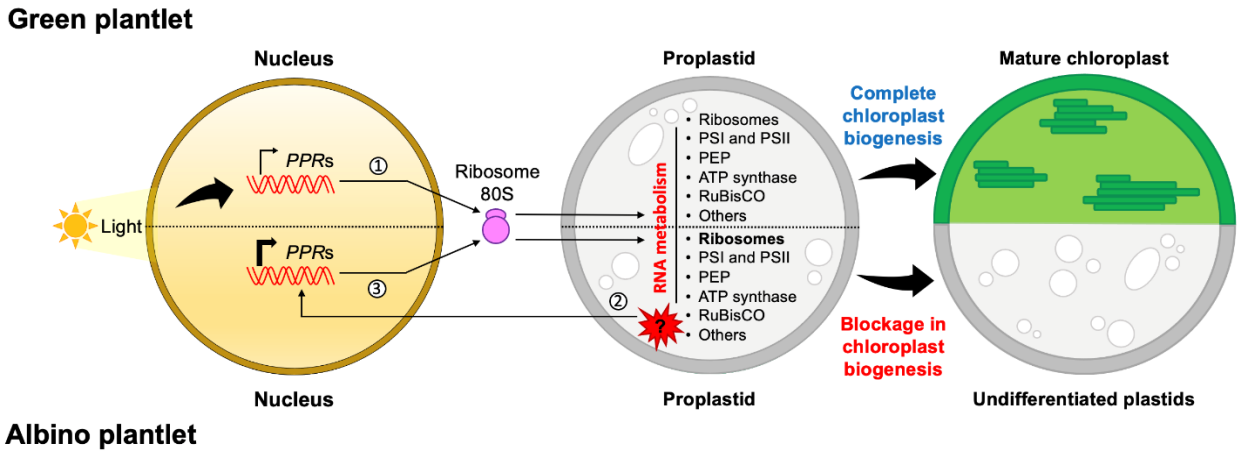


Figure 3.15. Model of retro-anterograde communication proposed for albino plantlets of *A. angustifolia*. In green plantlets, environmental factors, such as light, trigger chloroplast biogenesis through the activation of nuclear genes encoding chloroplastic proteins, among which are the *PPR*s (1). This interorganellar nucleus-plastid communication is part of anterograde signaling. *PPR* transcripts must be translated in cytoplasmic ribosomes and subsequently translocated to plastids, where they participate in RNA processing and in the maturation of the proplastids into mature and functional chloroplasts. On the other hand, in the numerous proplastids of the albino *Agave* plantlets, a possible signal of unknown origin (2) seems to activate a retrograde response (plastid-to-nucleus). We propose that when this signal reaches the nucleus, it activates the overexpression of *PPR*s, which is part of a compensatory retro-anterograde response (3) whose objective is to correct the blockage in plastid biogenesis. It is possible that the overexpression of *PPR*s whose targets are RNAs of genes related to translation and the ribosome could play a central role in this corrective action. However, this compensatory response is unable to remove the proplastid from this immature and undifferentiated stage and, therefore, does not reverse the albino phenotype.

The high expression levels of *PPR* in the albino plantlets (**Figure 3.12** and **3.13**) could reveal a key role of these genes during early biogenesis of the plastid, specifically during interorganellar plastid-nucleus communication**.** Recently, a novel interaction involving these two organelles was described in the virescent *cue8* mutant. In the proplastids of this *Arabidopsis* mutant, an unknown retrograde signal triggered a reduction in the expression of sigma factor genes, photosynthesis-associated nuclear genes (PhANGs), and their regulators, and promoted the expression of NEP**.** This resulted in a corrective anterograde response that maintained the replication of the plastome, suppressed the expression of PEP-dependent genes, and retained the plastid in a state of juvenile development, whose maturation process was slower but successful. This process was called corrective retro-anterograde communication (Loudya *et al.*, 2020). We propose that the high expression of *PPR* could be part of a retro-anterograde compensatory response, very similar to that of the *cue8* mutant. However, this response would fail in the attempt to remove the proplastid from its juvenile developmental stage and reverse the albino phenotype of *Agave* plantlets. At this point, the retrograde signal that activates and maintains *PPR* gene expression in AL is still unknown (**Figure 3.15**). However, the evidence available for the *Agave* albino variant and other albino variants with this phenotype seems to indicate that this retrograde signal could involve the biosynthetic pathways of tetrapyrroles and carotenoids (Moreno *et al.*, 2021; Shimizu y Masuda, 2021).

One aspect that could reveal more information about the importance of the *PPR* genes in this retro-antergrade response focuses on their RNA targets in the plastid. In our search to identify the hypothetical RNA targets of the 12 PPRs using RegExp, targets for nine of them were identified. These targets were related to ribosomes, photosystems, ATP synthase, plastid-encoded RNA polymerase (PEP) and RuBisCO (**Figure 3.14**). The information available on the RNA targets of orthologous proteins in plant models such as *A. thaliana* reveals that some of the targets identified in *A. angustifolia* are conserved (**Table 3.4**) (Zhang *et al.*, 2019; Wagoner *et al.*, 2015; Zhang *et al.*, 2015; Pyo *et al.*, 2013; Zoschke *et al.*, 2013; Wu y Zhang, 2010; Hammani *et al.*, 2009; Beick *et al.*, 2008). For instance, both AaPPR2 and its orthologous protein OTP81/QED1 have as one of their targets the RNA *rps12*, where OTP81/QED1 exerts C-to-U editing (Wagoner *et al.*, 2015). In the case of AaPPR5 and its orthologous protein SEL1/PDM1, the RNA of the *trnK* gene has been identified as a target, where SEL1/PDM1 participates in splicing (Zhang *et al.*, 2015). Finally, AaPPR6 and its orthologous protein SVR7 share as a target the mRNA *atpE*, where SVR7 regulates the transcription and translational activation of dicistron *atpB/E* (Zoschke *et al.*, 2013). A particular case is AaPPR10, to which the orthologous protein

AmPPR5 binds in *Z. mays* and protects the *trnG-UCC* precursor from the action of endonucleases; this activity indirectly impacts the processing and accumulation of rRNAs and the conformation of the plastid ribosome (Beick *et al.*, 2008). This data could be linked to one of the target RNAs of AaPPR10, the RNA of the *rrn23* gene, a finding which reveals a possible site of action that has not been previously identified in other plant models.

Table 3.4. Closest homologues to the nine PPR proteins with identified RNA targets.

Transcript ID	RNA targets	Homologs in *A. thaliana*	PPR subfamily (class)	*E*-value	Identity (%)	RNA targets	References
AaPPR1	*ycf1*	AT4G35130	PLS (DYW)	0	51	-	-
AaPPR2	*rpoC1, rps12*	AT2G29760 (OTP81/QED1)	PLS (DYW)	0	55	·Editing of *accD*, *matK*, *ndnB*, *rpoB* and *rps12.*	(Wagoner *et al.*, 2015)
AaPPR5	*trnK-UUU*	AT4G18520 (SEL1/PDM1)	PLS (PLS)	0	54	·Processing of *rpoA.* ·Editing of *accD.* ·Splicing of *ndhA* and *trnK.*	(Zhang *et al.*, 2015; Pyo *et al.*, 2013; Wu y Zhang, 2010)
AaPPR6	*rps16, petN, atpE*	AT4G16390 (SVR7)	P (SMR)	0	63	·Transcription of ATP synthase subunits *(atpB/E, atpH* and *atpF) and psaJ.* ·Accumulation of ATP synthase subunits *(atpA, atpB, atpE and atpF).* *·Translation rbcL and atpB/E.*	(Zoschke *et al.*, 2013)
AaPPR10	*rpoC1, psbC, petA, rrn23*	GRMZM2G025409 (ZmPPR5)	P	0	70	·Stabilizing unspliced precursor of *trnG-UCC* by inhibiting an endonucleolytic cleveage event. ·Possible association with *rpl16.* ·Indirect association with reduction in rRNA processing.	(Beick *et al.*, 2008)
AaPPR11	*ycf2*	AT4G30825 (BFA2)	P	0	60	·Barrier to prevent the *atpH/F* transcript degradation by exoribonucleases by binding to	(Zhang *et al.*, 2019)

						the consensus of the *atpF/A* intergenic region	
AaPPR13	*atpA, rps14, trnI-GAU, ycf1*	AT5G13770	P	0	50	-	-
AaPPR15	*rbcL, rpl33*	AT2G02980 (OTP85)	PLS (DYW)	0	57	·Editing of *ndhD*	(Hammani *et al.*, 2009)
AaPPR18	*ycf2, psbD, psbC, psaB, ndhG*	AT4G02750	PLS (DYW)	1.00E-117	45	-	-

The identification of more than one RNA target per protein is not a new event. Currently, PPRs that act on more than one target are known, such as OTP81 / QED1, which performs an editing action on five different sites (*accD*, *matK*, *ndhB*, *rpoB* and *rps12 transcripts*) (Wagoner *et al.*, 2015). Another example is the ZmPPR5 protein that has the RNA of the *trnG-UCC* gene as its main ligand, as well as other ligands with which it is weakly associated, such as the RNA of the *rpl16* gene (Beick *et al.*, 2008). A similar case occurs with the ATP4 protein in corn (orthologous of SVR7 in *A. thaliana*), where minor ligands have been reported (Zoschke et al., 2012). The information on the targets of these four *PPRs* (APPR2, APPR5, APPR6 and APPR10) could indicate that the ribosome, and particularly the translation of proteins in the cells of albino leaf tissue, could be compromised and, therefore, anterograde corrective action of PPRs could try to compensate for their defects. However, we cannot discard the possibility that the rest of the RNA targets identified for the PPRs are linked to other protein complexes that act in the plastid and that participate in transcription, photosynthetic metabolism, and ATP synthesis.

Plants with albino phenotypes exhibit a blockage in chloroplast biogenesis, which positions them as unique and novel models for understanding the mechanisms that regulate nucleus-plastid signaling. Here, we showed that the expression of chloroplastic *PPR*s is dependent on the state of differentiation of the plastid, being higher in the early phases of chloroplast biogenesis; that is, when the proplastid phase dominates. Chloroplastic *PPR* genes in the leaf tissue of albino plantlets exhibited an increased expression because of stagnation of plastid development in leaf mesophyll cells. These results reveal the unexpected finding of high expression levels of chloroplastic *PPRs* in albino plants; this expression could be part of a retro-anterograde

communication, where these genes are playing a compensatory function that tries to restore the normal process of development and maturation of proplastids.

CAPÍTULO IV

Proteome of *Agave angustifolia* Haw.: uncovering metabolic alterations, over-accumulation of amino acids, and compensatory pathways in chloroplast-deficient albino plantlets

Artículo publicado:

Andrade-Marcial, M., Ruíz-May, E., Elizalde-Contreras, J.M., Pacheco, N., Herrera-Pool, E., De-la-Peña, C. (2023). Proteome of *Agave angustifolia* Haw.: Uncovering metabolic alterations, over-accumulation of amino acids, and compensatory pathways in chloroplast-deficient albino plantlets. *Plant Physiology and Biochemistry*, 201, 107902.

4.1. INTRODUCTION

The chloroplast is a semiautonomous organelle of cyanobacterial origin present in land plants and algae. It has the ability to convert solar energy into chemical energy, a process known as photosynthesis (Cackett *et al.*, 2022; Kirchhoff, 2019). Since the discovery of the chloroplast, research focused on its physiology has been limited to its role in photosynthetic metabolism. However, in recent years, this reductionist approach has changed, placing the chloroplast as the most important organelle in the plant cell. In addition to photosynthesis, the chloroplast also regulates the response to biotic and abiotic stimuli and plant immunity. More importantly, this organelle is the center of primary and secondary plant metabolism (Littlejohn *et al.*, 2021). For instance, chloroplasts are responsible for the assimilation of carbon, nitrogen and sulfur, as well as the biosynthesis of sugars, fatty acids, lipids, isoprenoids, tetrapyrroles, hormone precursors, vitamins, pigments, alkaloids, and amino acids (AAs) (Cackett *et al.*, 2022; Füssy y Oborník, 2018). AAs are of vital importance because they are the monomeric units that are linked through peptide bonds to form proteins. However, AAs also participate in different stress responses (Hildebrandt, 2018), signaling (Qiu *et al.*, 2020), the energy state of the cell (Izumi y Ishida, 2019), and the biosynthesis of many organic compounds (Oliva *et al.*, 2021; Jez, 2019).

The chloroplast plays a critical role in the metabolism of 17 of the 20 AAs used by cells to synthesize proteins. At least ten AAs are only synthesized in this organelle (Arg, Lys, Thr, Leu, Ile, Val, Trp, Phe, Tyr and His). Seven AAs are synthesized both in the chloroplast and in other parts of the cell (Asp, Cys, Gln, Glu, Gly, Ser and Met) (Rolland *et al.*, 2018). The AA biosynthetic process is dependent on five carbon precursors derived from three pathways of primary

metabolism: phosphoenolpyruvate (PEP), 3-phosphoglycerate and pyruvate from glycolysis, erythrose-4-phosphate and phosphoribosylpyrophosphate (PRPP) derived from the pentose phosphate pathway, and oxaloacetate from the tricarboxylic acid (TCA) cycle (Chen *et al.*, 2018b). Under such circumstances, it is clear that homeostasis in AA metabolism in plants is strongly linked to plastids.

In plants with reduced photosynthetic pigments due to a blockage in chloroplast development and maturation, AA metabolism and concentration are drastically affected. The variety of albino and chlorotic genotypes in the tea plant (*Camellia sinensis*) has revealed some clues about the impact of the absence of functional chloroplasts on AA metabolism. Interestingly, most reports agree that albino and chlorotic cultivars have a higher AA concentration than their green counterparts. The causes of this increase are still a matter of discussion. However, some hypotheses have been proposed to explain this increase in AA content. These include an imbalance in carbon and nitrogen metabolism, activation of AA biosynthetic and degradation pathways, protein breakdown mechanisms such as the ubiquitin-proteasome system and autophagy, and low N consumption and stagnation of the urea cycle (Yamashita *et al.*, 2021; Li *et al.*, 2019b; Lu *et al.*, 2019; Li *et al.*, 2018a). In model plants such as *Arabidopsis thaliana*, AA metabolism and concentration in mutants with alterations in their pigmentation has been superficially explored. For example, *apg2*, *apg3*, *cla1* and *ch42* mutants with an albino/pale-green phenotype showed an increase in the concentration of Asn, Gln, Glu, Arg and His. This increase has been related to a strong activation of nitrogen assimilation and protein degradation (Satou *et al.*, 2014). A proteomic analysis investigating three *apg* mutants revealed a significant increase in the abundance of proteins associated with protein degradation, possibly due to cellular stress. Additionally, these mutants had higher levels of chloroplast proteins involved in AA metabolism and biosynthesis processes (Motohashi *et al.*, 2012). These findings suggest the activation of AA metabolic pathways in albino/pale-green heterotrophic models lacking mature chloroplasts.

The activation of AA metabolic pathways and the altered AA concentration in albino/chlorotic plants seems to be influenced by tissue development and plastid type. Studies have shown that different plastid types exhibit changes in AA metabolism and concentration compared to chloroplasts. In tobacco, the transition from etioplast to chloroplast differentiation leads to decreased AA concentration due to increased light exposure time and the synthesis of photosynthetic protein complexes, such as RuBisCO, photosystems and light-harvesting complexes (Armarego-Marriott *et al.*, 2019). In maize leaves, a development gradient from the base to the tip revealed high concentrations of certain AAs, including Gly, Glu, Ser, Lys, Met, Val,

Ile, and Asp at the leaf base. This tissue is considered photosynthetically immature, with active nitrogen metabolism (Wang *et al.*, 2014). Proteomic studies in maize reported high activity of biosynthetic enzymes for aromatic AAs, Met, Ser, and BCAAs at the leaf base, where proplastids are abundant (Majeran *et al.*, 2010). Subsequent studies indicated a two-fold reduction in AA metabolism in chloroplasts compared to proplastids (Majeran *et al.*, 2011). Similar increases in AA metabolic enzyme activity have been observed in other plastid types such as chromoplasts (Rödiger *et al.*, 2021). In our laboratory, variegated and albino somaclonal variants of *A. angustifolia* were found to have partial or total loss of chloroplasts, respectively. These variants harbor immature, uncolored and non-photosynthetic plastids similar to proplastids in the albino regions (Hernández-Castellano *et al.*, 2020). This could indicate that in phenotypes lacking chlorophyll and chloroplasts, pathways associated with heterotrophic plastid metabolism are highly active.

In this research, we conducted a thorough assessment of AA metabolism in albino and variegated somaclonal variants of *A. angustifolia*. This involved a comprehensive quantitative proteomic analysis to identify enzymes involved in AA metabolic pathways. Additionally, we performed expression analysis to gain insights into the intricate relationship between transcriptional and translational mechanisms, providing a deeper understanding of the complexity involved. Additionally, we employed bioinformatic analyses to gain insights into the regulatory mechanisms underlying the observed alterations in AA metabolism. Our findings indicate that changes in AA concentration and accumulation of key enzymes involved in AA metabolism are part of a broader metabolic and stress response reprogramming. These adaptations enable the albino and variegated plantlets to better cope with environmental conditions.

4.2. MATERIALS AND METHODS

4.2.1. Plant materials

Three somaclonal lines with the same genetic background but different phenotypes [green (G), variegated (V) and albino (A)] (Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016), were used (**Figure 4.1A**). Plantlets from each phenotype were micropropagated in Magenta boxes containing 50 ml of Murashige and Skoog (MS) medium (Murashige y Skoog, 1962) modified in its nitrogen content with 18 mM KNO_3 and 5 mM NH_4NO_3. MS medium was supplemented with 6-benzyladenine (BA; 22.2 µM) and 2,4-dichlorophenoxyacetic acid (2,4-D; 0.11 µM), and solidified with 0.175 % (w/v) agar and 0.175 % (w/v) gel-rite (Robert *et al.*, 2006). The plantlets

were cultured in a growth chamber at 27°C for 12 hours with a photoperiod of 12 hours (40 $\mu mol/m^{-2}/s^{-1}$). Plantlets of ~2.5 cm tall were collected, frozen with liquid nitrogen, and stored at -80ºC.

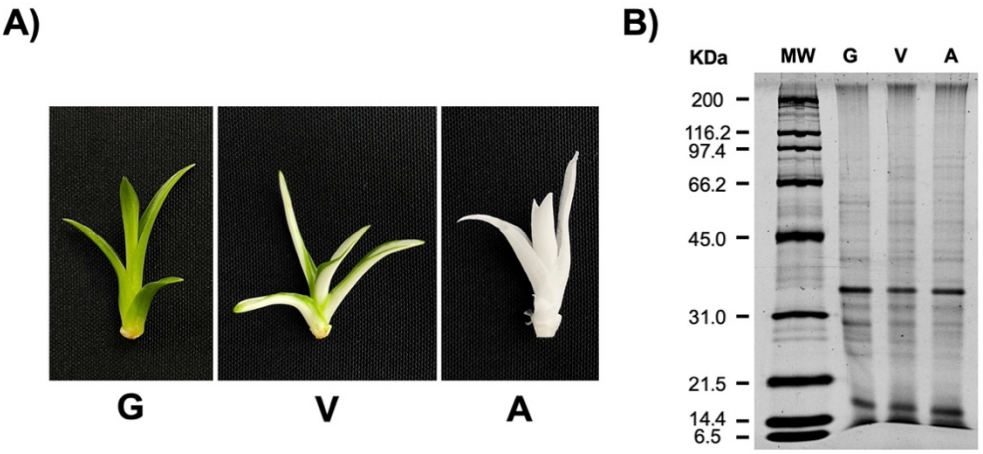


Figure 4.1. Quality evaluation of total protein extracts from *A. angustifolia* G, V and A plantlets. A) Photographs of the G, V and A plantlets. **B)** Protein profiles of the plantlets viewed by 1D-SDS-PAGE and stained with SYPRO Ruby®. Each sample lane was loaded with 10 µg of total protein. M= molecular mass marker, KDa= kilodalton, G: green plantlet, V: variegated plantlet, A: albino plantlet.

4.2.2. Total protein extraction and quality evaluation

Two-and-a-half grams of each phenotype (G, V and A) were homogenized with liquid nitrogen until obtaining a fine powder with the help of a mortar and pestle. Each phenotype had three biological replicates. Protein extraction was performed according to the protocol of Juarez-Escobar *et al.* (2021). The supernatant containing the total proteins was recovered. The sample was cleaned with Protein Desalting Spin Columns (Thermo-Fisher Scientific) in order to remove salts and small molecules. The crude protein extracts were stored at -80ºC until analysis. Total protein was determined using Pierce™ BCA Protein Assay Kit (Thermo-Fisher Scientific) according to the manufacturer's instructions. One hundred micrograms of total protein from each phenotype were used to visualize the protein profiles by 1D-SDS-PAGE (with 10% SDS) and stained with SYPRO™ Ruby (Invitrogen™). SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad Laboratories) was employed as the protein molecular weight marker.

4.2.3. Protein digestion and TMT labeling

Total protein was reduced, alkylated and digested according to Monribot-Villanueva *et al.* (2022). Tandem Mass Tag (TMT) 10-plex reagents (Thermo-Fisher Scientific) were utilized as

isobaric labels for the comparative quantification of proteins between samples. The labels used were as follows: 128N, 128C, and 129N for plantlets with the A phenotype; 130C, 129C and 130N for plantlets with the V phenotype; and 126, 127C, and 127N for samples with the G phenotype.

4.2.4. Nano LC-MS/MS analysis, synchronous precursor selection (SPS)-MS^3 and data analysis

The labeled peptides were dissolved in solvent A (0.1% formic acid in Mass Spectrometry grade water). 20 µl of the solution was injected into a nanoviper C18 trap column (3 µm, 75 µm × 2 cm, Dionex) at a flow rate of 3 µl min^{-1} in an UltiMate 3000 RSLC system (Dionex, Sunnyvale). The peptides were then separated on an EASY spray C18 RSLC column (2 µm, 75 µm × 25 cm), using a 100 min gradient with a flow rate of 300 nl min^{-1}. The gradient used was previously described by Monribot-Villanueva *et al.* (2022). The nanoLC platform was coupled to a Thermo-Fisher Scientific Orbitrap Fusion Tribid mass spectrometer with a "EASY Spray" nano ion source (Thermo-Fisher Scientific). The mass spectrometer was operated under the conditions provided by Juarez-Escobar *et al.* (2021). Full MS scans were performed with the Orbitrap analyzer under the settings and fragmentation parameters outlined in Juarez-Escobar *et al.* (2021). MS^3 spectra were collected using synchronous precursor selection (SPS) of 10 isolation notches, as described in McAlister *et al.* (2014). MS^3 precursors were fragmented at the conditions proposed by Juarez-Escobar *et al.* (2021).

The generated mass spectra were analyzed with Proteome Discoverer 2.4 (PD, Thermo Fisher Scientific) utilizing the AMANDA (Dorfer *et al.*, 2014) and SEQUEST HT (Eng *et al.*, 1994) search engines with scores of ≥25 and ≥2, respectively. The search was conducted against the translated transcriptome of *A. angustifolia* Haw. (unpublished data). This research adhered to the search parameters and static and dynamic adjustments established by Juarez-Escobar *et al.* (2021). Tolerances of ±10 ppm and ± 0.6 Da were applied to the SPS-MS^3 method, in which identification was conducted at a lower resolution in the linear ion trap. Using the Percolator algorithm, the resultant peptide hits were filtered for a maximum of 1% false discovery rate (FDR) (Käll *et al.*, 2007). Only proteins with at least two distinct peptides were chosen for additional analysis. The reporter ion quantification for tandem mass tags (TMT) was obtained using the Proteome Discoverer software template at the MS^3 level with mass tolerances of ±10 ppm for the most confident centroid and a precursor co-isolation filter of 75%. The proteomic method used isobaric labeling with the TMT and the SPS-MS^3, where MS2 (CID) identified peptides and MS^3 (HCD) quantified TMT reporter ions. The differentially accumulated proteins (DAPs) were selected

using a *p*-value < 0.05 and a Log2 fold change (LFC) > 0.58 and < -0.58 for the up- and down-accumulated proteins, respectively (raw data are available via ProteomeXchange with identifier PXD039512). The number of DAPs in each of the three phenotypic comparisons was represented in Venn diagrams using the InteractiVenn tool (Heberle *et al.*, 2015). To show DAP abundance in each phenotype, ComplexHeatmap software (vers. 2.4.3) was used (Gu *et al.*, 2016).

4.2.5. Annotation methods

Nearest orthologs of the identified proteins were searched by running a local BLASTp (vers. 2.10.1+) (Altschul *et al.*, 1990) against the TAIR11 proteome database of *A. thaliana* (downloaded on February 2023) and the RefSeq protein database (directed to Embryophyta [taxid:3193]) (NCBI, February 2023). The default settings were used for the local BLASTp search. Sequence pairwise alignments with an identity percentage greater than 50% were chosen. Only the best match was chosen and reported for sequence alignments with multiple hits. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) (2021 update) (Sherman *et al.*, 2022) was used to perform KEGG pathway enrichment analysis. DAVID used the *A. thaliana* orthologs proteins to perform this analysis. The strategy followed in this proteomic analysis is summarized in **Figure 4.2**.

4.2.6. Gene expression analysis by qRT-PCR

To identify changes in the expression of genes encoding enzymes involved in AA biosynthesis, relative quantification was performed by quantitative reverse transcription PCR (qRT-PCR). Total RNA was extracted from G, V and A plantlets using TRI Reagent® (Sigma-Aldrich) according to the manufacturer's instructions. cDNA was synthesized using RevertAid First Strand cDNA Synthesis Kit (Thermo-Fisher Scientific). *BCAT*, *CM*, *MS*, *TS* and *CGS* genes linked with Trp, Phe, Tyr, Val, Leu, Ile and Met metabolism were selected to validate their expression. *Actin* was used as housekeeping gene. The list of oligonucleotides synthesized for this study are presented in **Table 4.1**. qRT-PCR reactions were conducted using a StepOnePlus Real-Time PCR System and Maxima SYBR Green qPCR Master Mix (2x) (Applied Biosystems). Each qRT-PCR reaction was prepared according to the supplier's instructions. The thermocycler program consisted of initial denaturation at 95 °C for 10 min, followed by 40 cycles each with 15 s denaturation at 95 °C, 30 s annealing at 58 °C and 30 s extension at 72 °C. Three biological replicates were used for each phenotype of interest. The relative expression levels of the genes were calculated using the 2–ΔΔCT method (Livak y Schmittgen, 2001).

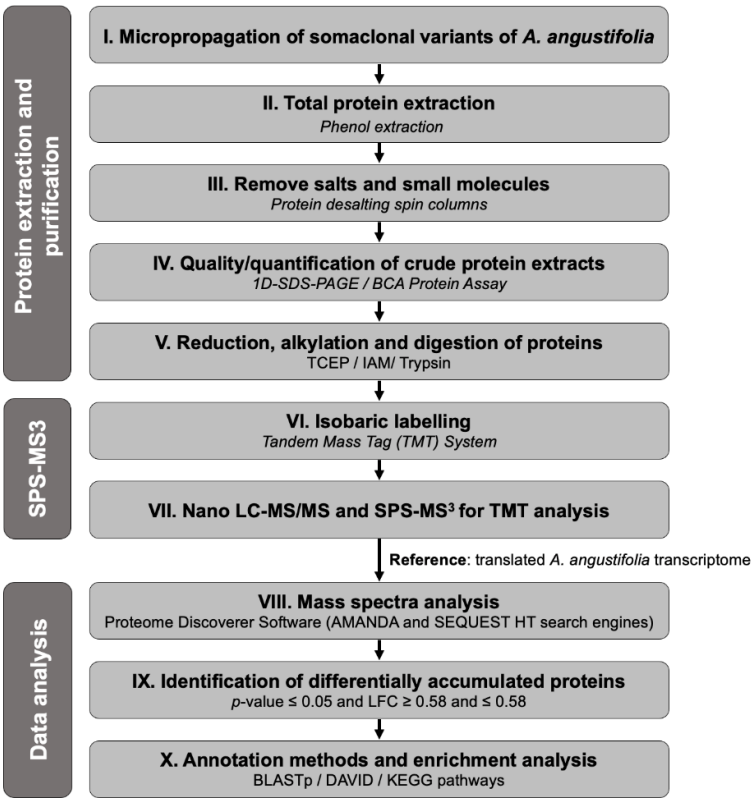


Figure 4.2. Representation of the workflow of this proteomic analysis.

Table 4.1. Oligonucleotides designed from AAs biosynthesis genes for expression analysis by qRT-PCR.

Gene	Oligonucleotide name	Sequence
CGS *(cystathionine gamma-synthetase)*	CGSF	5'-GATTACGGCCACTGTCATTG-3'
	CGSR	5'-ACCAGAGCTCCTTTGCTATG-3'
BCAT *(branched-chain aminotransferase)*	BCATF	5'-CTAAGGCCAATGGTTACTCTG-3'
	BCATR	5'-CCAGGCATTCTTCAACCTCG-3'
MS *(methionine synthase)*	MSF	5'-GGTCAAGTTGCAGGAAGAGC-3'
	MSR	5'-GCTTAACGCACCGTGAACCA-3'
CM *(chorismate mutase)*	CMF	5'-CTCGTTACCTCTTCTTCTTTCC-3'
	CMR	5'-GAATTCGCTGGAGATCTTGTC-3'
TS *(tryptophan synthase)*	TSF	5'-GTGTTCATGGACTTGTGGTG-3'
	TSR	5'-CGAACTTACAGATGCACGAG-3'
Actin	ACTF	5'-GTCGTACAACTGGTATTGTGCTGGA-3'
	ACTR	5'-GTAACCACGCTCAGTCAGGATCTTC-3'

4.2.7. Amino acid extraction

Frozen plant samples were lyophilized to extract free AA using the method described by Herrera-Pool *et al.* (2021). Fifty mg of each sample was weighed and placed into 50-ml conical tubes. 10 ml of borate buffer at 0.05 M (pH 8.05) was added. Samples were treated in an ultrasonic bath extraction (Bransonic® 3510R-MT, Branson Ultrasonics, Co., Danburi, CT, USA) for 60 min. The samples were then centrifuged for 15 min at 4 °C at 4,000 rpm and filtered through 0.20 µm acrodisc filter membranes (Phenex, NY).

4.2.7.1. Derivatization of standards and samples

To prepare the standards, the Amino Acid Standard H (Thermo-Fisher Scientific) that contains 2.5 mM of each AA (His, Ser, Arg, Gly, Asp, Glu, Thr, Ala, Pro, Lys, Tyr, Met, Val, Ile, Leu, Phe) except for Cys (1.5 mM) was used. A volume of 200 µl standard solution was diluted in 800 µl of Milli-Q water. The diluted solution was used to prepare a calibration curve with a concentration from 12.5 to 500 pmol/µl of each AA (12.5, 25, 62.5, 125,187.5, 250 and 500 pmol/µl), except for Cys (6.25, 12.5, 31.25, 125, 62.5, 93.75, 125 and 250 pmol/µl). All the calibration standards and samples were derivatized using an AccQ·Tag™ Ultra derivatization kit (Waters Corporation, Milford, MA, USA) following the protocol established in UPLC® Amino Acid Analysis Solution System Guide.

4.2.7.2. UPLC-PDA Analysis

The chromatographic profiles were acquired using a ACQUITY UPLC H-Class System (Waters Corporation, Milford, MA, USA) equipped with a quaternary pump (UPQSM), an autosampler injector (UPPDALTC), and a PDA λ photodiode array detector (UPPDALTC). Chromatographic separation was performed using a ACQUITY UPLC BEH C18 column (1.7 µm, 100 x 2.1 mm i.d.) (Waters Corporation, Milford, MA, USA) at 60°C and two solvents as mobile phase: AccQ·Tag™ Ultra Eluent A diluted 1:10 (v/v) with Milli-Q water (mobile phase A) and AccQ·Tag™ Ultra Eluent B (mobile phase B). The elution gradient with respect to the percentage of mobile phase A was as follows: from 0.00 to 0.54 min at 99.9% (isocratic); from 0.54 to 5.74 min in a linear gradient from 99.9% to 90.9%; from 5.74 to 7.74 min in a linear gradient from 90.9% to 78.8%; from 7.74 to 8.04 min in a linear gradient from 78.8% to 40.4%; from 8.04 to 8.05 min in a linear gradient from 40.4 to 10.0%; from 8.05 to 8.64 min at 10.00% (isocratic); from 8.64 to 8.73 min in a linear gradient from 10.00% to 99.9%; from 8.73 to 9.50 min at 99.9% (isocratic). The area integration of the chromatographic peaks (µV*sec) for each detected compound was carried out by taking the analytical signal at 260 nm.

4.2.8. Statistical analysis

The relative expression data generated by qRT-PCR and AA concentration were subjected to one-way analysis of variance (ANOVA) ($p < 0.05$) and a Tukey *post-hoc* test to determine statistical differences between phenotypes. The square root function was used when the assumption of data normality was not met. These analyses were performed using RStudio software (vers. 1.4.1106) (RStudio, 2021). Graphs were designed using Graph Pad Prism (v9.2.0) (Graph Pad software, www.graphpad.com) and *ggplot2* (v3.3.5).

4.3. RESULTS

4.3.1. Identification of total proteins in the somaclonal variants of *A. angustifolia*

A total of 2,442 distinct proteins were found after a proteome analysis was conducted on the three somaclonal variants (G, V and A) of the *A. angustifolia* Haw. species. Protein profiles analyzed by 1D-SDS-PAGE showed similar patterns among the phenotypes. However, the differences in the intensity of the bands could indicate changes in the abundance of specific proteins (**Figure 4.1B**). Pearson's correlation analysis revealed a high similarity among the replicates of each phenotype, with correlation coefficients equal to or greater than 0.99 (**Figure 4.3**). Multidimensional scaling (MDS) confirmed this similarity by grouping the replicates in the plot and showing differences between the proteomes of the phenotypes (**Figure 4.4**).

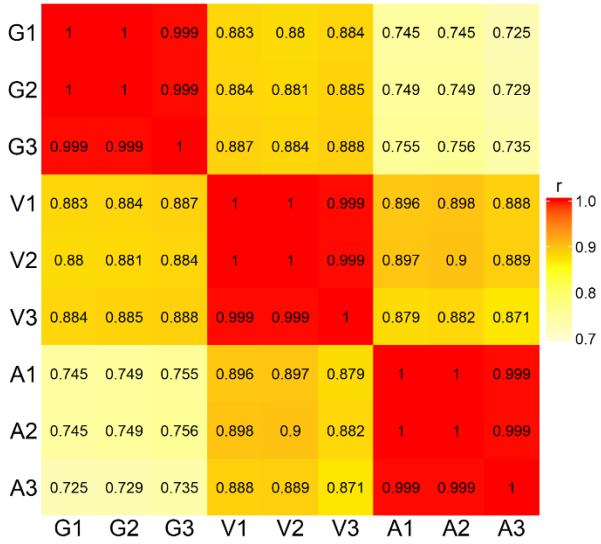


Figure 4.3. Heatmap showing the values of the Pearson correlation coefficient for the replicates of the studied phenotypes. The color scale was established based on the coefficient of determination (r). G:

green plantlet, V: variegated plantlet, A: albino plantlet. The numbers indicated after each phenotype (e.g., A1) show the replicate number.

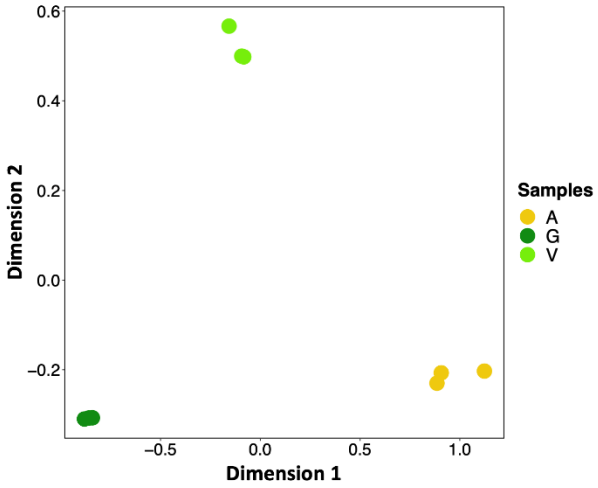


Figure 4.4. Multidimensional scaling (MDS) plot of the replicates of G, V and A phenotypes of *A. angustifolia*. This MDS shows that replicates cluster according to phenotype. G: green plantlet, V: variegated plantlet, A: albino plantlet.

4.3.2. Proteins differentially accumulated in the somaclonal variants

The three phenotypes of *Agave* plantlets were compared to identify DAPs (G vs V, G vs A and V vs A). A total of 1,028 DAPs were identified in the three comparisons (**Figure 4.5A** and **B**). The G vs A comparison showed the highest number of DAPs, with a total of 748 (443 over-accumulated and 305 down-accumulated). The V vs A comparison showed 565 DAPs (351 over-accumulated and 214 down-accumulated). Finally, the G vs V comparison showed the lowest number of DAPs, with a total of 527 (302 over-accumulated and 225 down-accumulated) (**Figure 4.5B** and **D**). The G vs A comparison also had the highest number of unique DAPs with 168, followed by V vs A with 98 and G vs V with 82 (**Figure 4.5C**).

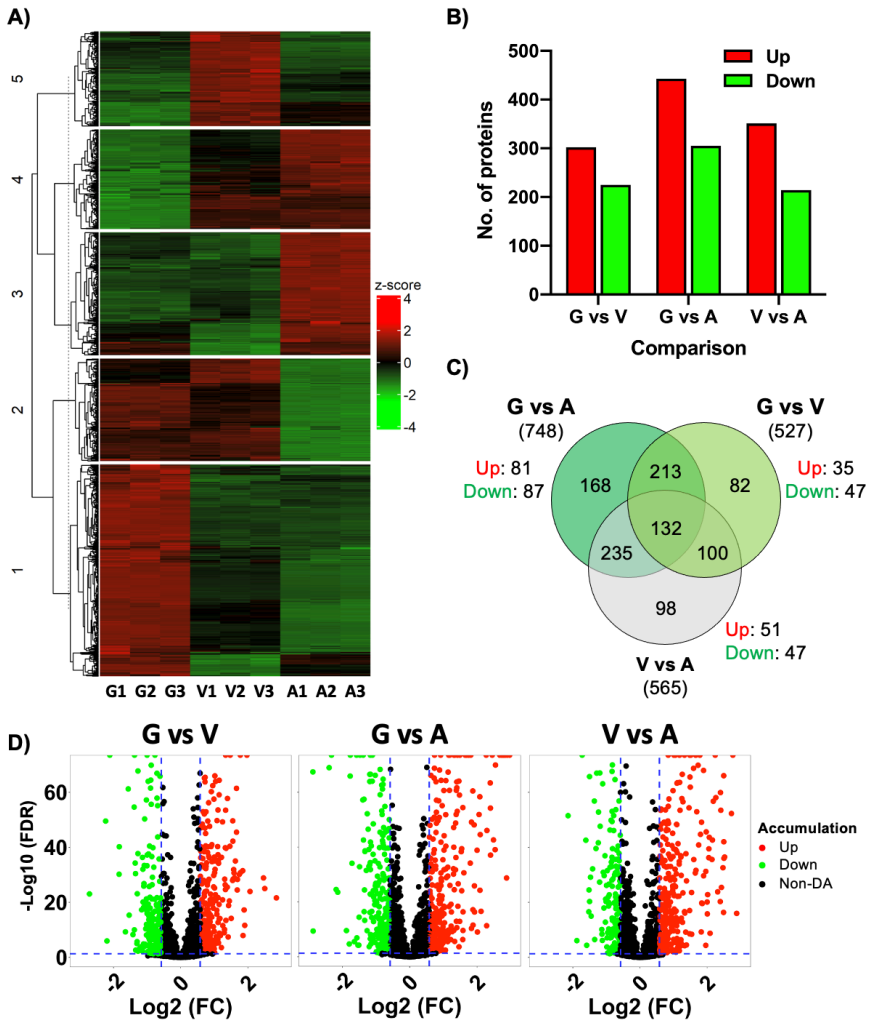


Figure 4.5. Differentially accumulated proteins (DAPs) identified in the proteome of the three somaclonal variants of *A. angustifolia*. A) Heatmap depicting the accumulation profiles of 1,028 DAPs in *A. angustifolia* G, V, and A plantlets. The bottom of the figure depicts the three phenotypes and replicates. Clustering was used to categorize the DAPs according to their accumulation levels. The created clusters are presented on the left with bars and numbers (1–5). **B)** The number of DAPs in each of the examined comparisons is displayed (G vs V, G vs A and V vs A). In each comparison, the initial phenotype serves as a reference to determine whether the proteins are accumulating more (represented by red bars) or less (marked by green bars). **C)** Venn diagram showing the DAPs identified in each comparison. The number of DAPs identified in each comparison is presented in parentheses. The number of specific DAPs in each comparison, both the up- and down-accumulated proteins, are indicated on the outside of each circle. **D)** Volcano plots showing the distribution of DAPs in the three comparisons. The x-axis shows the LFC calculated for each of the three comparisons and the -Log10 FDR on the y-axis. Green dots indicate proteins down-accumulated and red dots indicate proteins up-accumulated in each comparison. G: green

plantlet, V: variegated plantlet, A: albino plantlet, Up: up-accumulated proteins, Down: down- accumulated proteins.

4.3.3. Proteins involved in AA metabolism are strongly enriched in the proteomes of somaclonal variants

To determine the main metabolic pathways enriched using the 1,028 DAPs identified in the three comparisons between phenotypes, DAPs were mapped to the KEGG Pathway Database. The 20 most enriched KEGG pathways (*p*-value < 0.05) identified are presented in (**Figure 4.6**). These pathways include photosynthesis, phenylpropanoid biosynthesis, carbon fixation, glutathione metabolism, pentose phosphate pathway, fructose and mannose metabolism. However, among them, pathways related to AA metabolism were also found, such as Ala, Asp and Glu metabolism, Arg biosynthesis, Cys and Met metabolism, and particularly the AA biosynthesis.

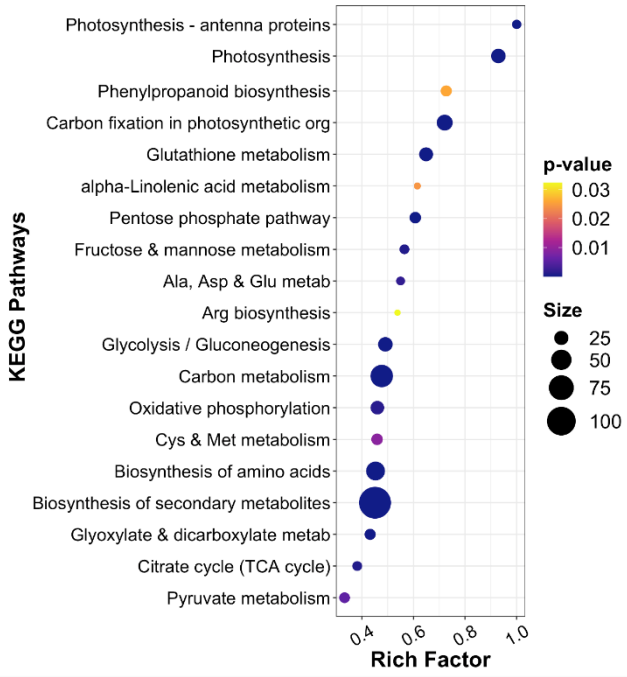


Figure 4.6. Bubble map of KEGG pathway enrichment analysis of total differentially accumulated proteins identified in somaclonal variants G, V and A of *A. angustifolia*. The y-axis represents the pathway name, while the x-axis represents the enriched factor for each pathway. The size of the bubble represents the number of DAPs identified along each pathway. A color gradient illustrates the *p*-value.

4.3.4. AA biosynthesis plays a central role in somaclonal variants V and A of *A. angustifolia*

To analyze the changes in AA metabolic pathways in the three phenotypes (G, V and A), particularly those related to their biosynthesis, the enzymes involved in these pathways and detected in the proteomes of each phenotype were identified. A total of 82 proteins participating in AA biosynthesis were identified in this study. Additionally, the concentration of each of the 20 AAs was quantified (**Figure 4.7, 4.8, 4.9, 4.10 and 4.11**).

4.3.4.1. AAs derived from 3-phosphoglycerate: Ser, Gly and Cys

The AAs Ser, Gly and Cys, which all have 3-phosphoglycerate as a precursor, were found in high concentrations in the V phenotype compared to the other phenotypes. Cys was severely reduced in the A phenotype (**Figure 4.7B**). Two of the three enzymes required for Ser biosynthesis, *Aa*PGDH and *Aa*PSAT, were identified in the proteomic analysis. Four putative chloroplast isoforms of *Aa*PGDH, named *Aa*PGDH1-4, were identified. Only *Aa*PGDH3 was over-accumulated in the A phenotype compared to V. Both enzymes are part of the pathway known as the phosphorylated pathway, which takes place in the chloroplast. For the interconversion of Gly to Ser, there are two alternative pathways. The first of these was the glycolate pathway, which takes place in the mitochondria and is associated with photorespiration. The enzyme *Aa*SHMT is responsible for catalyzing this reaction. (Toujani *et al.*, 2013). In this study, of the five isoforms detected for *Aa*SHMT (designated *Aa*SHMT1-5), *Aa*SHMT1, 3, and 5 were up-accumulated in the A phenotype. The second pathway is catalyzed by *Aa*SGAT, which was up-accumulated in G and V phenotypes in comparison to A. Finally, the enzymes that regulate Cys biosynthesis (*Aa*ASAT and *Aa*CS) were not detected or did not show differences in their accumulation (**Figure 4.7A**).

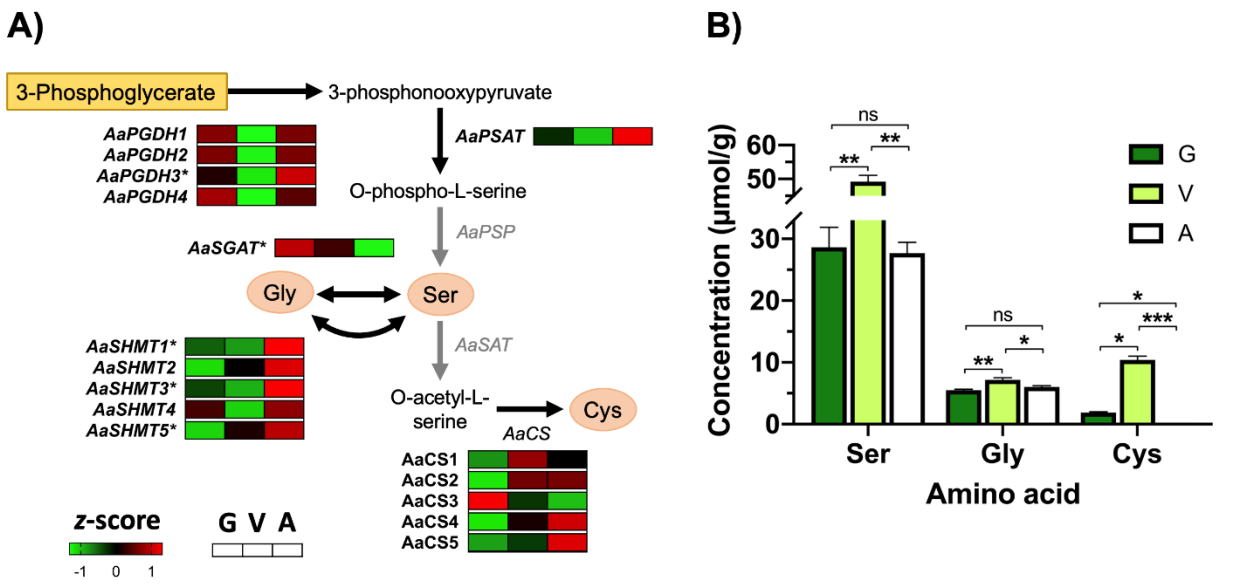


Figure 4.7. **Ser, Gly and Cys biosynthesis pathways derived from 3-phosphoglycerate**. **A)** Accumulation of the enzymes catalyzing Ser, Gly and Cys biosynthesis pathways identified in the proteomic analysis of the three somaclonal variants of *A. angustifolia*. Yellow rectangle indicates the metabolic precursor. Enzymes detected in the proteomic analysis are indicated by black arrows. Enzymes not detected are indicated in gray. Numbers in the name of each enzyme indicate the number of each isoform found. Asterisks indicate differentially accumulated enzyme isoforms. Orange ovals indicate AAs synthesized in both the chloroplast and other parts of the cell. Heatmaps attached to each enzyme show its accumulation level in all three phenotypes. **B)** Ser, Gly and Cys concentration in the three somaclonal variants. The values of the mean and the standard error are represented in each graph. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns: non-significant differences. G: green plantlet, V: variegated plantlet, A: albino plantlet, Ser: serine, Gly: glycine, Cys: cysteine, Aa: *Agave angustifolia*, PGDH: phosphoglycerate dehydrogenase (EC:1.1.1.95), PSAT: phosphoserine transaminase (EC:2.6.1.52), PSP: phosphoserine phosphatase (EC:3.1.3.3), SHMT: L-serine hydroxymethyltransferase (EC:2.1.2.1), SGAT: serine:glyoxylate aminotransferase (EC:2.6.1.45), SAT: serine acetyltransferase (EC:2.3.1.30), CS: cysteine synthase (EC:2.5.1.47).

4.3.4.2. AAs derived from D-erythrose 4-phosphate and phosphoenolpyruvate: Tyr, Phe and Trp

The aromatic AAs (Tyr, Phe and Trp), which are synthesized from D-erythrose 4-phosphate and phosphoenolpyruvate (PEP), were found to be in high concentrations in the V phenotype, followed by A and finally G (**Figure 4.8B**). Two enzymes of this biosynthetic pathway, *Aa*CM and *Aa*TSB, were identified as DAPs in this study. The enzyme *Aa*CM, which regulates

the synthesis of prephenate from chorismate prior to the formation of Tyr and Phe, was up-accumulated in the A phenotype compared to the other phenotypes. In Trp biosynthesis, only the enzyme *Aa*TSB was found, which catalyzes the synthesis of this AA from indole. This enzyme was highly accumulated in the G phenotype compared to the other phenotypes (**Figure 4.8A**).

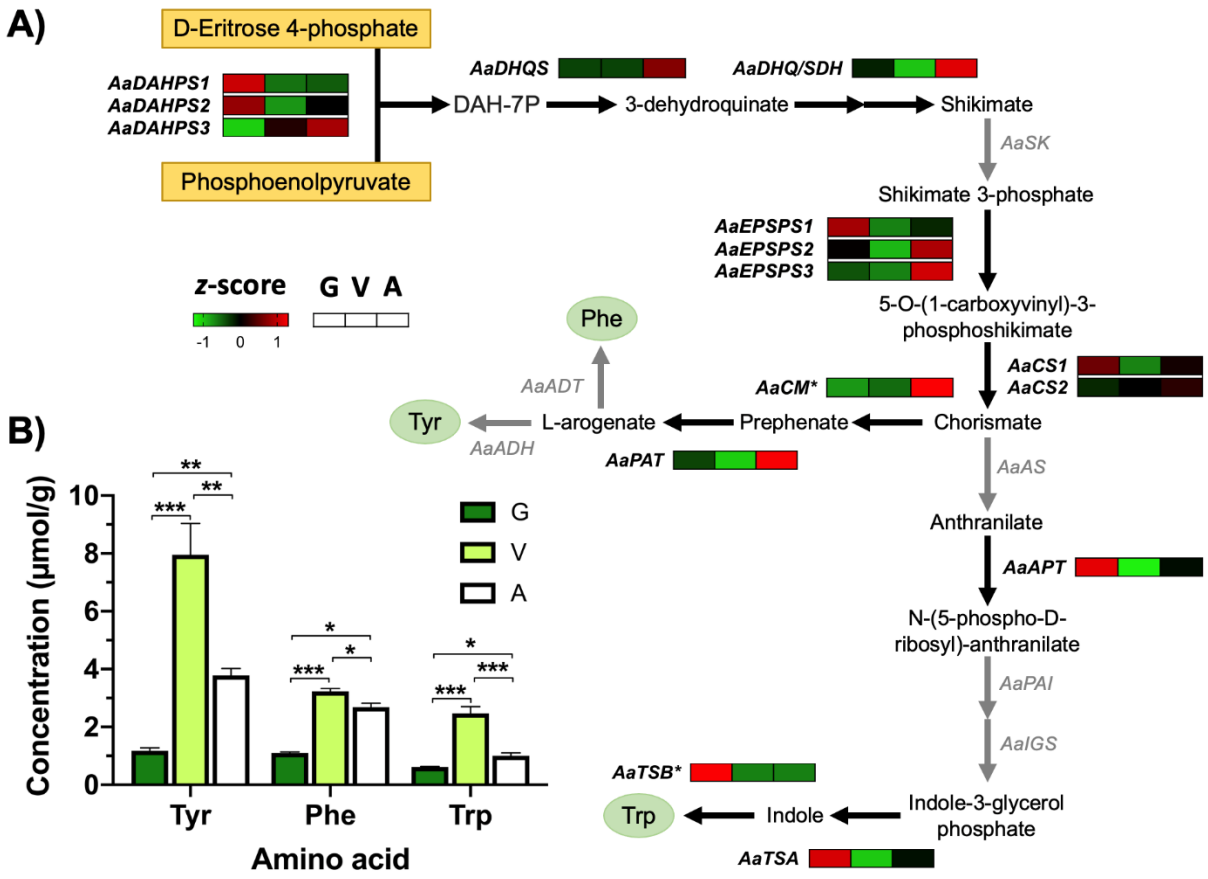


Figure 4.8. Tyr, Phe and Trp biosynthesis pathways derived from D-erythrose 4-phosphate and PEP. **A)** Accumulation of the enzymes catalyzing Tyr, Phe and Trp biosynthesis pathways identified in the proteomic analysis of the three somaclonal variants of *A. angustifolia*. Yellow rectangles indicate the metabolic precursors. Enzymes detected in the proteomic analysis are indicated by black arrows. Enzymes not detected are indicated in gray. Numbers in the name of each enzyme indicate the number of each isoform found. Asterisks indicate differentially accumulated enzyme isoforms. Green ovals indicate AAs only synthesized in the plastids. Heatmaps attached to each enzyme show its accumulation level in all three phenotypes. **B)** Tyr, Phe and Trp concentration in the three somaclonal variants. The values of the mean and the standard error are represented in each graph. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. G: green plantlet, V: variegated plantlet, A: albino plantlet, Phe: phenylalanine, Tyr: tyrosine, Trp: tryptophan, Aa: *Agave angustifolia*, DAHPS: 3-deoxy-7-phosphoheptulonate synthase (EC:2.5.1.54), DHQS: 3-dehydroquinate synthase (EC:4.2.3.4), DHQ/SDH: bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase

(EC:1.1.1.25 and EC:4.2.1.10), SK: shikimate kinase (EC:2.7.1.71), EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase (EC:2.5.1.19), CS: chorismate synthase (EC:4.2.3.5), CM: chorismate mutase (EC:5.4.99.5), PAT: prephenate aminotransferase (EC:2.6.1.78 and EC:2.6.1.79), ADT: arogenate dehydratase (EC:4.2.1.91), ADH: arogenate dehydrogenase (EC:1.3.1.43), AS: anthranilate synthase (EC:4.1.3.27), APT: anthranilate phosphoribosyltransferase (EC:2.4.2.18), PAI: phosphoribosylanthranilate isomerase (EC:5.3.1.24), IGPS: indole-3-glycerol-phosphate synthase (EC:4.1.1.48), TSA: tryptophan synthase alpha subunit (EC:4.2.1.20), TSB: tryptophan synthase beta subunit (EC:4.2.1.20), DAH-7P: 3-Deoxy-D-arabinoheptulosonate 7-phosphate.

4.3.4.3. AAs derived from pyruvate: Val, Leu and Ala

Val and Leu, which are derived from pyruvate, were detected in high concentrations in V and A phenotypes. On the other hand, Ala was over-accumulated in the G phenotype (**Figure 4.9B**). Proteomic analysis identified the presence of the *Aa*ALT enzyme, which is involved in Ala biosynthesis from pyruvate. This enzyme was over-accumulated in the A phenotype. Of the enzymes that regulate Val and Leu biosynthesis, only two showed differences in their accumulation: *Aa*ALS1 and *Aa*BCAT. The *Aa*ALS1 isoform, which regulates the first step of Val and Leu biosynthesis, was more abundant in the G phenotype. This enzyme also participates in the second step of the Ile biosynthesis pathway. On the other hand, *Aa*BCAT, which catalyzes the individual transamination of 2-oxo-acids to form the three branched-chain amino acids (BCAAs), was more abundant in the A phenotype compared to G. Finally, of the enzymes required for Leu biosynthesis, only a DAP corresponding to *Aa*IPMS1 was identified. This enzyme, which is responsible for regulating the condensation reaction of acetyl-CoA with 2-oxoisovalerate to form 2-isopropylmalate, was up-accumulated in the G phenotype respect to A (**Figure 4.9A**).

The AA His, which derives from ribose-5-phosphate, did not show differences in its concentration between phenotypes (**Figure 4.9B**). In the proteomic analysis, only the *Aa*HIN2 enzyme was identified. However, this did not show changes in its accumulation (**Figure 4.9C**).

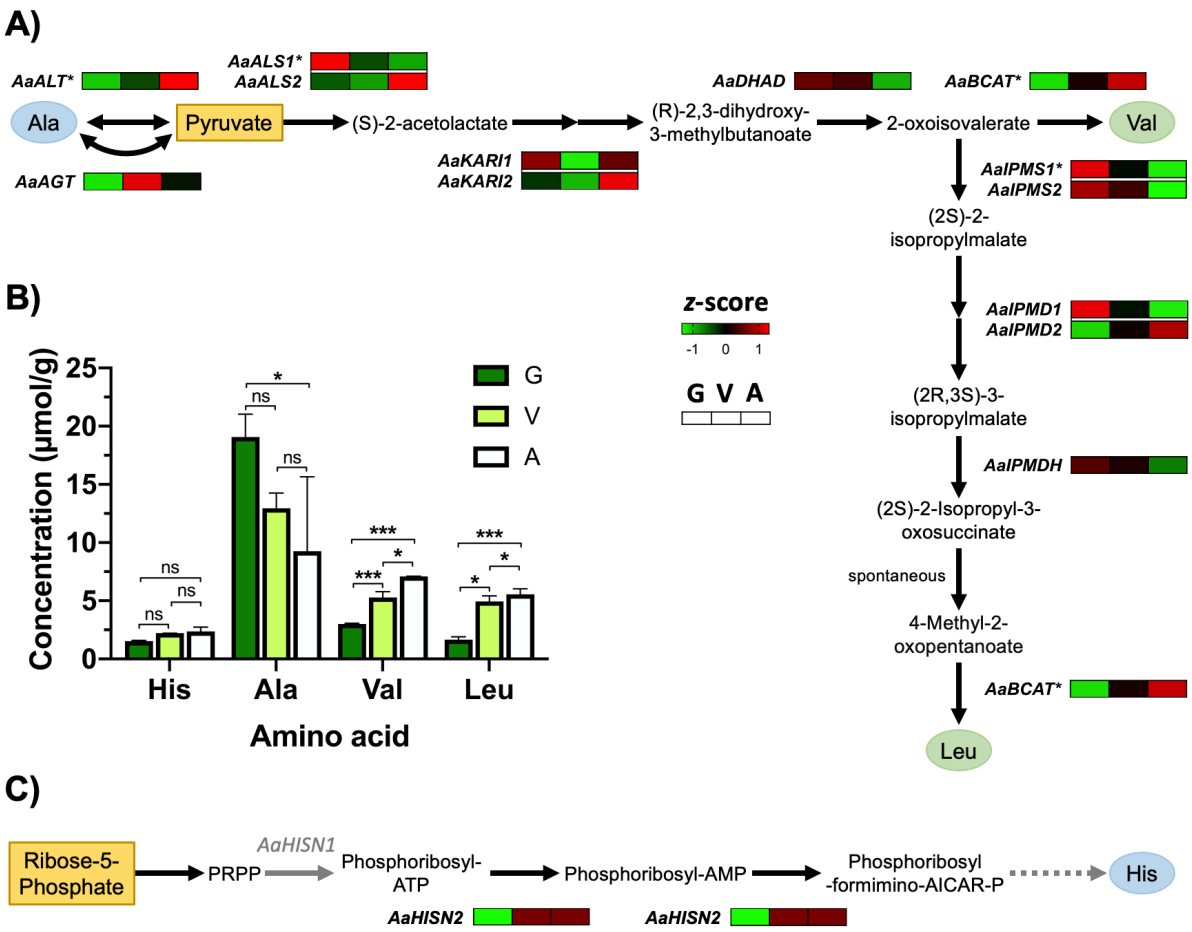


Figure 4.9. Ala, Val and Leu biosynthesis pathways derived from pyruvate. A) Accumulation of the enzymes catalyzing the Ala, Val and Leu biosynthesis pathways identified in the proteomic analysis of the three somaclonal variants of A. angustifolia. Yellow rectangles indicate the metabolic precursors. Enzymes detected in the proteomic analysis are indicated by black arrows. Enzymes not detected are indicated in gray. Numbers in the name of each enzyme indicate the number of each isoform found. Asterisks indicate differentially accumulated enzyme isoforms. Green ovals indicate AAs only synthesized in the plastids and blue ovals indicate AAs with questionable synthesis. Heatmaps attached to each enzyme show its accumulation level in all three phenotypes. **B)** Ala, Val and Leu concentration in the three somaclonal variants. The values of the means and the standard error are represented in each graph. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: non-significant differences. C) Accumulation of the enzymes catalyzing the His biosynthesis pathway. G: green plantlet, V: variegated plantlet, A: albino plantlet, His: histidine, Ala: alanine, Val: valine, Leu: leucine, Aa: Agave angustifolia, ALT: alanine transaminase (EC 2.6.1.2), AGT: alanine--glyoxylate transaminase (EC 2.6.1.44), ALS: acetolactate synthase (EC:2.2.1.6), KARI: ketol-acid reductoisomerase (EC:1.1.1.86), DHAD: dihydroxyacid dehydratase (EC:4.2.1.9), BCAT: branched-chain-amino-acid aminotransferase (EC:2.6.1.42), IPMS: isopropylmalate synthase (EC:2.3.3.13), IPMD: isopropylmalate dehydratase (EC:4.2.1.33), IPMDH: isopropylmalate dehydrogenase (EC:1.1.1.85),

HISN1: ATP-phosphoribosyltransferase (EC:2.4.2.17), HISN2: phosphoribosyl ATP pyrophosphohydrolase (EC:3.6.1.31), PRPP: Phosphoribosyl diphosphate.

4.3.4.4. AAs derived from α-ketoglutarate: Glu, Gln, Pro and Arg

Of the AAs derived from α-ketoglutarate, only Pro was identified in high concentrations in V and A phenotypes (**Figure 4.10B**). However, the results of the proteomic analysis identified differences in the accumulation of at least five enzymes of this pathway. The enzyme *Aa*GDH, which catalyzes the conversion of α-ketoglutarate to Glu, showed an increase in its abundance in the G phenotype. *Aa*AST2, another enzyme catalyzing the same reaction, showed an over-accumulation in the G phenotype compared to V. Three isoforms of the *Aa*GLN enzyme (*Aa*GLN1, 2, and 4), which catalyze the conversion of Glu to Gln, increased in abundance in the A and V phenotypes. Of the enzymes responsible for Arg biosynthesis, four of six enzymes were identified: *Aa*AGPR, *Aa*ACOAT, *Aa*OTC and *Aa*ASS. Only the latter, which catalyzes the conversion of citrulline to argininosuccinate, was up-accumulated in the A phenotype respect to G. Finally, the *Aa*P5CS enzyme, which catalyzes the first two steps of the Pro biosynthesis pathway to obtain L-glutamate 5-semialdehyde, was up-accumulated in the G and A phenotypes (**Figure 4.10A**).

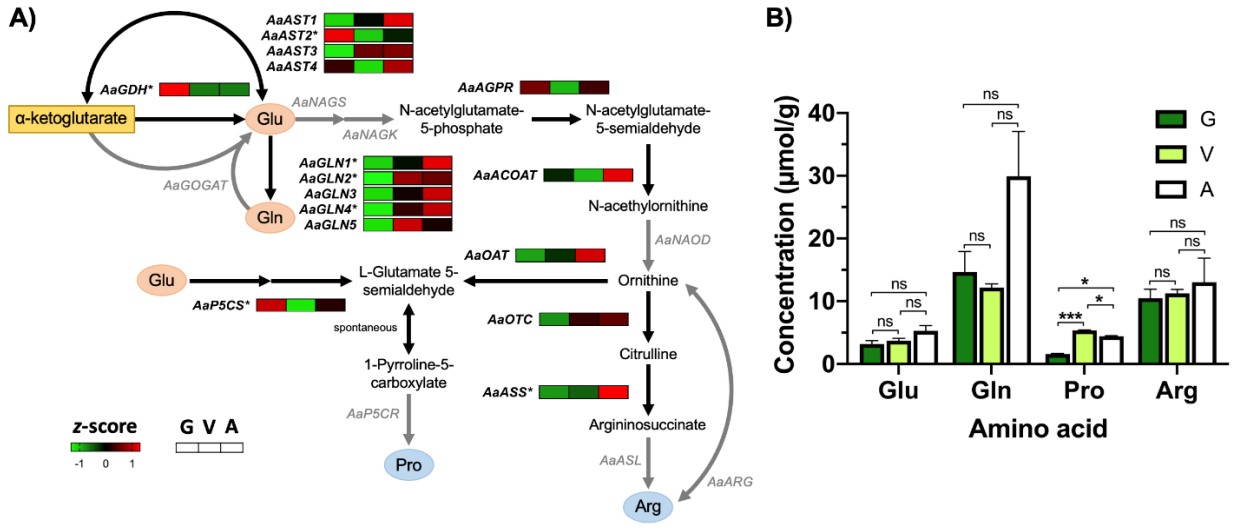


Figure 4.10. Glu, Gln, Pro and Arg biosynthesis pathways derived from α-ketoglutarate. A) Accumulation of the enzymes catalyzing the Glu, Gln, Pro and Arg biosynthesis pathways identified in the proteomic analysis of the three somaclonal variants of *A. angustifolia*. Yellow rectangle indicates the metabolic precursors. Enzymes detected in the proteomic analysis are indicated by black arrows. Enzymes not detected are indicated in gray. Numbers in the name of each enzyme indicate the number of each isoform found. Asterisks indicate differentially accumulated enzyme isoforms. Orange ovals indicate AAs

synthesized in both the chloroplast and other parts of the cell and blue ovals indicate AAs with questionable synthesis. Heatmaps attached to each enzyme show its accumulation level in all three phenotypes. **B)** Glu, Gln, Pro and Arg concentration in the three somaclonal variants. The values of the means and the standard error are represented in each graph. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: non-significant differences. G: green plantlet, V: variegated plantlet, A: albino plantlet, Glu: glutamate, Gln: glutamine, Pro: proline, Arg: arginine, Aa: *Agave angustifolia*, GDH: glutamate dehydrogenase (EC:1.4.1.4), GOGAT: glutamate synthase (EC:1.4.1.14), AST: aspartate transaminase (EC:2.6.1.1), GLN: glutamine synthetase (EC:6.3.1.2), NAGS: N-acetylglutamate synthase (EC:2.3.1.1), NAGK: N-acetylglutamate kinase (EC:2.7.2.8), AGPR: N-acetyl-gamma-glutamyl-phosphate reductase (EC:1.2.1.38), ACOAT: acetylornithine transaminase (EC:2.6.1.11), NAOD: acetylornithine deacetylase (EC:3.5.1.16), OTC: ornithine carbamoyltransferase (EC:2.1.3.3), ASS: argininosuccinate synthase (EC:2.1.3.3), ASL: argininosuccinate lyase (EC:4.3.2.1), ARG: Arginase (EC 3.5.3.1), OAT: ornithine aminotransferase (EC:2.6.1.13), P5CS: Delta-1-pyrroline-5-carboxylate synthase (EC:2.7.2.11 /EC:1.2.1.41), P5CR: pyrroline-5-carboxylate reductase (EC:1.5.1.2).

4.3.4.5. AAs derived from oxaloacetate: Asp, Asn, Lys, Met, Thr and Ile

The AAs derived from oxaloacetate (Asn, Asp, Lys, Met, Thr and Ile) showed different levels of accumulation in the three phenotypes. Asn was the only AA that did not show any changes in its concentration across all phenotypes. Met, Lys and Ile were found in high concentrations in the V and A phenotypes, while Thr was in high concentration in the V phenotype and Asp in the A phenotype (**Figure 4.11B**). The proteomic analysis showed showed that only one isoform of the enzyme *Aa*AST (*Aa*AST2), responsible for the conversion of oxaloacetate to Asp, was over-accumulated in the G phenotype with respect to V. Similarly, the enzyme *Aa*ASN, which catalyzes the conversion of Asp to Asn, was also over-accumulated in the G and V phenotypes compared to A, with a greater increase in G. For the biosynthesis pathway of Lys from Asp, it was possible to identify five of the seven pathway enzymes: *Aa*AK-HD, *Aa*ASADH, *Aa*LL-DAPAT, *Aa*DAPF and *Aa*DAPDC. Only *Aa*ASADH, which regulates the conversion of 4-phospho-L-aspartate to L-aspartate 4-semialdehyde, showed an increase in its accumulation in the G phenotype with respect to V. The product of the reaction catalyzed by *Aa*ASADH is an intermediate in the biosynthesis of Lys, Met and Thr. None of the enzymes of the Thr biosynthetic pathway showed changes in their accumulation. On the other hand, from the Met biosynthesis pathway an increase in the accumulation of the enzyme *Aa*CGS was found in the A phenotype with respect to G. This enzyme is responsible for the conversion of O-phospho-L-homoserine and Cys to L-cystathionine. In addition, isoforms of the enzyme *Aa*MS showed an over-accumulation in the A phenotype with respect to V. The *Aa*MS enzyme catalyzes the transfer of a methyl group

from 5-methyltetrahydrofolate to homocysteine to generate Met. Finally, four of the enzymes involved in Ile biosynthesis (which depends on the prior formation of Thr) were identified (**Figure 4.11A**). The accumulation of these enzymes was previously described as they also participate in the biosynthesis of Val and Leu (**Figure 4.11B**).

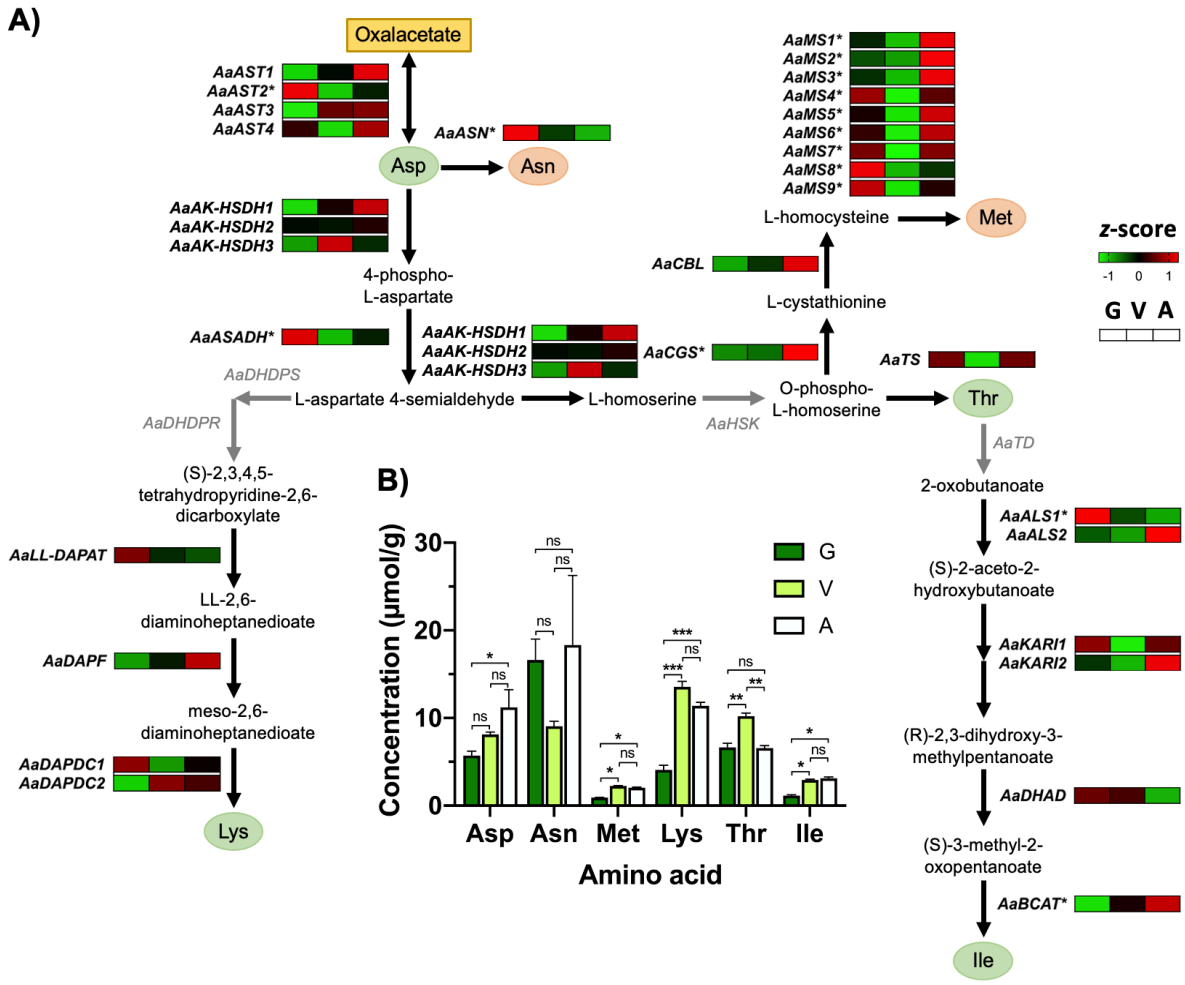


Figure 4.11. Asp, Asn, Lys, Met, Thr and Ile biosynthesis pathways derived from oxalacetate. A) Accumulation of the enzymes catalyzing the Asp, Asn, Lys, Met, Thr and Ile biosynthesis pathways identified in the proteomic analysis of the three somaclonal variants of *A. angustifolia*. Yellow rectangle indicates the metabolic precursor. Enzymes detected in the proteomic analysis are indicated by black arrows. Enzymes not detected are indicated in gray. Numbers in the name of each enzyme indicate the number of each isoform found. Asterisks indicate differentially accumulated enzyme isoforms. Green ovals indicate AAs only synthesized in the plastids and orange ovals indicate AAs synthesized in both the chloroplast and other parts of the cell. Heatmaps attached to each enzyme show its accumulation level in all three phenotypes. **B)** Asp, Asn, Lys, Met, Thr and Ile concentration in the three somaclonal variants.

The values of the mean and the standard error are represented in each graph. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: non-significant differences. G: green plantlet, V: variegated plantlet, A: albino plantlet, Asp: aspartate, Asn: asparagine, Met: methionine, Lys: lysine, Thr: threonine, Ile: isoleucine, Aa: *Agave angustifolia*, AST: aspartate transaminase (EC:2.6.1.1), ASN: asparagine synthase (EC:6.3.5.4), AK-HSDH: aspartate kinase-homoserine dehydrogenase (EC:2.7.2.4/EC:1.1.1.3), ASADH: aspartate-semialdehyde dehydrogenase (1.2.1.11), DHDPS: dihydrodipicolinate synthetase (EC:4.3.3.7), DHDPR: dihydrodipicolinate reductase (EC:1.17.1.8), LL-DAPAT: LL-diaminopimelate aminotransferase (EC:2.6.1.83), DAPF: diaminopimelate epimerase (EC:5.1.1.7), DAPDC: diaminopimelate decarboxylase (EC:4.1.1.20), CGS: cystathionine gamma-synthase (EC:2.5.1.48), CBL: cystathionine beta-lyase (EC:4.4.1.13), MS: methionine synthase (EC:2.1.1.14), TS: threonine synthase (EC:4.2.3.1), TD: threonine deaminase (EC:4.3.1.19), ALS: acetolactate synthase (EC:2.2.1.6), KARI: ketol-acid reductoisomerase (EC:1.1.1.86), DHAD: dihydroxyacid dehydratase (EC:4.2.1.9), BCAT: branched-chain-amino-acid aminotransferase (EC:2.6.1.42).

4.3.5. Expression analysis of genes encoding key enzymes in AA metabolism

To validate the proteomic data, we examined the expression profiles of five genes (*AaBCAT*, *AaMS*, *AaCGS*, *AaTS* and *AaCM*) encoding key enzymes in AA metabolism in somaclonal variants. This analysis revealed notable changes in the expression of four out of the five selected genes (**Figure 4.12**). Specifically, the *AaBCAT* gene was found to be over-expressed in the A phenotype, while the *AaMS* and *AaCGS* genes showed higher expression levels in the G phenotype. The *AaCM* gene exhibited overexpression in both the G and A phenotypes, with the A phenotype showing the highest expression. On the other hand, the *AaTS* gene did not show significant differences in expression between the variants.

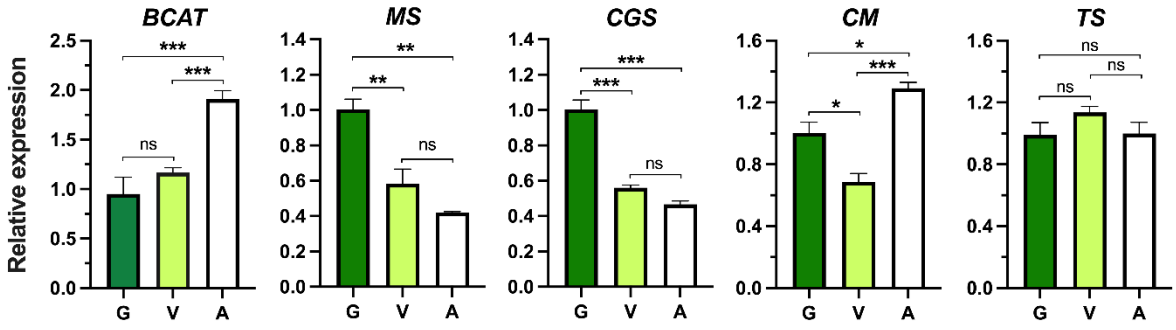


Figure 4.12. Expression profiles of five genes encoding key enzymes in AA biosynthesis pathways in *A. angustifolia* somaclonal variants. The expression levels were normalized using 2–ΔΔCT, with the G phenotype as the reference expression level. The values of the mean and the standard error of three biological replicates are represented in each graph. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: non-significant

differences, G: green plantlet, V: variegated plantlet, A: albino plantlet, *BCAT*: branched-chain-amino-acid aminotransferase, *CGS*: cystathionine gamma-synthase, *MS*: methionine synthase, *CM*: chorismate mutase, *TS*: threonine synthase.

Comparing the expression profiles obtained through quantitative real-time PCR (qRT-PCR) with the proteomic data for the *AaBCAT* and *AaCM* genes, we observed consistent trends at both the mRNA and protein level, indicating their accumulation in the A phenotype. However, for the *AaCGS*, *AaMS* and *AaTS* genes, there was no such correspondence between expression profiles and proteomic data. These findings suggest a need for further investigation into the intricate interplay between transcription and translation regulation in AA metabolism.

4.4. DISCUSSION

One of the advantages of working with plantlets that lack chloroplasts and an active photosynthetic metabolism is that these models are capable to modulate their metabolism to response to stress conditions and adapt to heterotrophy (de Luna-Valdez *et al.*, 2014). One of the most novel approaches to study of these albino/chlorotic models is proteomics. However, most of the previous reports have focused on aspects related to photosynthetic metabolism and the biosynthesis of chlorophyll and accessory pigments (Yang *et al.*, 2020; Shi *et al.*, 2017). This has neglected aspects of primary and/or secondary metabolism that could provide information about how a plant transitions from autotrophic to heterotrophic metabolism. In our study, we used a proteome approach to assess changes in the three somaclonal varieties of *A. angustifolia* (G, V, and A). The general results revealed that the DAPs found were involved in diverse metabolic pathways. These included pathways related to the biosynthesis of phenylpropanoids, glutathione metabolism, pathways involved with sugars, such as the pentose phosphate pathway and glycolysis, fructose and mannose metabolism, the TCA cycle, and AA metabolism (**Figure 4.6**).

AAs are monomers that allow protein biosynthesis. They also play a crucial role as building blocks in the production of numerous secondary metabolites (Heinemann y Hildebrandt, 2021). In plants, unlike in most other organisms, AA metabolism occurs almost entirely within the domain of the chloroplast (Rolland *et al.*, 2018). However, the presence of plant phenotypes characterized by alterations in plastid maturation and manifesting as a lack of pigmentation raises a number of intriguing issues concerning the homeostasis of these plants' AA metabolism. An increase in AA concentration has been reported in the most studied albino/chlorotic models, including *Camellia sinensis* genotypes and *A. thaliana* mutants (Lu *et al.*, 2019; Li *et al.*, 2018a; Satou *et al.*, 2014). Though the prevalence of AAs is a controversial subject, so are the factors that have led to their

rise. In this study, it has been possible to determine the status of proteins involved in the AA biosynthesis pathways and the specific concentration of each AA present in somaclonal *A. angustifolia* plantlets.

4.4.1. Over-accumulation of aromatic AA in A and V phenotypes suggests increases in biosynthesis of molecules with photoprotective functions

Our analysis revealed that the V and A phenotypes had higher concentrations of the three aromatic AAs (Phe, Tyr and Trp) than the G phenotype (**Figure 4.8B**). This is despite the fact that these AAs are synthesized *de novo* in the chloroplast (Lynch y Dudareva, 2020). Aromatic AAs are precursors for many secondary metabolites. Trp is a key precursor for auxins (including indole 3-acetic acid), indolamines such as serotonin and melatonin, phytoalexins, and indole glucosinolates (Corpas *et al.*, 2021; Negri *et al.*, 2021). Similarly, Tyr is a substrate for the synthesis of tocopherols, plastoquinone and ubiquinone (Xu *et al.*, 2020). Phe is the AA precursor of more than 8,000 phenolic compounds, mainly those that belong to the phenylpropanoid pathway and metabolites produced from it, therefore requiring the largest carbon flux (Lynch *et al.*, 2020).

The shikimate pathway produces chorismate during aromatic AA biosynthesis. Chorismate acts as an intermediate in the metabolism of aromatic AA, and from this metabolite the pathway diverges into two branches, the Phe and Tyr and the Trp biosynthesis pathways (Parthasarathy *et al.*, 2018). In our proteomic analysis, six of the seven shikimate pathway enzymes were detected, although none of them showed differences in their abundance among phenotypes (**Figure 4.8A**). In the Phe and Tyr biosynthesis pathway, the A and V phenotype over-accumulated *Aa*CM (**Figure 4.8A**). Interestingly, the orthologous protein to this enzyme in *A. thaliana* corresponds to CHORISMATE MUTASE 2 (*At*CM2) (AT5G10870), which is a nonallosteric form. The *At*CM2 enzyme forms part of the microbial-like phenylpyruvate pathway, which derives from the chloroplastic arogenate pathway, and it is responsible for the biosynthesis of Phe in the cytosol (Lynch *et al.*, 2020; Westfall *et al.*, 2014). It has been found in *Petunia* petals that Phe is mostly produced via the plastidial arogenate pathway (Maeda *et al.*, 2010). However, carbon flux can be redirected through the cytosolic phenylpyruvate pathway if the input route to the plastidial pathway is blocked (Yoo *et al.*, 2013). In V and A phenotypes, where chloroplasts display abnormalities, phenylalanine biosynthesis may be shifted towards the cytosolic pathway. Therefore, it is possible that the high accumulation of Phe in A and V phenotypes is the result of

an activation of the cytosolic pathway as a result of the blockage in the chloroplast pathway due to the absence of this organelle.

Changes in CO_2 fixation, decreased expression of genes encoding subunits of ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and proteins composing the LHC, morphological abnormalities, etc., have all been documented for A and V phenotypes by our group (Hernández-Castellano *et al.*, 2020; Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016). These photosynthetic impairments in V and A phenotypes may suggest that these plantlets are susceptible to photo-oxidative stress, which would require the presence of molecules with photoprotective properties. The rise in Tyr concentration is another indicator that V and A phenotypes must produce a bigger quantity of photoprotective molecules. One of the main products formed from this AA is plastoquinone (PQ), which acts as an electron carrier between photosystem II and cytochrome b6f complex during phosphorylation in the chloroplast (Van Eerden *et al.*, 2017). In times of stress, PQ has also been associated with the prevention of lipid peroxidation, protein oxidation and DNA damage (Liu y Lu, 2016). It has been discovered that in the chloroplasts of *A. thaliana*, PQ cooperates with other antioxidants to counteract the deleterious effects of singlet oxygen derived from exposure to excessive light (Kumar *et al.*, 2020). Therefore, a probable increase in PQ in V and A phenotypes as a result of higher Tyr, coupled with a possible increase in the content of specific flavonoids (**Figure 4.6**), would support the idea that V and A phenotypes may have a large accumulation of photoprotective compounds.

These results support the idea that if V and A phenotypes are more susceptible to light-induced stress, then an increase in the pool of aromatic AA may lead to an acceleration in the production of different molecules with photoprotective functions. Experiments centered on metabolomics may shed light on this intriguing subject.

4.4.2. Changes in the accumulation of *Aa*ALS and *Aa*BCAT proteins favor the high concentration of BCAAs in V and A phenotypes

BCAAs (Val, Leu and Ile), another group of AA produced in the chloroplast, were found in high concentrations in the variants V and A compared to the G (**Figure 4.9B and 4.11B**). Proteomic analysis showed that three DAPs, *Aa*ALS, *Aa*IPMS, and *Aa*BCAT, are involved in the biosynthesis of BCAAs (**Figure 4.9A and 4.11A**). The accumulation of the *Aa*ALS enzyme was reduced in the V phenotype compared to the G, and the lowest concentration was found in the A. It is important to highlight that this enzyme is one of the primary targets of ~50 agricultural

herbicides (Lonhienne *et al.*, 2022). *In vitro* reconstitution of the catalytic subunit of the *At*ALS with its regulatory subunit stimulates its catalytic activity, but also makes it susceptible to inhibition by all BCAAs, particularly the Leu and Val combination (Lee y Duggleby, 2001). The ALS enzyme has also been described as having a redox regulation, in which its activity is inhibited in the presence of ubiquinone-type molecules that oxidize the flavin adenine dinucleotide (FAD) cofactor (Lonhienne *et al.*, 2017). Possible explanations for the decrease in *Aa*ALS accumulation include oxidative stress produced by the disruption to photosynthetic metabolism and an increase in quinone-type molecules that oxidize the cofactor of this enzyme. This may be supported in part by the increase in Tyr (**Figure 4.8B**), a precursor to ubiquinone. Additionally, high concentrations of Leu and Val above Ile could inhibit *Aa*ALS activity alone or in combination.

In contrast to the *Aa*ALS enzyme, the accumulation of *Aa*BCAT was greater in the A phenotype compared to the G (**Figure 4.9B and 4.11B**). *Aa*BCAT plays a dual role in BCAA biosynthesis and degradation (Lee *et al.*, 2019a). The orthologous protein to *Aa*BCAT in *A. thaliana* corresponds to BCAT-3 (AT3G49680). *At*BCAT-3 is a plastid-localized enzyme that is involved in both primary and secondary metabolism via the biosynthesis of BCAAs and glucosinolates, respectively (Knill *et al.*, 2008). Therefore, it is possible that the higher concentration of Val, Leu and Ile in the A variants is affected in part by the elevated accumulation and activity of *Aa*BCAT during BCAA biosynthesis.

4.4.3. Changes in Met and Cys pools in A and V phenotypes suggest increased formation of glutathione and S-adenosylmethionine

The sulfur-containing AAs (Cys and Met) showed different concentrations among the three phenotypes of *A. angustifolia* (**Figure 4.7B and 4.11B**). Cys is a central AA in sulfur metabolism. It acts as the main sulfur donor compound and is a precursor for the synthesize of metabolites such as Met, vitamins, cofactors, iron-sulfur (Fe-S) clusters, ethylene, and glutathione (GSH) (Li *et al.*, 2020; Romero *et al.*, 2014). The proteomic analysis did not reveal changes in the accumulation of the two enzymes involved in Cys biosynthesis (**Figure 4.7A**). However, the A phenotype presented almost undetectable Cys levels compared to the G and V phenotypes (**Figure 4.7B**). One of the plant dilemmas related to sulfur is whether the Cys pool will be directed towards protein synthesis, which is associated with plant growth, or towards the formation of glutathione (GSH) (Speiser *et al.*, 2018). GSH is a major scavenger of reactive oxygen species (ROS) (Filiz *et al.*, 2019). In our previous study, we reported that A and V phenotypes exhibit a certain retardation in their growth with respect to G (Duarte-Aké *et al.*, 2016). This leads us to

think that part of the Cys pool could go directly towards the formation of GSH, where it would act as an antioxidant.

On the other hand, Met is an AA constituent of proteins, essential in the early stages of translation and a precursor to S-adenosylmethionine (SAM). SAM is a molecule involved in many metabolic pathways, including the biosynthesis of ethylene, biotin, polyamines (spermidine and spermine), nicotinamide, phytosiderophores, and others (Sehar *et al.*, 2022; Astolfi *et al.*, 2021). One of the known functions of SAM is that it acts as the main donor of methyl groups in various methylation reactions, including the methylation of histones, DNA and RNA (Ouyang *et al.*, 2020). Previous studies focused on detecting epigenetic changes in histone modifications in the A phenotype of *A. angustifolia* reported an increase in euchromatin (H3K4me2 and H3K36me2) and heterochromatin (H3K9me2 and H3K27me3)-related methylation marks (Duarte-Aké *et al.*, 2016). Additionally, in the albino *Agave*, there are changes in the levels of global DNA methylation strongly related to the state of development and maturation of the plantlets as well as their shoots (Us-Camas *et al.*, 2017). The high increase in Met in A plantlets appears to be fueled by increased biosynthesis. However, further studies are needed to reveal the ultimate metabolic fate of this AA. Due to the central role of SAM in aspects related to the biosynthesis of chlorophyll, lignins and suberins of the cell wall, flavonoids and other secondary metabolites (Amir, 2010), we believe that this metabolite may be of great interest for further studies.

4.4.4. Accumulation of Lys, BCAAs and aromatic AAs in V and A phenotypes suggests activation of catabolic pathways for alternative energy production

Asp is a precursor in the aspartate metabolic pathway, which allows the biosynthesis of Asn, Lys, Thr, Met and Ile (Han *et al.*, 2021). Of the four isoforms identified of the *Aa*AST enzyme, which catalyzes Asp production, only *Aa*AST2 was over-accumulated in the G phenotype compared to V**.** However, the rest of the isoforms (*Aa*AST1, 3 and 4) have a slight tendency to over-accumulate in the A phenotype (**Figure 4.11A**). Therefore, the high concentration of Asp in the A phenotype (**Figure 4.11B**) could act as a reserve pool that sustains the high production of the rest of AAs that make up this aspartate metabolic pathway.

Two AAs that showed high concentrations in V and A plantlets compared to G were Pro and Lys (**Figure 4.10B** and **4.11B**). In the case of Lys, the proteome of the somaclonal variants revealed only an over-accumulation of the enzyme *Aa*ASADH in the G phenotype respect to V (**Figure 4.11B**). Like other AAs, Lys levels also increase in response to abiotic stress (Obata y

Fernie, 2012). For example, *A. thaliana* plants subjected to sugar starvation induced by dark treatment showed an increase in the content of Lys as well as BCAAs (Leu, Val and Ile) and aromatic AAs (Phe, Tyr and Trp) (Hirota *et al.*, 2018; Avin-Wittenberg *et al.*, 2015; Araújo *et al.*, 2010). High concentrations of these AAs have also been found in our V and A phenotypes and other albino/chlorotic models (Yamashita *et al.*, 2021; Lu *et al.*, 2019; Li *et al.*, 2018a). It has been reported that the enrichment of these AAs and the activation of their catabolic pathways gives rise to alternative substrates that can enter the respiratory chain for energy production in the mitochondria (producing 24 or more ATP molecules) (Izumi y Ishida, 2019; Hildebrandt *et al.*, 2015). It is likely that, in addition to biosynthetic events raising the AA pool in variants V and A, catabolic pathways of these AA may also be activated as an alternative energy production mechanism.

4.4.5. Photo-oxidative stress in V and A plantlets promotes the accumulation of Pro

Proline is an AA that has been studied for its ability to accumulate under different abiotic stress conditions (Ghosh *et al.*, 2022; Forlani *et al.*, 2019). Two enzymes are involved in the biosynthesis of Pro from Glu. Only the *Aa*P5CS enzyme was found to over-accumulate in the G and A phenotypes (**Figure 4.10A**). High concentrations of Pro can repress *P5CS* (Hellmann *et al.*, 2000; Peng *et al.*, 1996), which may explain the lower levels of *Aa*P5CSS in A and V phenotypes. Pro has been described as a stabilizing agent for proteins and cell membranes that protects the cell from over-accumulation of ROS under stress conditions (Ghosh *et al.*, 2022; Forlani *et al.*, 2019). For example, when *Pisum sativum* leaf discs treated with methyl violagen (an oxidant targeting the chloroplast) and subjected to high light-induced photo-oxidative stress, an increase in ROS production was observed, which led to an increase in the level of antioxidants such as ascorbate, GSH, and Pro, particularly the latter (Aswani *et al.*, 2019). Thus, the high levels of Pro observed in the V and A phenotypes may be due to the photo-oxidative stress to which they are subjected (**Figure 4.10B**).

4.4.6. mRNA expression and protein accumulation in genes of AA biosynthesis

The expression analysis of five key genes in AA biosynthesis (*AaBCAT*, *AaCM, AaMS*, *AaCGS*, and *AaTS*) revealed a discrepancy between mRNA expression and protein accumulation for the last three (**Figure 4.12**). This lack of integration has been described previously (Lyu *et al.*, 2021; Chen *et al.*, 2017) and linked to regulatory mechanisms ranging from the post-transcriptional and translational to the post-translational level (Tian *et al.*, 2004). The mechanisms

governing these genes in albino models remain unknown and warrant further investigation. Identifying such mechanisms could shed light on the molecular basis of AA metabolism in these models and facilitate the development of targeted interventions to improve their metabolic efficiency.

4.4.7. Metabolic reprogramming in albino and variegated *Agave* plantlets

The increases in the concentration of AAs that are either exclusively or partially synthesized in the chloroplast (Phe, Tyr, Trp, Leu, Val, Ile, Met, Lys, Asp, Thr, Cys, Gly, and Ser) indicate that the total or partial absence of this organelle in A and V plantlets has a significant impact on AA metabolic pathways (**Figure 4.13**). Our results also suggest that the enzymes *Aa*CM, *Aa*ALS, *Aa*BCAT, *Aa*IPMS1, *Aa*SHMT, *Aa*AST, *Aa*CGS, and *Aa*MS play critical roles in the survival of chloroplast-deficient *Agave* plantlets. In proteomic and metabolite analyses performed on heterotrophic members of the plastid family, such as proplastids, etioplasts and chromoplasts, a strong protein activity linked to AA metabolism and an increased AA concentration has been reported (Rödiger *et al.*, 2021; Armarego-Marriott *et al.*, 2019; Majeran *et al.*, 2011; Majeran *et al.*, 2010). In our study models, both the albino and variegated *Agave* plantlet leaves harbor numerous non-photosynthetic and immature plastids similar to proplastids (Hernández-Castellano *et al.*, 2020). Considering that the proplastid is an organelle in high proliferation, growth and translation (Loudya *et al.*, 2021)**,** the activation and reprogramming of AA metabolic pathways in A and V plantlets could be key to sustaining such processes (Hildebrandt, 2018).

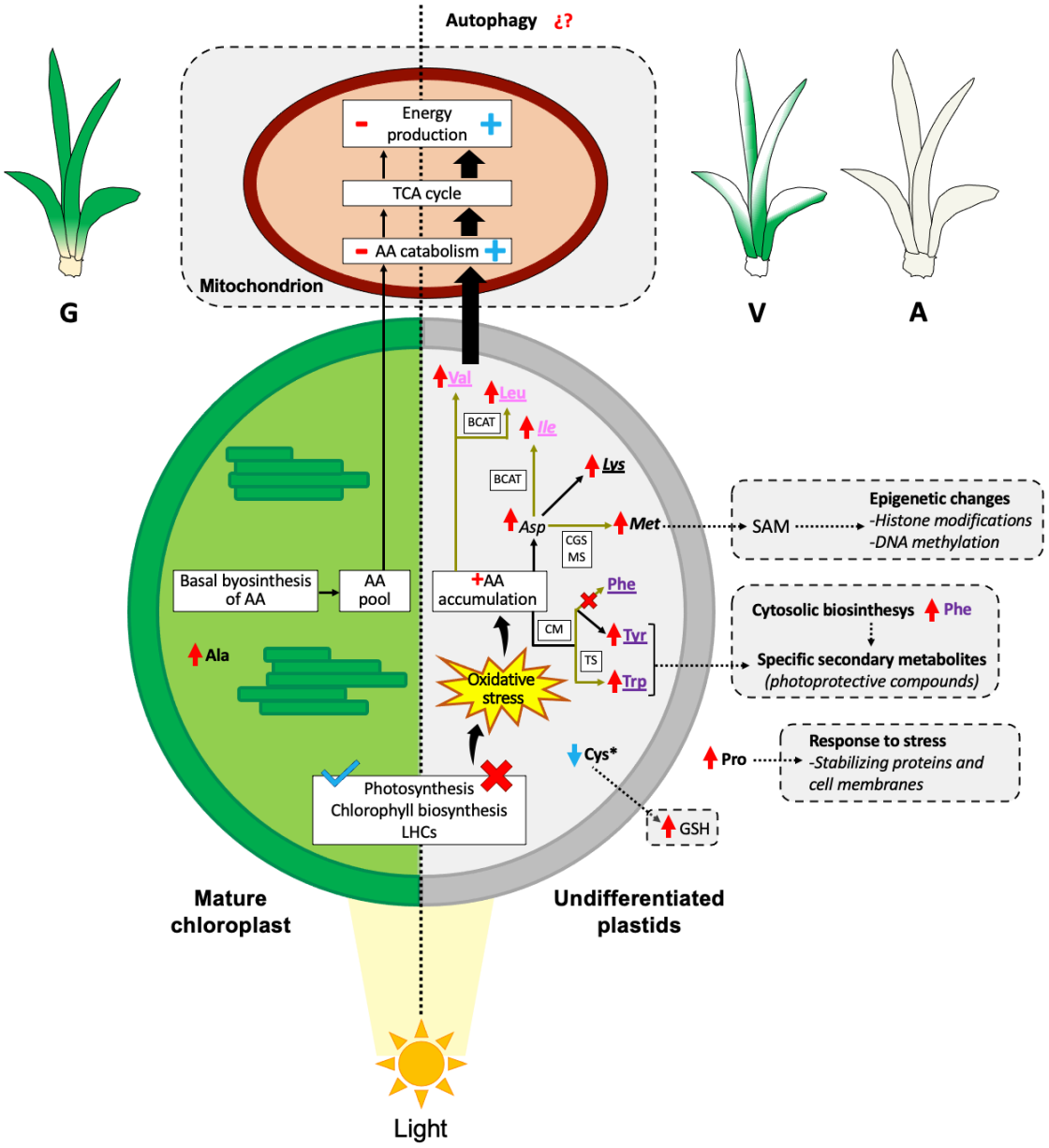


Figure 4.13. Model for AA metabolism in albino and variegated somaclonal variants of *A. angustifolia*. In green plantlets with mature chloroplasts, the chlorophyll and carotenoid biosynthesis pathways, photosynthetic metabolism, and light harvesting complexes (LHCs) are active. This allows the chloroplast to obtain the necessary energy to carry out its functions, among which AA biosynthesis stands out. However, in undifferentiated plastids present in both albino and variegated plantlets, the ability to produce photosynthetic pigments and perform photosynthesis is impaired. Under these conditions, exposure to light and other environmental factors triggers oxidative stress conditions, which also activate certain AA biosynthetic pathways. Among them are BCAA (Val, Leu and Ile), Met and aromatic AA (Tyr, Phe and Trp) biosynthetic pathways. Pro was also found in high concentrations in these plants. Additionally,

hypothetical destinations for each of these AA are presented. The hypothetical autophagy pathway that could be key for alternative energy generation under oxidative stress conditions in albino and variegated plants is presented at the top of the figure. Full lines with arrows indicate AA biosynthetic pathways. Olive green lines indicate those AA biosynthetic pathways that appear to influence AA accumulation. Dotted lines indicate hypothetical destinations of AA. White rectangles indicate key enzymes in the AA biosynthesis pathways found in the proteomic analysis of the phenotypes. AAs in pink indicate BCAAs. AAs in italics indicate AAs that are derived from the aspartate metabolic pathway. AAs in purple indicate aromatic AAs. Underlined AAs indicate AAs previously linked to autophagy mechanisms. Red and blue arrows next to each AA indicate a higher or lower concentration, respectively. Red cross represents a blockage/corruption in that path or process. * indicates an AA found only in low concentrations in A phenotype. TCA: tricarboxylic acid, AA: amino acid, BCAT: branched-chain-amino-acid aminotransferase, CGS: cystathionine gamma-synthase, MS: methionine synthase, CM: chorismate mutase, TS: threonine synthase, GSH: glutathione, SAM: S-adenosylmethionine.

On the other hand, it is also important to consider that the increase in aromatic AAs, BCAAs, Lys and Pro in A and V plantlets, which are markers linked to AA catabolic processes typical of autophagy, open up new questions in the field of AA metabolism (Hirota *et al.*, 2018). As discussed before, the oxidative stress conditions in which the V and A variants are found could also be activating some of these AA catabolic pathways for alternative energy production (Izumi y Ishida, 2019; Hildebrandt *et al.*, 2015). In light of this evidence, it is likely that in addition to biosynthetic events raising the AA pool of V and A phenotypes, other mechanisms such as autophagy could actively participate in albino plantlets for survival of these plants. In any of the scenarios proposed, these albino phenotypes seem to be subject to metabolic reprogramming that acts as a survival strategy against adverse environmental conditions in the absence of the most important organelle of the plant cell.

While we have identified key enzymes that suggest a reprogramming of certain AA metabolic pathways, we cannot rule out the activation of other metabolic pathways that do not fully involve the chloroplast. These include the mithochondrial glycolate pathway and the cytosolic phenylpyruvate pathway. Further studies are needed to determine the final fate of these AAs and their impact on plantlets physiology. This will help us to better understand how these somaclonal variants respond to the challenge of not presenting mature and functional chloroplasts.

Albino and variegated plantlets of *A. angustifolia* accumulate elevated levels of free AAs, specifically those exclusively or partially synthesized in the chloroplast. This accumulation indicates a potential compensatory mechanism for the lack of photosynthesis in these

phenotypes. This survival mechanism would be characterized by an enhanced activity of specific enzymes involved in the metabolic pathways of aromatic AAs (chorismate mutase and tryptophan synthase), BCAAs (acetolactate synthase and branched-chain-amino-acid aminotransferase) and Met (cystathionine gamma-synthase and methionine synthase). These enzymes play a crucial role in adapting to adverse environmental conditions and sustaining metabolic homeostasis. Validation of the proteomic data through gene expression analysis revealed consistent trends at both the mRNA and protein level for the *AaBCAT* and *AaCM* genes, indicating an over-accumulation in albino plantlets. However, discrepancies between certain genes, including *AaCGS*, *AaMS*, and *AaTS,* were detected. This suggests the presence of regulatory mechanisms at the post-transcriptional, translational and/or post-translational level, which contribute to the fine-tuning of AA metabolism in chloroplast-deficient plantlets. Our findings highlight the importance of reprogramming AA metabolic pathways for the survival of photosynthetically inactive plants lacking chloroplasts. This adaptation allows them to cope with challenging environmental conditions and maintain essential metabolic functions. We acknowledge that there is still much to explore in understanding the complete range of compensatory mechanisms and specific enzyme interactions in these plantlets. Future studies could focus on unraveling the regulatory networks that underlie these compensatory mechanisms and further investigate the interplay between different metabolic pathways. By addressing these aspects, we can gain a more comprehensive understanding of the complex adaptations exhibited by plants lacking chloroplasts.

CAPÍTULO V

DISCUSIÓN, CONCLUSIONES GENERALES Y PERSPECTIVAS

5.1. DISCUSIÓN

Las plantas con fenotipos albinos son modelos invaluables y poco apreciados para el estudio de la biogénesis del cloroplasto. Los estudios transcriptómicos previos hechos en estas plantas han revelado algunas pistas sobre las principales alteraciones metabólicas y fisiológicas que resultan de la ausencia parcial o total de cloroplastos (Yan *et al.*, 2022; Wang *et al.*, 2020a; Li *et al.*, 2018b; Li *et al.*, 2017; Shi *et al.*, 2017; Satou *et al.*, 2014). El estudio transcriptómico realizado aquí abordó un análisis comparativo de los perfiles de expresión de genes vinculados a la biogénesis del cloroplasto en el tejido foliar y meristemático de plántulas de *Agave* verde y albina. Los resultados revelaron una intensa reprogramación transcripcional caracterizada por un aumento en la expresión de la mayoría de los genes estudiados en los tejidos de la plántula de *Agave* albina.

La elevada actividad transcripcional en la hoja albina estuvo vinculada a la división del plástido, la maquinaria de importación de proteínas, la actividad transcripcional de la PEP, la regulación postranscripcional, la actividad ribosomal plastídica, el ensamble de los PSI y PSII y el metabolismo fotosintético, entre otros. Este comportamiento transcripcional inusual en el *Agave* albino fue contrastado con el primer mapa transcriptómico realizado en un gradiente de desarrollo de una hoja de trigo verde (Loudya *et al.*, 2021). Este mapa profundizó en los principales cambios transcripcionales relacionados con la diferenciación de la célula y el plástido partiendo desde el meristemo (tejido rico en proplástidos) hasta la punta de la hoja (tejido rico en cloroplastos maduros). Esta comparación reveló que los perfiles de expresión de los genes en la hoja albina eran similares a los observados en los primeros estadios de desarrollo, los cuales coinciden con el meristemo y la base de la hoja de trigo. Por otro lado, los perfiles de expresión en el *Agave* albino fueron muy similares a los reportados para las mutantes albinas *apg2* y *apg3* de *A. thaliana* (Satou *et al.*, 2014).

La reprogramación transcripcional observada en los tejidos del *Agave* albino es comparable a la descrita para la mutante *cue8* de *A. thaliana* por Loudya *et al.* (2020). Es posible que un mecanismo de corrección anterógrado esté actuando en nuestro modelo albino con el objetivo de compensar los defectos durante la diferenciación temprana del plástido. En esta

hipotética respuesta correctiva, el proplástido de la planta de *Agave* albino estaría enviando señales retrógradas de origen desconocido al núcleo, para indicarle que no ha logrado diferenciarse. Esto desencadenaría una reprogramación transcripcional nuclear que involucra *PhANGs*, al gen *GLK*, los factores sigma, al gen que codifica NEP así como a los genes *PPR*. Más estudios son necesarios para confirmar la posible actuación de este mecanismo correctivo en nuestro modelo.

Uno de los resultados más interesantes para nosotros fue el incremento generalizado en la expresión de los genes que codifican para proteínas PPR en los tejidos del *Agave* albino. Estos genes son considerados jugadores protagónicos en el metabolismo de ARN de los plástidos, particularmente durante sus estadios tempranos de desarrollo (Wang *et al.*, 2021; Lurin *et al.*, 2004). Sin embargo, el estadio de diferenciación de la célula también determina los niveles de expresión de genes *PPR*. Por ejemplo, el gen *GUN1*, un regulador central de la señalización retrógrada, se expresa intensamente en hojas jóvenes y en crecimiento, mientras que en hojas maduras, tallos y raíces, su expresión disminuye (Wu *et al.*, 2018). Casos similares de expresión han sido descritos en otros genes *PPR* como *PDM3* y *AtDPG1* (Zhang *et al.*, 2017; Liu *et al.*, 2016). Las mutantes con una reducción parcial o completa de la expresión de genes *PPR* generalmente presentan una reducción en la producción de pigmentos fotosintéticos, afectaciones en las membranas tilacoidales y en la capacidad fotosintética, reducción en la síntesis de proteínas y en la actividad de la PEP (Huang *et al.*, 2020; Lee y Kang, 2020; Yuan *et al.*, 2019; Tadini *et al.*, 2018; Wang *et al.*, 2018b).

No nos queda duda que el incremento en la expresión de genes *PPR* esta estrechamente ligado al bloqueo en la biogénesis del cloroplasto. En el mapa transcriptómico de Loudya *et al.* (2021), los genes *PPR* mostraron incrementos en su expresión en sitios de la hoja donde abundan los proplástidos. Conforme avanza la diferenciación del proplástido, la expresión de los genes *PPR* empieza a decaer. Por su parte, las mutantes *apg* mencionadas previamente, soportan nuestra idea al no presentar cloroplastos y presentar una fuerte expresión de los genes *PPR* (Satou *et al.*, 2014; Motohashi *et al.*, 2012). Curiosamente, al comparar los perfiles de expresión de estos genes en las mutantes *apg2* y *apg3* con los de la mutante clorótica *ch42*, el incremento en la actividad transcripcional de los genes *PPR* es más discreta. Estos estudios soportan la estrecha relación entre el estadio de desarrollo del plástido y la actividad transcripcional de estos genes. Nuestro estudio revela un papel clave de los genes *PPR* durante las fases tempranas de la biogénesis del cloroplasto y propone que su alta expresión en el *Agave*

albino podría formar parte del mecanismo de corrección anterógrado descrito previamente, el cual intenta remover al proplástido de su estadio juvenil.

Por otro lado, el incremento en la concentración de los AA que son exclusiva o parcialmente sintetizados en el plástido (Phe, Tyr, Trp, Leu, Val, Ile, Met, Lys, Asp, Thr, Cys, Gly and Ser), revela que el bloqueo en la biogénesis del cloroplasto conduce a una reprogramación en el metabolismo de los AA en las plántulas de *Agave* albina y variegada. Nuestro análisis proteómico también reveló que las enzimas AaCM, AaALS, AaBCAT, AaIPMS1, AaSHMT, AaAST, AaCGS y AaMS juegan roles importantes para la sobrevivencia de estos fenotipos que carecen de cloroplastos funcionales.

Por ejemplo, el daño en el metabolismo fotosintético de estas plántulas (Hernández-Castellano *et al.*, 2020; Us-Camas *et al.*, 2017; Duarte-Aké *et al.*, 2016) podría indicar que son susceptibles a estrés fotooxidativo y que requieren moléculas con propiedades fotoprotectoras, las cuales podrían provenir de AA aromáticos como la Tyr y la Phe (Kumar *et al.*, 2020; Lynch *et al.*, 2020; Liu y Lu, 2016). El incremento de Met en las plántulas variegadas y albinas es interesante ya que este AA es un precursor de SAM. Este metabolito está involucrado en la biosíntesis de etileno, biotina, poliaminas, nicotinamida, fitosideróforos y participa como donador de grupos metilo en procesos de metilación de histonas, ADN y ARN (Sehar *et al.*, 2022; Astolfi *et al.*, 2021; Ouyang *et al.*, 2020). La Pro, un AA clave en la respuesta a estrés abiótico (Ghosh *et al.*, 2022; Forlani *et al.*, 2019), también estuvo en altas concentraciones en los *Agaves* albino y variegado lo que podría deberse al estrés fotooxidativo al que se encuentran sujetas.

Finalmente, considerando que el proplástido es un organelo en alta proliferación, crecimiento y actividad traduccional (Loudya *et al.*, 2021) y que es dominante en los fenotipos albino y variegado de *Agave*, la activación de las rutas metabólicas de los AA en estas plantas parece ser indispensable para mantener estos procesos activos. El incremento en la concentración de los AA aromáticos, BCAAs, Lys y Pro en los fenotipos albino y variegado, los cuales son marcadores para procesos catabólicos de AA típicos de la autofagia, podrían ser evidencia para su posible activación para la producción alternativa de energía (Izumi y Ishida, 2019; Hirota *et al.*, 2018; Hildebrandt *et al.*, 2015).

5.2. CONCLUSIONES GENERALES

- Las plántulas albinas de *Agave* exhiben una reprogramación transcripcional caracterizada por una fuerte activación en la expresión de genes vinculados con la división del plástido, el complejo TOC/TIC de importación de proteínas, la regulación transcripcional y traduccional en el plástido, la biosíntesis de los tetrapirroles, la fotomorfogénesis, el metabolismo fotosintético y la fijación de carbono, entre otros.

- Los perfiles de expresión de los diversos grupos de genes analizados revelan que los tejidos foliares albinos presentan plástidos que se conservan en un estadio de desarrollo juvenil similar al de un proplástido.

- El número de copias de los genomas plastídicos es mayor en los tejidos albinos respecto a su contraparte verde.

- La alta expresión de los genes que codifican para las enzimas NEP y PEP y subunidades del ribosoma plastídico en el *Agave* albino, sugiere que estos elementos juegan papeles clave para entender las adaptaciones de la plántula albina al bloqueo en la biogénesis del cloroplasto.

- Se identificó un total de 1,581 transcritos *PPR* en el transcriptoma de las variantes somaclonales verde y albino de *A. angustifolia*, de los cuales 282 estuvieron asociados al cloroplasto. La expresión de estos genes depende tanto del estadio de diferenciación celular como el del plástido.

- Los tejidos albinos, particularmente el tejido foliar albino, presentaron una mayor actividad transcripcional de los genes *PPR* cloroplásticos. Este incremento puede estar provocado por las alteraciones en la comunicación entre el núcleo y cloroplasto durante la biogénesis del plástido.

- El incremento en la expresión de los genes *PPR* cloroplásticos, así como el de otros genes claves de la biogénesis del cloroplasto, podría formar parte de una respuesta compensatoria correctiva que intenta remover al proplástido de su estadio juvenil y revertir el fenotipo albino. Esta respuesta correctiva por parte del proplástido, podría estar ligada a la adaptación y la supervivencia de esta planta heterótrofa.

- La identificación de los ARN blanco de las proteínas hipotéticas codificadas por doce genes *PPR* seleccionados para su validación por qRT-PCR, revelan su posible actuación en los ARN de genes ribosomales, de las subunidades de la ATP sintasa, de la PEP, de la RuBisCO y de los fotosistemas.

- Las plántulas de *Agave* variegadas y albinas acumulan altos niveles de AA libres, principalmente aquellos sintetizados parcial o totalmente en el cloroplasto, destacando BCAA, AA aromáticos, sí como Pro, Lys, Met y Asp.

- El aumento en la concentración de AA en plántulas de *Agave* variegadas y albinas podría estar relacionado con una mayor acumulación y actividad de las enzimas AaBCAT, AaALS, AaIPMS, AaSHMT, AaAST, AaTS, AaCGS, AaMS y AaCM. Estas enzimas podrían jugar papeles fundamentales en la adaptación al ambiente y mantenimiento de la homeostasis metabólica.

- Si bien, para los casos de las enzimas AaBCAT y AaCM se encontró una integración entre su acumulación a nivel de ARNm y de proteínas, la existencia de discrepancias para otras enzimas a estos dos niveles, revela posibles mecanismos de regulación desconocidos que podrían estar actuando en nuestro modelo de estudio.

- La reprogramación en el metabolismo de los AA juega un papel clave en la respuesta de las plantas de *Agave* albinas y variegadas que son inactivas fotosintéticamente.

5.3. PERSPECTIVAS

- Profundizar y confirmar si las diferentes funciones que a nivel transcriptómico parecen ser claves en la superviviencia de las plántulas de *Agave* albina, están realmente activas a otros niveles. Entre ellos podemos mencionar la división del plástido, el complejo TOC/TIC de importación de proteínas, regulación de la fotomorfogénesis, entre otros.

- Determinar que mecanismos de regulación epigenéticos, postranscripcionales, traduccionales o postraduccionales podrían estar actuando en las plántulas de *Agave* albinas y variegadas.

- Continuar con el estudio de la actividad de la NEP y de la PEP así como de los ribosomas cloroplásticos en el *Agave* albino.

- Secuenciar el plastoma de la plántula de *Agave* albino con el objetivo de identificar los cambios que puede alojar esta variante somaclonal respecto a su contraparte verde.

- Determinar el tipo de mecanismo de regulación postranscripcional ejercido por las proteínas codificadas por los genes *PPR* seleccionados en este estudio.

- Profundizar en el estudio del modelo de respuesta retro-anterógrada propuesto para el *Agave* albino, identificando sus similitudes y particularidades respecto a lo descrito en otros modelos vegetales.

- Debido al papel de las proteínas PPR en el metabolismo del ARN, es necesario continuar su caracterización con el fin de profundizar y madurar la idea de su uso como una herramienta biotecnológica en la edición del ARN del cloroplasto.

- Profundizar en el impacto del aumento en la concentración de los AA sobre otras rutas metabólicas que tienen a estos elementos como precursores. De especial interes es su impacto en las rutas de flavonoides, de melatonina y serotonina, síntesis de SAM y derivados, entre otros.

- Evaluar si el aumento en la concentración de AA en las plántulas de *Agave* albinas y variegadas responde también a la activación de rutas catabólicas de AA.

- Determinar la activación de los mecanismos de autofagia, particularmente clorofagia, en las plántulas de *Agave* albinas y variegadas como alternativa para la producción de energía.

- Evaluar el efecto de ciertos inhibidores en las rutas de biosíntesis y catabolismo de AA en las plántulas de *Agave* albina.

- Explorar la posibilidad de explotar biotecnológicamente el alto contenido de AA en las plántulas de *Agave* albinas y variegadas.

BIBLIOGRAFÍA

Abdallah, F., Salamini, F., Leister, D. (2000). A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. *Trends in Plant Science*, 5, 141-142.

Adams, J. (2008). The Proteome: Discovering the Structure and Function of Proteins. *Nature Education*, 1, 6.

Ahmad, P., Abdel Latef, A. A. H., Rasool, S., Akram, N. A., Ashraf, M., Gucel, S. (2016). Role of Proteomics in Crop Stress Tolerance. *Frontiers in Plant Science*, 7.

Ahsan, N., Nanjo, Y., Sawada, H., Kohno, Y., Komatsu, S. (2010). Ozone stress-induced proteomic changes in leaf total soluble and chloroplast proteins of soybean reveal that carbon allocation is involved in adaptation in the early developmental stage. *PROTEOMICS*, 10, 2605-2619.

Allen, J. F. (2018). Translating photosynthesis. *Nature Plants*, 4, 199-200.

Almagro-Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life science alliance*, 2, e201900429.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.

Álvarez-Chávez, J., Villamiel, M., Santos-Zea, L., Ramírez-Jiménez, A. K. (2021). Agave By-Products: An Overview of Their Nutraceutical Value, Current Applications, and Processing Methods. *Polysaccharides*, 2, 720-743.

Amir, R. (2010). Current understanding of the factors regulating methionine content in vegetative tissues of higher plants. *Amino Acids*, 39, 917-931.

An, H., Ke, X., Li, L., Liu, Y., Yuan, S., Wang, Q., Hou, X., Zhao, J. (2023). ALBINO EMBRYO AND SEEDLING is required for RNA splicing and chloroplast homeostasis in Arabidopsis. *Plant Physiology*, 193, 483-501.

Andrade-Marcial, M., Pacheco-Arjona, R., Góngora-Castillo, E., De-la-Peña, C. (2022). Chloroplastic pentatricopeptide repeat proteins (PPR) in albino plantlets of Agave angustifolia Haw. reveal unexpected behavior. *BMC Plant Biology*, 22, 352.

Andrade-Marcial, M., Ruíz-May, E., Elizalde-Contreras, J. M., Pacheco, N., Herrera-Pool, E., De-la-Peña, C. (2023). Proteome of Agave angustifolia Haw.: Uncovering metabolic alterations, over-accumulation of amino acids, and compensatory pathways in chloroplast-deficient albino plantlets. *Plant Physiology and Biochemistry*, 201, 107902.

Araújo, W. L., Ishizaki, K., Nunes-Nesi, A., Larson, T. R., Tohge, T., Krahnert, I., Witt, S., Obata, T., Schauer, N., Graham, I. A., Leaver, C. J., Fernie, A. R. (2010). Identification of the 2-Hydroxyglutarate and Isovaleryl-CoA Dehydrogenases as Alternative Electron Donors Linking Lysine Catabolism to the Electron Transport Chain of Arabidopsis Mitochondria. *Plant Cell*, 22, 1549-1563.

Archibald, J. M. (2009). The Puzzle of Plastid Evolution. *Current Biology*, 19, R81-R88.

Archibald, J. M. (2015). Endosymbiosis and Eukaryotic Cell Evolution. *Current Biology*, 25, R911-R921.

Armarego-Marriott, T., Kowalewska, Ł., Burgos, A., Fischer, A., Thiele, W., Erban, A., Strand, D., Kahlau, S., Hertle, A., Kopka, J., Walther, D., Reich, Z., Schöttler, M. A., Bock, R. (2019). Highly Resolved Systems Biology to Dissect the Etioplast-to-Chloroplast Transition in Tobacco Leaves. *Plant Physiology*, 180, 654-681.

Armbruster, U. y Strand, D. D. (2020). Regulation of chloroplast primary metabolism. *Photosynthesis Research*, 145, 1-3.

Arzate-Fernández, A., Piña-Escutia, J., Norman-Mondragón, T., Reyes-Díaz, J., Guevara-Suárez, K., Vásquez-García, L. (2016). Regeneraión de Agave mezcalero (*Agave angustifolia* Haw.) a partir de embriones somáticos encapsulados. *Revista Fitotecnia Mexicana*, 39, 359-366.

Astolfi, S., Celletti, S., Vigani, G., Mimmo, T., Cesco, S. (2021). Interaction Between Sulfur and Iron in Plants. *Frontiers in Plant Science*, 12.

Aswani, V., Rajsheel, P., Bapatla, R. B., Sunil, B., Raghavendra, A. S. (2019). Oxidative stress induced in chloroplasts or mitochondria promotes proline accumulation in leaves of pea (*Pisum sativum*): another example of chloroplast-mitochondria interactions. *Protoplasma*, 256, 449-457.

Avin-Wittenberg, T., Bajdzienko, K., Wittenberg, G., Alseekh, S., Tohge, T., Bock, R., Giavalisco, P., Fernie, A. R. (2015). Global Analysis of the Role of Autophagy in Cellular Metabolism and Energy Homeostasis in Arabidopsis Seedlings under Carbon Starvation. *Plant Cell*, 27, 306-322.

Bailey, T. L., Johnson, J., Grant, C. E., Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43, W39-W49.

Bairu, M. W., Aremu, A. O., Van Staden, J. (2011). Somaclonal variation in plants: causes and detection methods. *Plant Growth Regulation*, 63, 147-173.

Barkan, A. y Small, I. (2014). Pentatricopeptide Repeat Proteins in Plants. *Annual Review of Plant Biology*, 65, 415-442.

Barrientos-Rivera, g., Esparza-Ibarra, E., Segura-Pacheco.Héctor, Talavera-Mendoza, Ó., Sampedro-Rosas, M., Hernández-Castro, E. (2019). Caracterización morfológica de *Agave angustifolia* y su conservación en Guerrero, México. *Revista Mexicana de Ciencias Agrícolas*, 10, 655-668.

Bayat, A. (2002). Bioinformatics. *BMJ*, 324, 1018-1022.

Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., Barkan, A. (2008). The Pentatricopeptide Repeat Protein PPR5 Stabilizes a Specific tRNA Precursor in Maize Chloroplasts. *Molecular and Cellular Biology*, 28, 5337-5347.

Benešová, M., Holá, D., Fischer, L., Jedelský, P. L., Hnilička, F., Wilhelmová, N., Rothová, O., Kočová, M., Procházková, D., Honnerová, J., Fridrichová, L., Hniličková, H. (2012). The Physiology and Proteomics of Drought Tolerance in Maize: Early Stomatal Closure as a Cause of Lower Tolerance to Short-Term Dehydration? *PLOS ONE*, 7, e38017.

Berry, J. O., Yerramsetty, P., Zielinski, A. M., Mure, C. M. (2013). Photosynthetic gene expression in higher plants. *Photosynthesis Research*, 117, 91-120.

Bock, R. (2017). Witnessing Genome Evolution: Experimental Reconstruction of Endosymbiotic and Horizontal Gene Transfer. *Annual Review of Genetics*, 51, 1-22.

Bohne, A. V., Schwenkert, S., Grimm, B., Nickelsen, J. (2016). Roles of Tetratricopeptide Repeat Proteins in Biogenesis of the Photosynthetic Apparatus. *International Review of Cell and Molecular Biology*, 324, 187-227.

Bölter, B. (2018). En route into chloroplasts: preproteins' way home. *Photosynthesis Research*, 138, 263-275.

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

Boussardon, C., Salone, V., Avon, A., Berthomé, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., Lurin, C. (2012). Two Interacting Proteins Are Necessary for the Editing of the NdhD-1 Site in Arabidopsis Plastids. *The Plant Cell*, 24, 3684-3694.

Cackett, L., Luginbuehl, L. H., Schreier, T. B., Lopez-Juez, E., Hibberd, J. M. (2022). Chloroplast development in green plant tissues: the interplay between light, hormone, and transcriptional regulation. *New Phytologist*, 233, 2000-2016.

Canonge, J., Roby, C., Hamon, C., Potin, P., Pfannschmidt, T., Philippot, M. (2021). Occurrence of albinism during wheat androgenesis is correlated with repression of the key genes required for proper chloroplast biogenesis. *Planta*, 254, 123.

Caredda, S., Doncoeur, C., Devaux, P., Sangwan, R. S., Clément, C. (2000). Plastid differentiation during androgenesis in albino and non-albino producing cultivars of barley (*Hordeum vulgare* L.). *Sexual Plant Reproduction*, 13, 95-104.

Chang, S. H., Lee, S., Um, T. Y., Kim, J.-K., Do Choi, Y., Jang, G. (2017). pTAC10, a Key Subunit of Plastid-Encoded RNA Polymerase, Promotes Chloroplast Development. *Plant Physiology*, 174, 435-449.

Chen, G., Zou, Y., Hu, J., Ding, Y. (2018a). Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC Genomics*, 19, 720.

Chen, J. y Henny, R. J. (2006). Somaclonal variation: an important source for cultivar development of floricultural crops, en: Floriculture, Ornamental and Plant Biotechnology, II, Da Silva, J. A. T. (Ed.). Global Science Books, London, UK, pp. 244-253.

Chen, L., Huang, Y., Xu, M., Cheng, Z., Zheng, J. (2017). Proteomic Analysis Reveals Coordinated Regulation of Anthocyanin Biosynthesis through Signal Transduction and Sugar Metabolism in Black Rice Leaf. *International Journal of Molecular Sciences*, 18, 2722.

Chen, Y., Chen, X., Hu, F., Yang, H., Yue, L., Trigiano, R. N., Cheng, Z.-M. (2014). Micropropagation of *Agave americana*. *HortScience*, 49, 320-327.

Chen, Y., Zhou, B., Li, J., Tang, H., Tang, J., Yang, Z. (2018b). Formation and Change of Chloroplast-Located Plant Metabolites in Response to Light Conditions. *International Journal of Molecular Sciences*, 19, 654.

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., Castleden, I., Song, Y., Song, B., Huang, J., Liu, X., Xu, X., Lim, B. L., Bond, C. S., Yiu, S.-M., Small, I. (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *The Plant Journal*, 85, 532-547.

Chi, W., Feng, P., Ma, J., Zhang, L. (2015a). Metabolites and chloroplast retrograde signaling. *Current Opinion in Plant Biology*, 25, 32-38.

Chi, W., He, B., Mao, J., Jiang, J., Zhang, L. (2015b). Plastid sigma factors: Their individual functions and regulation in transcription. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1847, 770-778.

Clarke, H. J., Kumari, M., Khan, T. N., Siddique, K. H. M. (2011). Poorly formed chloroplasts are barriers to successful interspecific hybridization in chickpea following in vitro embryo rescue. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 106, 465-473.

Corpas, F. J., Gupta, D. K., Palma, J. M. (2021). Tryptophan: A Precursor of Signaling Molecules in Higher Plants, en: Hormones and Plant Response, Gupta, D. K. y Corpas, F. J. (Eds.). Springer, Cham, pp. 273-289.

Cushing, D. A., Forsthoefel, N. R., Gestaut, D. R., Vernon, D. M. (2005). Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta*, 221, 424-436.

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

Das, T. (1992). Micropropagation of Agave sisalana. *Plant Cell, Tissue and Organ Culture*, 31, 253-255.

Davidson, N. M. y Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novoassembled transcriptomes. *Genome Biology*, 15, 410.

Davis, S. C., Simpson, J., Gil-Vega, K. d. C., Niechayev, N. A., Tongerlo, E. v., Castano, N. H., Dever, L. V., Búrquez, A. (2019). Undervalued potential of crassulacean acid metabolism for current and future agricultural production. *Journal of Experimental Botany*, 70, 6521-6537.

de Luna-Valdez, L. A., Martínez-Batallar, A. G., Hernández-Ortiz, M., Encarnación-Guevara, S., Ramos-Vega, M., López-Bucio, J. S., León, P., Guevara-García, A. A. (2014). Proteomic analysis of chloroplast biogenesis (clb) mutants uncovers novel proteins potentially involved in the development of Arabidopsis thaliana chloroplasts. *Journal of Proteomics*, 111, 148-164.

De-la-Peña, C., León, P., Sharkey, T. D. (2022). Editorial: Chloroplast Biotechnology for Crop Improvement. *Frontiers in Plant Science*, 13.

Delgado-Aceves, L., Portillo, L., Folgado, R., de Jesús Romo-Paz, F., González-Arnao, M. T. (2022). New approaches for micropropagation and cryopreservation of Agave peacockii, an endangered species. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 150, 85-95.

Dewir, Y. H., Nurmansyah, Naidoo, Y., Teixeira da Silva, J. A. (2018). Thidiazuron-induced abnormalities in plant tissue cultures. *Plant Cell Reports*, 37, 1451-1470.

Dobrogojski, J., Adamiec, M., Luciński, R. (2020). The chloroplast genome: a review. *Acta Physiologiae Plantarum*, 42, Article 98.

Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., Mechtler, K. (2014). MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *Journal of Proteome Research*, 13, 3679-3684.

Duarte-Aké, F., Castillo-Castro, E., Pool, F. B., Espadas, F., Santamaría, J. M., Robert, M. L., De-la-Peña, C. (2016). Physiological differences and changes in global DNA methylation levels in Agave angustifolia Haw. albino variant somaclones during the micropropagation process. *Plant Cell Reports*, 35, 2489-2502.

Duta-Cornescu, G., Constantin, N., Pojoga, D.-M., Nicuta, D., Simon-Gruita, A. (2023). Somaclonal Variation—Advantage or Disadvantage in Micropropagation of the Medicinal Plants. *International Journal of Molecular Sciences*, 24, 838.

Ebihara, T., Matsuda, T., Sugita, C., Ichinose, M., Yamamoto, H., Shikanai, T., Sugita, M. (2019). The P-class pentatricopeptide repeat protein PpPPR_21 is needed for accumulation of the psbI-ycf12 dicistronic mRNA in Physcomitrella chloroplasts. *The Plant Journal*, 97, 1120-1131.

Echevarría-Machado, I., Sánchez-Cach, L. A., Hernández-Zepeda, C., Rivera-Madrid, R., Moreno-Valenzuela, O. A. (2005). A simple and efficient method for isolation of DNA in high mucilaginous plant tissues. *Molecular Biotechnology*, 31, 129-135.

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLOS Computational Biology*, 7, e1002195.

Eguiarte, L. E., Jiménez Barrón, O. A., Aguirre-Planter, E., Scheinvar, E., Gámez, N., Gasca-Pineda, J., Castellanos-Morales, G., Moreno-Letelier, A., Souza, V. (2021). Evolutionary ecology of Agave: distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). *American Journal of Botany*, 108, 216-235.

Eng, J. K., McCormack, A. L., Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5, 976-989.

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y., Luo, K. (2015). Efficient CRISPR/Cas9-mediated Targeted Mutagenesis in Populus in the First Generation. *Scientific Reports*, 5, 12217.

Ferreira, M. d. S., Rocha, A. d. J., Nascimento, F. d. S., Oliveira, W. D. d. S., Soares, J. M. d. S., Rebouças, T. A., Morais Lino, L. S., Haddad, F., Ferreira, C. F., Santos-Serejo, J. A. d., Fernández, J. S., Amorim, E. P. (2023). The Role of Somaclonal Variation in Plant Genetic Improvement: A Systematic Review. *Agronomy*, 13, 730.

Fey, V., Wagner, R., Bräutigam, K., Pfannschmidt, T. (2005). Photosynthetic redox control of nuclear gene expression. *Journal of Experimental Botany*, 56, 1491-1498.

Filiz, E., Ozyigit, I. I., Saracoglu, I. A., Uras, M. E., Sen, U., Yalcin, B. (2019). Abiotic stress-induced regulation of antioxidant genes in different Arabidopsis ecotypes: microarray data evaluation. *Biotechnology & Biotechnological Equipment*, 33, 128-143.

Flores-Abreu, I. N., Trejo-Salazar, R. E., Sánchez-Reyes, L. L., Good, S. V., Magallón, S., García-Mendoza, A., Eguiarte, L. E. (2019). Tempo and mode in coevolution of Agave sensu lato (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae (Phyllostomidae). *Molecular Phylogenetics and Evolution*, 133, 176-188.

Floris, D. y Kühlbrandt, W. (2021). Molecular landscape of etioplast inner membranes in higher plants. *Nature Plants*, 7, 514-523.

Forlani, G., Trovato, M., Funck, D., Signorelli, S. (2019). Regulation of Proline Accumulation and Its Molecular and Physiological Functions in Stress Defence, en: Osmoprotectant-Mediated Abiotic Stress Tolerance in Plants: Recent Advances and Future Perspectives, Hossain, M. A., Kumar, V., Burritt, D. J., Fujita, M., y Mäkelä, P. S. A. (Eds.). Springer, Cham, pp. 73-97.

Fucile, G., Falconer, S., Christendat, D. (2008). Evolutionary Diversification of Plant Shikimate Kinase Gene Duplicates. *PLOS Genetics*, 4, e1000292.

Fujii, S. y Small, I. (2011). The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytologist*, 191, 37-47.

Füssy, Z. y Oborník, M. (2018). Complex Endosymbioses I: From Primary to Complex Plastids, Multiple Independent Events, en: Plastids: Methods and Protocols, Maréchal, E. (Ed.). Springer US, New York, NY, pp. 17-35.

Gao, H., Kadirjan-Kalbach, D., Froehlich, J. E., Osteryoung, K. W. (2003). ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proceedings of the National Academy of Sciences*, 100, 4328-4333.

Gao, H., Sage, T. L., Osteryoung, K. W. (2006). FZL, an FZO-like protein in plants, is a determinant of thylakoid and chloroplast morphology. *Proceedings of the National Academy of Sciences*, 103, 6759-6764.

Gao, Z.-P., Yu, Q.-B., Zhao, T.-T., Ma, Q., Chen, G.-X., 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.

Garber, M., Grabherr, M. G., Guttman, M., Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*, 8, 469-477.

García, A., Aguado, E., Parra, G., Manzano, S., Martínez, C., Megías, Z., Cebrián, G., Romero, J., Beltrán, S., Garrido, D., Jamilena, M. (2018). Phenomic and Genomic Characterization of a Mutant Platform in Cucurbita pepo. *Frontiers in Plant Science*, 9.

García-Alcázar, M., Giménez, E., Pineda, B., Capel, C., García-Sogo, B., Sánchez, S., Yuste-Lisbona, F. J., Angosto, T., Capel, J., Moreno, V., Lozano, R. (2017). Albino T-DNA tomato mutant reveals a key function of 1-deoxy-D-xylulose-5-phosphate synthase (DXS1) in plant development and survival. *Scientific Reports*, 7, 45333.

Ghosh, U. K., Islam, M. N., Siddiqui, M. N., Cao, X., Khan, M. A. R. (2022). Proline, a multifaceted signalling molecule in plant responses to abiotic stress: understanding the physiological mechanisms. *Plant Biology*, 24, 227-239.

Głowacka, K., Kromdijk, J., Salesse-Smith, C. E., Smith, C., Driever, S. M., Long, S. P. (2023). Is chloroplast size optimal for photosynthetic efficiency? *New Phytologist*, 239, 2197-2211.

Gomez-Casati, D. F., Barchiesi, J., Busi, M. V. (2022). Mitochondria and chloroplasts function in microalgae energy production. *PeerJ*, 10, e14576.

Good-Avila, S. V., Souza, V., Gaut, B. S., Eguiarte, L. E. (2006). Timing and rate of speciation in Agave (Agavaceae). *Proceedings of the National Academy of Sciences*, 103, 9124-9129.

Goodwin, S., McPherson, J. D., McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17, 333-351.

Gothandam, K. M., Kim, E.-S., Cho, H., Chung, Y.-Y. (2005). OsPPR1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis. *Plant Molecular Biology*, 58, 421-433.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36, 3420-3435.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644-652.

Gu, Z., Eils, R., Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32, 2847-2849.

Gutmann, B., Royan, S., Schallenberg-Rüdinger, M., Lenz, H., Castleden, I. R., McDowell, R., Vacher, M. A., Tonti-Filippini, J., Bond, C. S., Knoop, V., Small, I. D. (2020). The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Molecular Plant*, 13, 215-230.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494-1512.

Hammani, K., Okuda, K., Tanz, S. K., Chateigner-Boutin, A.-L., Shikanai, T., Small, I. (2009). A Study of New Arabidopsis Chloroplast RNA Editing Mutants Reveals General Features of Editing Factors and Their Target Sites. *The Plant Cell*, 21, 3686-3699.

Hammani, K., Takenaka, M., Miranda, R., Barkan, A. (2016). A PPR protein in the PLS subfamily stabilizes the 5′-end of processed rpl16 mRNAs in maize chloroplasts. *Nucleic Acids Research*, 44, 4278-4288.

Han, M., Zhang, C., Suglo, P., Sun, S., Wang, M., Su, T. (2021). L-Aspartate: An Essential Metabolite for Plant Growth and Stress Acclimation. *Molecules*, 26, 1887.

Han, Z., Qin, Y., Li, X., Yu, J., Li, R., Xing, C., Song, M., Wu, J., Zhang, J. (2020). A genome-wide analysis of pentatricopeptide repeat (PPR) protein-encoding genes in four Gossypium species with an emphasis on their expression in floral buds, ovules, and fibers in upland cotton. *Molecular Genetics and Genomics*, 295, 55-66.

He, G. (2022). Engineering chloroplasts for insect pest control. *Proceedings of the National Academy of Sciences*, 119, e2205125119.

Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16, 169.

Heinemann, B. y Hildebrandt, T. M. (2021). The role of amino acid metabolism in signaling and metabolic adaptation to stress-induced energy deficiency in plants. *Journal of Experimental Botany*, 72, 4634-4645.

Hellmann, H., Funck, D., Rentsch, D., Frommer, W. B. (2000). Hypersensitivity of an Arabidopsis Sugar Signaling Mutant toward Exogenous Proline Application. *Plant Physiology*, 123, 779-789.

Hernández-Castellano, S., Garruña-Hernández, R., Rosa, U.-C., Rosa, Kú-Gonzalez, Á., De-la-Peña, C. (2020). Agave angustifolia albino plantlets lose stomatal physiology function by changing the development of the stomatal complex due to a molecular disruption. *Mol. Genet. Genom.*, 295, 787-805.

Herrera-Pool, E., Ramos-Díaz, A. L., Lizardi-Jiménez, M. A., Pech-Cohuo, S., Ayora-Talavera, T., Cuevas-Bernardino, J. C., García-Cruz, U., Pacheco, N. (2021). Effect of solvent polarity on the Ultrasound Assisted extraction and antioxidant activity of phenolic compounds from habanero pepper leaves (Capsicum chinense) and its identification by UPLC-PDA-ESI-MS/MS. *Ultrasonics Sonochemistry*, 76, 105658.

Hicks, J. L., Lassadi, I., Carpenter, E. F., Eno, M., Vardakis, A., Waller, R. F., Howe, C. J., Nisbet, R. E. R. (2019). An essential pentatricopeptide repeat protein in the apicomplexan remnant chloroplast. *Cellular Microbiology*, 21, e13108.

Higashi, H., Kato, Y., Fujita, T., Iwasaki, S., Nakamura, M., Nishimura, Y., Takenaka, M., Shikanai, T. (2021). The Pentatricopeptide Repeat Protein PGR3 Is Required for the Translation of petL and ndhG by Binding Their 5′ UTRs. *Plant and Cell Physiology*, 62, 1146-1155.

Hildebrandt, T. M. (2018). Synthesis versus degradation: directions of amino acid metabolism during Arabidopsis abiotic stress response. *Plant Molecular Biology*, 98, 121-135.

Hildebrandt, T. M., Nunes Nesi, A., Araújo, W. L., Braun, H.-P. (2015). Amino Acid Catabolism in Plants. *Molecular Plant*, 8, 1563-1579.

Hirota, T., Izumi, M., Wada, S., Makino, A., Ishida, H. (2018). Vacuolar Protein Degradation via Autophagy Provides Substrates to Amino Acid Catabolic Pathways as an Adaptive Response to Sugar Starvation in Arabidopsis thaliana. *Plant and Cell Physiology*, 59, 1363-1376.

Hölzer, M. y Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, 8.

Hooghvorst, I., López-Cristoffanini, C., Nogués, S. (2019). Efficient knockout of phytoene desaturase gene using CRISPR/Cas9 in melon. *Scientific Reports*, 9, 17077.

Hou, D.-Y., Xu, H., Du, G.-Y., Lin, J.-T., Duan, M., Guo, A.-G. (2009). Proteome analysis of chloroplast proteins in stage albinism line of winter wheat (Triticum aestivum) FA85. *BMB Reports*, 42, 450-455.

Huang, W., Zhang, Y., Shen, L., Fang, Q., Liu, Q., Gong, C., Zhang, C., Zhou, Y., Mao, C., Zhu, Y., Zhang, J., Chen, H., Zhang, Y., Lin, Y., Bock, R., Zhou, F. (2020). Accumulation of the RNA polymerase subunit RpoB depends on RNA editing by OsPPR16 and affects chloroplast development during early leaf development in rice. *New Phytologist*, 228, 1401-1416.

Huang, X., Xiao, M., Xi, J., He, C., Zheng, J., Chen, H., Gao, J., Zhang, S., Wu, W., Liang, Y., Xie, L., Yi, K. (2019). De Novo Transcriptome Assembly of Agave H11648 by Illumina Sequencing and Identification of Cellulose Synthase Genes in Agave Species. *Genes*, 10, 103.

Huang, X., Xu, B., Tan, S., Huang, Y., Xi, J., Qin, X., Chen, T., Chen, H., Yang, X., Yi, K. (2022). Transcriptome Sequencing of Agave angustifolia Reveals Conservation and Diversification in the Expression of Cinnamyl Alcohol Dehydrogenase Genes in Agave Species. *Agriculture*, 12, 1003.

Hwang, Y., Han, S., Yoo, C. Y., Hong, L., You, C., Le, B. H., Shi, H., Zhong, S., Hoecker, U., Chen, X., Chen, M. (2022). Anterograde signaling controls plastid transcription via sigma factors separately from nuclear photosynthesis genes. *Nature Communications*, 13, 7440.

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

Izumi, M. y Ishida, H. (2019). An additional role for chloroplast proteins—an amino acid reservoir for energy production during sugar starvation. *Plant Signaling & Behavior*, 14, 1552057.

Jan, M., Liu, Z., Rochaix, J.-D., Sun, X. (2022). Retrograde and anterograde signaling in the crosstalk between chloroplast and nucleus. *Frontiers in Plant Science*, 13.

Jarvis, P. y López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nature Reviews Molecular Cell Biology*, 14, 787-802.

Jez, J. M. (2019). Structural biology of plant sulfur metabolism: from sulfate to glutathione. *Journal of Experimental Botany*, 70, 4089-4103.

Jiang, J. y Dehesh, K. (2021). Plastidial retrograde modulation of light and hormonal signaling: an odyssey. *New Phytologist*, 230, 931-937.

Jiang, X., Zhao, H., Guo, F., Shi, X., Ye, C., Yang, P., Liu, B., Ni, D. (2020). Transcriptomic analysis reveals mechanism of light-sensitive albinism in tea plant Camellia sinensis 'Huangjinju'. *BMC Plant Biology*, 20, Article 216.

Jin, H., Fu, M., Duan, Z., Duan, S., Li, M., Dong, X., Liu, B., Feng, D., Wang, J., Peng, L., Wang, H.-B. (2018). LOW PHOTOSYNTHETIC EFFICIENCY 1 is required for light-regulated photosystem II biogenesis in Arabidopsis. *Proceedings of the National Academy of Sciences*, 115, E6075-E6084.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236-1240.

Jorrin-Novo, J. V. (2021). Proteomics and plant biology: contributions to date and a look towards the next decade. *Expert Review of Proteomics*, 18, 93-103.

Juarez-Escobar, J., Guerrero-Analco, J. A., Zamora-Briseño, J. A., Elizalde-Contreras, J. M., Bautista-Valle, M. V., Bojórquez-Velázquez, E., Loyola-Vargas, V. M., Mata-Rosas, M., Ruíz-May, E. (2021). Tissue-specific proteome characterization of avocado seed during postharvest shelf life. *Journal of Proteomics*, 235, 104112.

Julou, T., Burghardt, B., Gebauer, G., Berveiller, D., Damesin, C., Selosse, M.-A. (2005). Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of Cephalanthera damasonium. *New Phytologist*, 166, 639-653.

Jung, H.-S. y Chory, J. (2009). 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.

Kaiser, E., Correa-Galvis, V., Armbruster, U. (2019). Efficient photosynthesis in dynamic light environments: a chloroplast's perspective. *Biochemical Journal*, 476, 2725-2741.

Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods*, 4, 923-925.

Karim, R., Nuruzzaman, M., Khalid, N., Harikrishna, J. A. (2016). Importance of DNA and histone methylation in in vitro plant propagation for crop improvement: a review. *Annals of Applied Biology*, 169, 1-16.

Karpenahalli, M. R., Lupas, A. N., Söding, J. (2007). TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics*, 8, Article 2.

Katoh, K., Rozewicki, J., Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20, 1160-1166.

Kaur, A., Malhotra, P. K., Manchanda, P., Gosal, S. S. (2018). Micropropagation and Somatic Embryogenesis in Sugarcane, en: Biotechnologies of Crop Improvement, Volume 1: Cellular Approaches, Gosal, S. S. y Wani, S. H. (Eds.). Springer International Publishing, Cham, pp. 57-91.

Kirchhoff, H. (2019). Chloroplast ultrastructure in plants. *New Phytologist*, 223, 565-574.

Knill, T., Schuster, J., Reichelt, M., Gershenzon, J., Binder, S. (2008). Arabidopsis Branched-Chain Aminotransferase 3 Functions in Both Amino Acid and Glucosinolate Biosynthesis. *Plant Physiology*, 146, 1028-1039.

Kosová, K., Vítámvás, P., Urban, M. O., Prášil, I. T., Renaut, J. (2018). Plant Abiotic Stress Proteomics: The Major Factors Determining Alterations in Cellular Proteome. *Frontiers in Plant Science*, 9.

Krishna, H., Alizadeh, M., Singh, D., Singh, U., Chauhan, N., Eftekhari, M., Sadh, R. K. (2016). Somaclonal variations and their applications in horticultural crops improvement. *3 Biotech*, 6, 54.

Kumar, A., Prasad, A., Sedlářová, M., Ksas, B., Havaux, M., Pospíšil, P. (2020). Interplay between antioxidants in response to photooxidative stress in Arabidopsis. *Free Radical Biology and Medicine*, 160, 894-907.

Kumari, M., Clarke, H. J., Small, I., Siddique, K. H. M. (2009). Albinism in Plants: A Major Bottleneck in Wide Hybridization, Androgenesis and Doubled Haploid Culture. *Critical Reviews in Plant Sciences*, 28, 393-409.

Lee, J. H., Kim, Y.-C., Jung, Y., Han, J. H., Zhang, C., Yun, C.-W., Lee, S. (2019a). The overexpression of cucumber (Cucumis sativus L.) genes that encode the branched-chain amino acid transferase modulate flowering time in Arabidopsis thaliana. *Plant Cell Reports*, 38, 25-35.

Lee, K. y Kang, H. (2020). Roles of Organellar RNA-Binding Proteins in Plant Growth, Development, and Abiotic Stress Responses. *International Journal of Molecular Sciences*, 21, Article 4548.

Lee, K. y Kang, H. (2023). Engineering of pentatricopeptide repeat proteins in organellar gene regulation. *Frontiers in Plant Science*, 14.

Lee, K., Park, S. J., Han, J. H., Jeon, Y., Pai, H.-S., Kang, H. (2019b). A chloroplast-targeted pentatricopeptide repeat protein PPR287 is crucial for chloroplast function and Arabidopsis development. *BMC Plant Biology*, 19, Article 244.

Lee, Y.-T. y Duggleby, R. G. (2001). Identification of the Regulatory Subunit of Arabidopsis thaliana Acetohydroxyacid Synthase and Reconstitution with Its Catalytic Subunit. *Biochemistry*, 40, 6836-6844.

Legen, J., Dühnen, S., Gauert, A., Götz, M., Schmitz-Linneweber, C. (2023). A CRR2-Dependent sRNA Sequence Supports Papillomavirus Vaccine Expression in Tobacco Chloroplasts. *Metabolites*, 13, 315.

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

Li, B. y Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12, 323.

Li, C.-F., Ma, J.-Q., Huang, D.-J., Ma, C.-L., Jin, J.-Q., Yao, M.-Z., Chen, L. (2018a). Comprehensive Dissection of Metabolic Changes in Albino and Green Tea Cultivars. *Journal of Agricultural and Food Chemistry*, 66, 2040-2048.

Li, H., Cao, H., Yu, R.-p., Miao, Z., Wang, J.-h., Qu, S.-P., Yuan, Q., Li, S.-c. (2018b). De novo transcriptome analysis of an albino mutant Pasphiopedilum pacific shamrock reveals reduced expression of genes related to chloroplast biosynthesis and division. *Horticulture, Environment, and Biotechnology*, 59, 411-421.

Li, J., Cao, X., Jia, X., Liu, L., Cao, H., Qin, W., Li, M. (2021). Iron Deficiency Leads to Chlorosis Through Impacting Chlorophyll Synthesis and Nitrogen Metabolism in Areca catechu L. *Frontiers in Plant Science*, 12.

Li, M., Hensel, G., Mascher, M., Melzer, M., Budhagatapalli, N., Rutten, T., Himmelbach, A., Beier, S., Korzun, V., Kumlehn, J., Börner, T., Stein, N. (2019a). Leaf Variegation and Impaired Chloroplast Development Caused by a Truncated CCT Domain Gene in albostrians Barley. *The Plant Cell*, 31, 1430-1445.

Li, N.-N., Lu, J.-L., Li, Q.-S., Zheng, X.-Q., Wang, X.-C., Wang, L., Wang, Y.-C., Ding, C.-Q., Liang, Y.-R., Yang, Y.-J. (2019b). Dissection of Chemical Composition and Associated Gene Expression in the Pigment-Deficient Tea Cultivar 'Xiaoxueya' Reveals an Albino Phenotype and Metabolite Formation. *Frontiers in Plant Science*, 10, 1543.

Li, Q., Gao, Y., Yang, A. (2020). Sulfur Homeostasis in Plants. *International Journal of Molecular Sciences*, 21, 8926.

Li, X., Kanakala, S., He, Y., Zhong, X., Yu, S., Li, R., Sun, L., Ma, J. (2017). Physiological Characterization and Comparative Transcriptome Analysis of White and Green Leaves of Ananas comosus var. bracteatus. *PLOS ONE*, 12, e0169838.

Liang, Z., Zhu, N., Mai, K. K., Liu, Z., Tzeng, D., Osteryoung, K. W., Zhong, S., Staehelin, L. A., Kang, B.-H. (2018). Thylakoid-Bound Polysomes and a Dynamin-Related Protein, FZL, Mediate Critical Stages of the Linear Chloroplast Biogenesis Program in Greening Arabidopsis Cotyledons. *The Plant Cell*, 30, 1476-1495.

Lin, C.-S., Hsu, C.-T., Yang, L.-H., Lee, L.-Y., Fu, J.-Y., Cheng, Q.-W., Wu, F.-H., Hsiao, H. C.-W., Zhang, Y., Zhang, R., Chang, W.-J., Yu, C.-T., Wang, W., Liao, L.-J., Gelvin, S. B., Shih, M.-C. (2018). Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration. *Plant Biotechnology Journal*, 16, 1295-1310.

Lin, C.-S., Liu, N.-T., Liao, D.-C., Yu, J.-S., Tsao, C.-H., Lin, C.-H., Sun, C.-W., Jane, W.-N., Tsay, H. S., Jian-Wei Chen, J., Lai, E.-M., Lin, N.-S., Chang, W.-C., Lin, C.-C. (2008). Differential Protein Expression of Two Photosystem II Subunits, PsbO and PsbP, in an Albino Mutant of Bambusa edulis with Chloroplast DNA Aberration. *Journal of the American Society for Horticultural Science.*, 133, 270-277.

Lin, D., Gong, X., Jiang, Q., Zheng, K., Zhou, H., Xu, J., Teng, S., Dong, Y. (2015). The rice ALS3 encoding a novel pentatricopeptide repeat protein is required for chloroplast development and seedling growth. *Rice*, 8, 17.

Lin, Z., Xiong, Y., Xue, Y., Mao, M., Xiang, Y., He, Y., Rafique, F., Hu, H., Liu, J., Li, X., Sun, L., Huang, Z., Ma, J. (2019). Screening and characterization of long noncoding RNAs involved in the albinism of Ananas comosus var. bracteatus leaves. *PLOS ONE*, 14, e0225602.

Littlejohn, G. R., Breen, S., Smirnoff, N., Grant, M. (2021). Chloroplast immunity illuminated. *New Phytologist*, 229, 3088-3107.

Liu, D., Li, W., Cheng, J. (2016). The novel protein DELAYED PALE-GREENING1 is required for early chloroplast biogenesis in Arabidopsis thaliana. *Scientific Reports*, 6, 25742.

Liu, M. y Lu, S. (2016). Plastoquinone and Ubiquinone in Plants: Biosynthesis, Physiological Function and Metabolic Engineering. *Frontiers in Plant Science*, 7, 1898.

Liu, N.-T., Jane, W.-N., Tsay, H.-S., Wu, H., Chang, W.-C., Lin, C.-S. (2007). Chloroplast genome aberration in micropropagation-derived albino Bambusa edulis mutants, ab1 and ab2. *Plant Cell, Tissue and Organ Culture*, 88, 147-156.

Liu, X., An, J., Wang, L., Sun, Q., An, C., Wu, B., Hong, C., Wang, X., Dong, S., Guo, J., Feng, Y., Gao, H. (2021). A novel amphiphilic motif at the C-terminus of FtsZ1 facilitates chloroplast division. *The Plant Cell*, 34, 419-432.

Liu, X., Cao, P. H., Huang, Q. Q., Yang, Y. R., Tao, D. D. (2020). Disruption of a Rice Chloroplast-Targeted Gene OsHMBPP Causes a Seedling-Lethal Albino Phenotype. *Rice*, 13, 51.

Livak, K. J. y Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, 25, 402-408.

Lonhienne, T., Cheng, Y., Garcia, M. D., Hu, S. H., Low, Y. S., Schenk, G., Williams, C. M., Guddat, L. W. (2022). Structural basis of resistance to herbicides that target acetohydroxyacid synthase. *Nature Communications*, 13, 3368.

Lonhienne, T., Garcia, M. D., Guddat, L. W. (2017). The Role of a FAD Cofactor in the Regulation of Acetohydroxyacid Synthase by Redox Signaling Molecules. *Journal of Biological Chemistry*, 292, 5101-5109.

Loudya, N., Mishra, P., Takahagi, K., Uehara-Yamaguchi, Y., Inoue, K., Bogre, L., Mochida, K., López-Juez, E. (2021). Cellular and transcriptomic analyses reveal two-staged chloroplast biogenesis underpinning photosynthesis build-up in the wheat leaf. *Genome Biology*, 22, 151.

Loudya, N., Okunola, T., He, J., Jarvis, P., López-Juez, E. (2020). Retrograde signalling in a virescent mutant triggers an anterograde delay of chloroplast biogenesis that requires GUN1 and is essential for survival. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375, 20190400.

Loyola-Vargas, V. M. y Ochoa-Alejo, N. (2018). An Introduction to Plant Tissue Culture: Advances and Perspectives, en: Plant Cell Culture Protocols, Loyola-Vargas, V. M. y Ochoa-Alejo, N. (Eds.). Springer New York, New York, NY, pp. 3-13.

Lu, M., Han, J., Zhu, B., Jia, H., Yang, T., Wang, R., Deng, W.-W., Zhang, Z.-Z. (2019). Significantly increased amino acid accumulation in a novel albino branch of the tea plant (Camellia sinensis). *Planta*, 249, 363-376.

Lurin, C., Andreés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.-L., Mireau, H., Peeters, N., Renou, J.-P., Szurek, B., Taconnat, L., Small, I. (2004). Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. *The Plant Cell*, 16, 2089-2103.

Lv, J., Shang, L., Chen, Y., Han, Y., Yang, X., Xie, S., Bai, W., Hu, M., Wu, H., Lei, K., Yang, Y. n., Ge, S., Trinh, H. P., Zhang, Y., Guo, L., Wang, Z. (2020). OsSLC1 Encodes a Pentatricopeptide Repeat Protein Essential for Early Chloroplast Development and Seedling Survival. *Rice*, 13, Article 25.

Lynch, J. H. y Dudareva, N. (2020). Aromatic Amino Acids: A Complex Network Ripe for Future Exploration. *Trends in Plant Science*, 25, 670-681.

Lynch, J. H., Qian, Y., Guo, L., Maoz, I., Huang, X.-Q., Garcia, A. S., Louie, G., Bowman, M. E., Noel, J. P., Morgan, J. A., Dudareva, N. (2020). Modulation of auxin formation by the cytosolic phenylalanine biosynthetic pathway. *Nature Chemical Biology*, 16, 850-856.

Lyu, J., Wu, Y., Jin, X., Tang, Z., Liao, W., Dawuda, M. M., Hu, L., Xie, J., Yu, J., Calderón-Urrea, A. (2021). Proteomic analysis reveals key proteins involved in ethylene-induced adventitious root development in cucumber (Cucumis sativus L.). *PeerJ*, 9, e10887.

Ma, K., Deng, L., Wu, H., Fan, J. (2022). Towards green biomanufacturing of high-value recombinant proteins using promising cell factory: Chlamydomonas reinhardtii chloroplast. *Bioresources and Bioprocessing*, 9, 83.

Macadlo, L. A., Ibrahim, I. M., Puthiyaveetil, S. (2019). Sigma factor 1 in chloroplast gene transcription and photosynthetic light acclimation. *Journal of Experimental Botany*, 71, 1029-1038.

Maeda, H., Shasany, A. K., Schnepp, J., Orlova, I., Taguchi, G., Cooper, B. R., Rhodes, D., Pichersky, E., Dudareva, N. (2010). RNAi Suppression of Arogenate Dehydratase1 Reveals That Phenylalanine Is Synthesized Predominantly via the Arogenate Pathway in Petunia Petals. *Plant Cell*, 22, 832-849.

Majeran, W., Friso, G., Asakura, Y., Qu, X., Huang, M., Ponnala, L., Watkins, K. P., Barkan, A., van Wijk, K. J. (2011). Nucleoid-Enriched Proteomes in Developing Plastids and Chloroplasts from Maize Leaves: A New Conceptual Framework for Nucleoid Functions. *Plant Physiology*, 158, 156-189.

Majeran, W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., Zhang, C., Asakura, Y., Bhuiyan, N. H., Sun, Q., Turgeon, R., van Wijk, K. J. (2010). Structural and Metabolic Transitions of C4 Leaf Development and Differentiation Defined by Microscopy and Quantitative Proteomics in Maize. *The Plant Cell*, 22, 3509-3542.

Makowska, K. y Oleszczuk, S. (2014). Albinism in barley androgenesis. *Plant cell reports*, 33, 385-392.

Manchanda, P., Kaur, A., Gosal, S. S. (2018). Somaclonal Variation for Sugarcane Improvement, en: Biotechnologies of Crop Improvement, Volume 1: Cellular Approaches, Gosal, S. S. y Wani, S. H. (Eds.). Springer International Publishing, Cham, pp. 299-326.

Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. *Nature Protocols*, 12, 213-218.

Maréchal, E. (2018). Primary Endosymbiosis: Emergence of the Primary Chloroplast and the Chromatophore, Two Independent Events, en: Plastids: Methods and Protocols, Maréchal, E. (Ed.). Springer US, New York, NY, pp. 3-16.

Margulis, L. (1970). Origin of Eukaryotic Cells Evidence and Research Implications for a Theory of the Origin and Evolution of Microbial, Plant, and Animal Cells on the Precambrian Earth. Yale University Press.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., Kowallik, K. V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature*, 393, 162-165.

McAlister, G. C., Nusinow, D. P., Jedrychowski, M. P., Wühr, M., Huttlin, E. L., Erickson, B. K., Rad, R., Haas, W., Gygi, S. P. (2014). MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Analytical Chemistry*, 86, 7150-7158.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49, D412-D419.

Monribot-Villanueva, J. L., Altúzar-Molina, A., Aluja, M., Zamora-Briseño, J. A., Elizalde-Contreras, J. M., Bautista-Valle, M. V., Arellano de los Santos, J., Sánchez-Martínez, D. E., Rivera-Reséndiz, F. J., Vázquez-Rosas-Landa, M., Camacho-Vázquez, C., Guerrero-Analco, J. A., Ruiz-May, E. (2022). Integrating proteomics and metabolomics approaches to elucidate the ripening process in white Psidium guajava. *Food Chemistry*, 367, 130656.

Moreno, J. C., Mi, J., Alagoz, Y., Al-Babili, S. (2021). Plant apocarotenoids: from retrograde signaling to interspecific communication. *The Plant Journal*, 105, 351-375.

Moreton, J., Izquierdo, A., Emes, R. D. (2016). Assembly, Assessment, and Availability of De novo Generated Eukaryotic Transcriptomes. *Frontiers in Genetics*, 6.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35, W182-W185.

Morreeuw, Z. P., Escobedo-Fregoso, C., Ríos-González, L. J., Castillo-Quiroz, D., Reyes, A. G. (2021). Transcriptome-based metabolic profiling of flavonoids in Agave lechuguilla waste biomass. *Plant Science*, 305, 110748.

Motohashi, R., Rödiger, A., Agne, B., Baerenfaller, K., Baginsky, S. (2012). Common and Specific Protein Accumulation Patterns in Different Albino/Pale-Green Mutants

Reveals Regulon Organization at the Proteome Level *Plant Physiology*, 160, 2189-2201.

Mozgova, G. V., Zaitseva, O. I., Lemesh, V. A. (2012). Structural Changes in Chloroplast Genome Accompanying Albinism in Anther Culture of Wheat and Triticale. *Cereal Research Communications*, 40, 467-475.

Murashige, T. y Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum*, 15, 473-497.

Mustárdy, L. y Garab, G. (2003). Granum revisited. A three-dimensional model —where things fall into place. *Trends in Plant Science*, 8, 117-122.

Naim, F., Dugdale, B., Kleidon, J., Brinin, A., Shand, K., Waterhouse, P., Dale, J. (2018). Gene editing the phytoene desaturase alleles of Cavendish banana using CRISPR/Cas9. *Transgenic Research*, 27, 451-460.

Nakai, M. (2018). New Perspectives on Chloroplast Protein Import. *Plant and Cell Physiology*, 59, 1111-1119.

Nakamura, T., Yagi, Y., Kobayashi, K. (2012). Mechanistic Insight into Pentatricopeptide Repeat Proteins as Sequence-Specific RNA-Binding Proteins for Organellar RNAs in Plants. *Plant and Cell Physiology*, 53, 1171-1179.

Nawaz, G. y Kang, H. (2017). Chloroplast- or Mitochondria-Targeted DEAD-Box RNA Helicases Play Essential Roles in Organellar RNA Metabolism and Abiotic Stress Responses. *Frontiers in Plant Science*, 8.

Negri, S., Commisso, M., Avesani, L., Guzzo, F. (2021). The case of tryptamine and serotonin in plants: a mysterious precursor for an illustrious metabolite. *Journal of Experimental Botany*, 72, 5336-5355.

Newkirk, G. M., de Allende, P., Jinkerson, R. E., Giraldo, J. P. (2021). Nanotechnology Approaches for Chloroplast Biotechnology Advancements. *Frontiers in Plant Science*, 12.

Nishitani, C., Hirai, N., Komori, S., Wada, M., Okada, K., Osakabe, K., Yamamoto, T., Osakabe, Y. (2016). Efficient Genome Editing in Apple Using a CRISPR/Cas9 system. *Scientific Reports*, 6, 31481.

Núñez-Noriega, L. y Salazar-Solano, V. (2009). La producción y comercialización de bacanora como estrategia de desarrollo regional en la sierra sonorense. *Estudios Sociales*, 17, 205-219.

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., Small, I. (2008). On the Expansion of the Pentatricopeptide Repeat Gene Family in Plants. *Molecular Biology and Evolution*, 25, 1120-1128.

Obata, T. y Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences*, 69, 3225-3243.

Oliva, M., Guy, A., Galili, G., Dor, E., Schweitzer, R., Amir, R., Hacham, Y. (2021). Enhanced Production of Aromatic Amino Acids in Tobacco Plants Leads to Increased Phenylpropanoid Metabolites and Tolerance to Stresses. *Frontiers in Plant Science*, 11, 604349.

Ortelt, J. y Link, G. (2021). Plastid Gene Transcription: An Update on Promoters and RNA Polymerases, en: Chloroplast Biotechnology: Methods and Protocols, Maliga, P. (Ed.). Springer US, New York, NY, pp. 49-76.

Ouyang, Y., Wu, Q., Li, J., Sun, S., Sun, S. (2020). S-adenosylmethionine: A metabolite critical to the regulation of autophagy. *Cell Proliferation*, 53, e12891.

Park, H.-S., Jeon, J.-H., Cho, W., Lee, Y., Park, J. Y., Kim, J., Park, Y. S., Koo, H. J., Kang, J. H., Lee, T. J., Kim, S. H., Kim, J.-B., Kwon, H.-Y., Kim, S.-H., Paek, N.-C., Jang, G., Suh, J.-Y., Yang, T.-J. (2022). High-throughput discovery of plastid genes causing albino phenotypes in ornamental chimeric plants. *Horticulture Research*, 10.

Parthasarathy, A., Cross, P. J., Dobson, R. C. J., Adams, L. E., Savka, M. A., Hudson, A. O. (2018). A Three-Ring Circus: Metabolism of the Three Proteogenic Aromatic Amino Acids and Their Role in the Health of Plants and Animals. *Frontiers in Molecular Biosciences*, 5, 29.

Peng, Z., Lu, Q., Verma, D. P. S. (1996). Reciprocal regulation of Δ1-pyrroline-5-carboxylate synthetase and proline dehydrogenase genes controls proline levels during and after osmotic stress in plants. *Molecular Genetics and Genomics*, 253, 334-341.

Pérez-Zavala, M. d. L., Hernández-Arzaba, J. C., Bideshi, D. K., Barboza-Corona, J. E. (2020). Agave: a natural renewable resource with multiple applications. *Journal of the Science of Food and Agriculture*, 100, 5324-5333.

Pfalz, J., Bayraktar, O. A., Prikryl, J., Barkan, A. (2009). Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *The EMBO Journal*, 28, 2042-2052.

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.-J., Oelmüller, R. (2005). pTAC2, -6, and -12 Are Components of the Transcriptionally Active Plastid Chromosome That Are Required for Plastid Gene Expression. *The 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.

Pittermann, J., Cowan, J., Kaufman, N., Baer, A., Zhang, E., Kuty, D. (2018). The water relations and xylem attributes of albino redwood shoots (Sequioa sempervirens (D. Don.) Endl.). *PLOS ONE*, 13, e0191836.

Pogson, B. J. y Albrecht, V. (2011). Genetic Dissection of Chloroplast Biogenesis and Development: An Overview. *Plant Physiology*, 155, 1545-1551.

Pribil, M., Labs, M., Leister, D. (2014). Structure and dynamics of thylakoids in land plants. *Journal of Experimental Botany*, 65, 1955-1972.

Pruitt, K. D., Tatusova, T., Brown, G. R., Maglott, D. R. (2011). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research*, 40, D130-D135.

Puente-Garza, C. A., Gutiérrez-Mora, A., García-Lara, S. (2015). Micropropagation of Agave salmiana: Means to Production of Antioxidant and Bioactive Principles. *Frontiers in Plant Science*, 6.

Pyo, Y. J., Kwon, K.-C., Kim, A., Cho, M. H. (2013). Seedling Lethal1, a Pentatricopeptide Repeat Protein Lacking an E/E+ or DYW Domain in Arabidopsis, Is Involved in Plastid Gene Expression and Early Chloroplast Development. *Plant Physiology*, 163, 1844-1858.

Qin, X., Yang, X., Huang, X., Jin, G., Yang, X., Wu, M., Chen, T., Yi, K. (2021). The complete chloroplast genome of Agave angustifolia. *Mitochondrial DNA Part B*, 6, 3236-3237.

Qiu, X.-M., Sun, Y.-Y., Ye, X.-Y., Li, Z.-G. (2020). Signaling Role of Glutamate in Plants. *Frontiers in Plant Science*, 10, 1743.

Qiu, Z., Chen, D., Teng, L., Guan, P., Yu, G., Zhang, P., Song, J., Zeng, Q., Zhu, L. (2022). OsWHY1 Interacts with OsTRX z and is Essential for Early Chloroplast Development in Rice. *Rice*, 15, 50.

Rivals, E., Bruyère, C. m., Toffano-Nioche, C., Lecharny, A. (2006). Formation of the Arabidopsis Pentatricopeptide Repeat Family. *Plant Physiology*, 141, 825-839.

Robert, M. L., Herrera, J. L., Chan, J. L., Contreras, F. (1992). Micropropagation of Agave spp, en: High-Tech and Micropropagation III, Bajaj, Y. P. S. (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 306-329.

Robert, M. L., Herrera-Herrera, J. L., Castillo, E., Ojeda, G., Herrera-Alamillo, M. A. (2006). An Efficient Method for the Micropropagation of Agave Species, en: Plant Cell Culture Protocols, Loyola-Vargas, V. M. y Vázquez-Flota, F. (Eds.). Humana Press, Totowa, NJ, pp. 165-178.

Robinson, M. D., McCarthy, D. J., Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.

Robles, P. y Quesada, V. (2021). Research Progress in the Molecular Functions of Plant mTERF Proteins. *Cells*, 10, Article 205.

Rochaix, J.-D. y Ramundo, S. (2017). Chloroplast signaling and quality control. *Essays in Biochemistry*, 62, 13-20.

Rödiger, A., Agne, B., Dobritzsch, D., Helm, S., Müller, F., Pötzsch, N., Baginsky, S. (2021). Chromoplast differentiation in bell pepper (Capsicum annuum) fruits. *The Plant Journal*, 105, 1431-1442.

Rodríguez-Garay, B. y Rodríguez-Domínguez, J. M. (2018). Micropropagation of Agave Species, en: Plant Cell Culture Protocols, Loyola-Vargas, V. M. y Ochoa-Alejo, N. (Eds.). Springer New York, New York, NY, pp. 151-159.

Roelfsema, M. R. G., Konrad, K. R., Marten, H., Psaras, G. K., Hartung, W., Hedrich, R. (2006). Guard cells in albino leaf patches do not respond to photosynthetically active radiation, but are sensitive to blue light, CO2 and abscisic acid. *Plant, Cell & Environment*, 29, 1595-1605.

Rolland, N., Bouchnak, I., Moyet, L., Salvi, D., Kuntz, M. (2018). The Main Functions of Plastids, en: Plastids: Methods and Protocols, Maréchal, E. (Ed.). Springer US, New York, NY, pp. 73-85.

Romero, L. C., Aroca, M. Á., Laureano-Marín, A. M., Moreno, I., García, I., Gotor, C. (2014). Cysteine and Cysteine-Related Signaling Pathways in Arabidopsis thaliana. *Molecular Plant*, 7, 264-276.

Rovira, A. G. y Smith, A. G. (2019). PPR proteins – orchestrators of organelle RNA metabolism. *Physiologia Plantarum*, 166, 451-459.

RStudio, T. (2021). RStudio: Integrated Development Environment for R.

SAGARPA. (2016). *Agave angustifolia* Haw. Técnicas para el transplante de vitroplantas a condiciones de agostadero. *Plan Nacional de Recursos Fitogenéticos*1-18.

SAGARPA. (2017). Plan Agrícola Nacional. Agave tequilero y mezcalero mexicano.

Saha, D., Prasad, A. M., Srinivasan, R. (2007). Pentatricopeptide repeat proteins and their emerging roles in plants. *Plant Physiology and Biochemistry*, 45, 521-534.

Sahijram, L. (2015). Somaclonal Variation in Micropropagated Plants, en: Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology, Bahadur, B., Venkat Rajam, M., Sahijram, L., y Krishnamurthy, K. V. (Eds.). Springer India, New Delhi, pp. 407-416.

Salazar-Iribe, A. y De-la-Peña, C. (2020). Auxins, the hidden player in chloroplast development. *Plant Cell Reports*, 39, 1595-1608.

Sanger, F. y Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94, 441-448.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74, 5463-5467.

Sarwar, M. B., Ahmad, Z., Rashid, B., Hassan, S., Gregersen, P. L., Leyva, M. D. l. O., Nagy, I., Asp, T., Husnain, T. (2019). De novo assembly of Agave sisalana transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Scientific Reports*, 9, 396.

Satou, M., Enoki, H., Oikawa, A., Ohta, D., Saito, K., Hachiya, T., Sakakibara, H., Kusano, M., Fukushima, A., Saito, K., Kobayashi, M., Nagata, N., Myouga, F., Shinozaki, K., Motohashi, R. (2014). Integrated analysis of transcriptome and metabolome of Arabidopsisalbino or pale green mutants with disrupted nuclear-encoded chloroplast proteins. *Plant Molecular Biology*, 85, 411-428.

Schweer, J., Loschelder, H., Link, G. (2006). A promoter switch that can rescue a plant sigma factor mutant. *FEBS Letters*, 580, 6617-6622.

Sehar, Z., Gautam, H., Iqbal, N., Alvi, A. F., Jahan, B., Fatma, M., Albaqami, M., Khan, N. A. (2022). The Functional Interplay between Ethylene, Hydrogen Sulfide, and Sulfur in Plant Heat Stress Tolerance. *Biomolecules*, 12, 678.

Sharma, M. y Pandey, G. K. (2016). Expansion and Function of Repeat Domain Proteins During Stress and Development in Plants. *Frontiers in Plant Science*, 6.

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, 50, W216-W221.

Shi, K., Gu, J., Guo, H., Zhao, L., Xie, Y., Xiong, H., Li, J., Zhao, S., Song, X., Liu, L. (2017). Transcriptome and proteomic analyses reveal multiple differences associated with

chloroplast development in the spaceflight-induced wheat albino mutant mta. *PLOS ONE*, 12, e0177992.

Shi, L.-X. y Theg, S. M. (2013). The chloroplast protein import system: From algae to trees. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833, 314-331.

Shikanai, T. y Fujii, S. (2013). Function of PPR proteins in plastid gene expression. *RNA Biology*, 10, 1446-1456.

Shimizu, T. y Masuda, T. (2021). The Role of Tetrapyrrole- and GUN1-Dependent Signaling on Chloroplast Biogenesis. *Plants*, 10, 196.

Shyjan, A. W. y Buttow, R. A. (1993). Intracellular dialogue. *Current Biology*, 3, 398-400.

Sierra, J., Escobar-Tovar, L., Leon, P. (2023). Plastids: diving into their diversity, their functions, and their role in plant development. *Journal of Experimental Botany*, 74, 2508-2526.

Silva, L. A. S., Sampaio, V. F., Barbosa, L. C. S., Machado, M., Flores-Borges, D. N. A., Sales, J. F., de Oliveira, D. C., Mayer, J. L. S., Kuster, V. C., Rocha, D. I. (2020). Albinism in plants – far beyond the loss of chlorophyll: Structural and physiological aspects of wild-type and albino royal poinciana (Delonix regia) seedlings. *Plant Biology*, 22, 761-768.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210-3212.

Singhal, R., Pal, R., Dutta, S. (2023). Chloroplast Engineering: Fundamental Insights and Its Application in Amelioration of Environmental Stress. *Applied Biochemistry and Biotechnology*, 195, 2463-2482.

Slatko, B. E., Gardner, A. F., Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current protocols in molecular biology*, 122, e59-e59.

Small, I., Peeters, N., Legeai, F., Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *PROTEOMICS*, 4, 1581-1590.

Small, I. D. y Peeters, N. (2000). The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends in Biochemical Sciences*, 25, 45-47.

Solymosi, K., Lethin, J., Aronsson, H. (2018). Diversity and Plasticity of Plastids in Land Plants, en: Plastids: Methods and Protocols, Maréchal, E. (Ed.). Springer US, New York, NY, pp. 55-72.

Speiser, A., Silbermann, M., Dong, Y., Haberland, S., Uslu, V. V., Wang, S., Bangash, S. A. K., Reichelt, M., Meyer, A. J., Wirtz, M., Hell, R. (2018). Sulfur Partitioning between Glutathione and Protein Synthesis Determines Plant Growth. *Plant Physiology*, 177, 927-937.

Staehelin, L. A. (2003). Chloroplast structure: from chlorophyll granules to supra-molecular architecture of thylakoid membranes. *Photosynthesis Research*, 76, 185-196.

Stern, D. B., Hanson, M. R., Barkan, A. (2004). Genetics and genomics of chloroplast biogenesis: maize as a model system. *Trends in Plant Science*, 9, 293-301.

Stewart, J. R. (2015). Agave as a model CAM crop system for a warming and drying world. *Frontiers in Plant Science*, 6.

Su, N., Hu, M.-L., Wu, D.-X., Wu, F.-Q., Fei, G.-L., Lan, Y., Chen, X.-L., Shu, X.-L., Zhang, X., Guo, X.-P., Cheng, Z.-J., Lei, C.-L., Qi, C.-K., Jiang, L., Wang, H., Wan, J.-M. (2012). Disruption of a Rice Pentatricopeptide Repeat Protein Causes a Seedling-Specific Albino Phenotype and Its Utilization to Enhance Seed Purity in Hybrid Rice Production. *Plant Physiology*, 159, 227-238.

Suárez-González, E. M., López, M. G., Délano-Frier, J. P., Gómez-Leyva, J. F. (2014). Expression of the 1-SST and 1-FFT genes and consequent fructan accumulation in Agave tequilana and A. inaequidens is differentially induced by diverse (a)biotic-stress related elicitors. *Journal of Plant Physiology*, 171, 359-372.

Suetsugu, K., Yamato, M., Miura, C., Yamaguchi, K., Takahashi, K., Ida, Y., Shigenobu, S., Kaminaka, H. (2017). Comparison of green and albino individuals of the partially mycoheterotrophic orchid Epipactis helleborine on molecular identities of mycorrhizal fungi, nutritional modes and gene expression in mycorrhizal roots. *Molecular Ecology*, 26, 1652-1669.

Sun, B., Zheng, A., Jiang, M., Xue, S., Yuan, Q., Jiang, L., Chen, Q., Li, M., Wang, Y., Zhang, Y., Luo, Y., Wang, X., Zhang, F., Tang, H. (2018). CRISPR/Cas9-mediated mutagenesis of homologous genes in Chinese kale. *Scientific Reports*, 8, 16786.

Suzuki, R., Sugita, C., Aoki, S., Sugita, M. (2022). Physcomitrium patens pentatricopeptide repeat protein PpPPR_32 is involved in the accumulation of psaC mRNA encoding the iron sulfur protein of photosystem I. *Genes to Cells*, 27, 293-304.

Sveshnikova, N., Soll, J., Schleiff, E. (2000). Toc34 is a preprotein receptor regulated by GTP and phosphorylation. *Proceedings of the National Academy of Sciences*, 97, 4973-4978.

Tadini, L., Ferrari, R., Lehniger, M.-K., Mizzotti, C., Moratti, F., Resentini, F., Colombo, M., Costa, A., Masiero, S., Pesaresi, P. (2018). Trans-splicing of plastid rps12 transcripts, mediated by AtPPR4, is essential for embryo patterning in Arabidopsis thaliana. *Planta*, 248, 257-265.

Tan, J., Tan, Z., Wu, F., Sheng, P., Heng, Y., Wang, X., Ren, Y., Wang, J., Guo, X., Zhang, X., Cheng, Z., Jiang, L., Liu, X., Wang, H., Wan, J. (2014). A Novel Chloroplast-Localized Pentatricopeptide Repeat Protein Involved in Splicing Affects Chloroplast Development and Abiotic Stress Response in Rice. *Molecular Plant*, 7, 1329-1349.

Tanwar, N., Arya, S. S., Rookes, J. E., Cahill, D. M., Lenka, S. K., Bansal, K. C. (2022). Prospects of chloroplast metabolic engineering for developing nutrient-dense food crops. *Critical Reviews in Biotechnology*1-18.

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., Hood, L. E. (2004). Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Molecular & Cellular Proteomics*, 3, 960-969.

Tian, S., Jiang, L., Gao, Q., Zhang, J., Zong, M., Zhang, H., Ren, Y., Guo, S., Gong, G., Liu, F., Xu, Y. (2017). Efficient CRISPR/Cas9-based gene knockout in watermelon. *Plant cell reports*, 36, 399-406.

Toujani, W., Muñoz-Bertomeu, J., Flores-Tornero, M., Rosa-Téllez, S., Anoman, A. D., Alseekh, S., Fernie, A. R., Ros, R. (2013). Functional Characterization of the Plastidial 3-Phosphoglycerate Dehydrogenase Family in Arabidopsis. *Plant Physiology*, 163, 1164-1178.

Tozawa, Y., Teraishi, M., Sasaki, T., Sonoike, K., Nishiyama, Y., Itaya, M., Miyao, A., Hirochika, H. (2007). The plastid sigma factor SIG1 maintains photosystem I activity via regulated expression of the psaA operon in rice chloroplasts. *The Plant Journal*, 52, 124-132.

Tsunoyama, Y., Ishizaki, Y., Morikawa, K., Kobori, M., Nakahira, Y., Takeba, G., Toyoshima, Y., Shiina, T. (2004). Blue light-induced transcription of plastid-encoded psbD gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. *Proceedings of the National Academy of Sciences*, 101, 3304-3309.

Tsunoyama, Y., Morikawa, K., Shiina, T., Toyoshima, Y. (2002). Blue light specific and differential expression of a plastid σ factor, Sig5 in Arabidopsis thaliana. *FEBS Letters*, 516, 225-228.

Us-Camas, R., Castillo-Castro, E., Aguilar-Espinosa, M., Limones-Briones, V., Rivera-Madrid, R., Robert-Díaz, M. L., De-la-Peña, C. (2017). Assessment of molecular and epigenetic changes in the albinism of Agave angustifolia Haw. *Plant Science*, 263, 156-167.

Van Eerden, F. J., Melo, M. N., Frederix, P. W. J. M., Periole, X., Marrink, S. J. (2017). Exchange pathways of plastoquinone and plastoquinol in the photosystem II complex. *Nature Communications*, 8, 15214.

Wagner, G. P., Kin, K., Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, 131, 281-285.

Wagoner, J. A., Sun, T., Lin, L., Hanson, M. R. (2015). Cytidine Deaminase Motifs within the DYW Domain of Two Pentatricopeptide Repeat-containing Proteins Are Required for Site-specific Chloroplast RNA Editing *. *Journal of Biological Chemistry*, 290, 2957-2968.

Wang, B., Kumar, V., Olson, A., Ware, D. (2019). Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Frontiers in Genetics*, 10.

Wang, D., Liu, H., Zhai, G., Wang, L., Shao, J., Tao, Y. (2016). OspTAC2 encodes a pentatricopeptide repeat protein and regulates rice chloroplast development. *Journal of Genetics and Genomics*, 43, 601-608.

Wang, L., Czedik-Eysenberg, A., Mertz, R. A., Si, Y., Tohge, T., Nunes-Nesi, A., Arrivault, S., Dedow, L. K., Bryant, D. W., Zhou, W., Xu, J., Weissmann, S., Studer, A., Li, P., Zhang, C., LaRue, T., Shao, Y., Ding, Z., Sun, Q., Patel, R. V., Turgeon, R., Zhu, X., Provart, N. J., Mockler, T. C., Fernie, A. R., Stitt, M., Liu, P., Brutnell, T. P. (2014). Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. *Nature Biotechnology*, 32, 1158-1165.

Wang, M., Zhu, X., Li, Y., Xia, Z. (2020a). Transcriptome analysis of a new maize albino mutant reveals that zeta-carotene desaturase is involved in chloroplast development and retrograde signaling. *Plant Physiology and Biochemistry*, 156, 407-419.

Wang, P., Zhang, J., Sun, L., Ma, Y., Xu, J., Liang, S., Deng, J., Tan, J., Zhang, Q., Tu, L., Daniell, H., Jin, S., Zhang, X. (2018a). High efficient multisites genome editing in allotetraploid cotton (Gossypium hirsutum) using CRISPR/Cas9 system. *Plant Biotechnology Journal*, 16, 137-150.

Wang, X., An, Y., Xu, P., Xiao, J. (2021). Functioning of PPR Proteins in Organelle RNA Metabolism and Chloroplast Biogenesis. *Frontiers in Plant Science*, 12.

Wang, X., Huang, X., Chen, L., Xie, Z., Tan, S., Qin, X., Chen, T., Huang, Y., Xi, J., Chen, H., Yi, K. (2023). Transcriptome Sequencing of Agave amaniensis Reveals Shoot-Related Expression Patterns of Expansin A Genes in Agave. *Plants*, 12, 2020.

Wang, X., Yang, Z., Zhang, Y., Zhou, W., Zhang, A., Lu, C. (2020b). Pentatricopeptide repeat protein PHOTOSYSTEM I BIOGENESIS FACTOR2 is required for splicing of ycf3. *Journal of Integrative Plant Biology*, 62, 1741-1761.

Wang, Y., Yang, Z., Zhang, M., Ai, P. (2022). A chloroplast-localized pentatricopeptide repeat protein involved in RNA editing and splicing and its effects on chloroplast development in rice. *BMC Plant Biology*, 22, 437.

Wang, Z.-w., Lv, J., Xie, S.-z., Zhang, Y., Qiu, Z.-n., Chen, P., Cui, Y.-t., Niu, Y.-f., Hu, S.-k., Jiang, H.-z., Ge, S.-z., Trinh, H., Lei, K.-r., Bai, W.-q., Zhang, Y., Guo, L.-b., Ren, D.-y. (2018b). OsSLA4 encodes a pentatricopeptide repeat protein essential for early chloroplast development and seedling growth in rice. *Plant Growth Regulation*, 84, 249-260.

Weckx, S., Inzé, D., Maene, L. (2019). Tissue Culture of Oil Palm: Finding the Balance Between Mass Propagation and Somaclonal Variation. *Frontiers in Plant Science*, 10.

Westfall, C. S., Xu, A., Jez, J. M. (2014). Structural Evolution of Differential Amino Acid Effector Regulation in Plant Chorismate Mutases. *Journal of Biological Chemistry*, 289, 28619-28628.

Wilkins, M. R., Sanchez, J.-C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., Williams, K. L. (1996). Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It. *Biotechnology and Genetic Engineering Reviews*, 13, 19-50.

Williams-Carrier, R., Brewster, C., Belcher, S. E., Rojas, M., Chotewutmontri, P., Ljungdahl, S., Barkan, A. (2019). The Arabidopsis pentatricopeptide repeat protein LPE1 and its maize ortholog are required for translation of the chloroplast psbJ RNA. *The Plant Journal*, 99, 56-66.

Wu, G.-Z., Chalvin, C., Hoelscher, M., Meyer, E. H., Wu, X. N., Bock, R. (2018). Control of Retrograde Signaling by Rapid Turnover of GENOMES UNCOUPLED1. *Plant Physiology*, 176, 2472-2495.

Wu, H. y Zhang, L. (2010). The PPR protein PDM1 is involved in the processing of rpoA pre-mRNA in Arabidopsis thaliana. *Chinese Science Bulletin*, 55, 3485-3489.

Wu, J., Liu, H., Lu, S., Hua, J., Zou, B. (2021). Identification and expression analysis of chloroplast ribonucleoproteins (cpRNPs) in Arabidopsis and rice. *Genome*, 64, 515-524.

Xiao, H., Xu, Y., Ni, C., Zhang, Q., Zhong, F., Huang, J., Liu, W., Peng, L., Zhu, Y., Hu, J. (2018). A rice dual-localized pentatricopeptide repeat protein is involved in organellar RNA editing together with OsMORFs. *Journal of Experimental Botany*, 69, 2923-2936.

Xu, J.-J., Fang, X., Li, C.-Y., Yang, L., Chen, X.-Y. (2020). General and specialized tyrosine metabolism pathways in plants. *aBIOTECH*, 1, 97-105.

Yamashita, H., Kambe, Y., Ohshio, M., Kunihiro, A., Tanaka, Y., Suzuki, T., Nakamura, Y., Morita, A., Ikka, T. (2021). Integrated Metabolome and Transcriptome Analyses Reveal Etiolation-Induced Metabolic Changes Leading to High Amino Acid Contents in a Light-Sensitive Japanese Albino Tea Cultivar. *Frontiers in Plant Science*, 11, 611140.

Yan, C., Peng, L., Zhang, L., Qiu, Z. (2020). Fine mapping of a candidate gene for cool-temperature-induced albinism in ornamental kale. *BMC Plant Biology*, 20, Article 460.

Yan, J., Liu, B., Cao, Z., Chen, L., Liang, Z., Wang, M., Liu, W., Lin, Y. e., Jiang, B. (2022). Cytological, genetic and transcriptomic characterization of a cucumber albino mutant. *Frontiers in Plant Science*, 13.

Yan, J., Yao, Y., Hong, S., Yang, Y., Shen, C., Zhang, Q., Zhang, D., Zou, T., Yin, P. (2019). Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Research*, 47, 3728-3738.

Yang, P., Li, Y., He, C., Yan, J., Zhang, W., Li, X., Xiang, F., Zuo, Z., Li, X., Zhu, Y., Liu, X., Zhao, X. (2020). Phenotype and TMT-based quantitative proteomics analysis of Brassica napus reveals new insight into chlorophyll synthesis and chloroplast structure. *Journal of Proteomics*, 214, 103621.

Yao, J. L. y Cohen, D. (2000). Multiple gene control of plastome-genome incompatibility and plastid DNA inheritance in interspecific hybrids of Zantedeschia. *Theoretical and Applied Genetics*, 101, 400-406.

Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., Gonda, I., Lewinsohn, E., Rhodes, D., Dudareva, N. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nature Communications*, 4, 2833.

Yu, Q.-B., Li, G., Wang, G., Sun, J.-C., Wang, P.-C., Wang, C., Mi, H.-L., Ma, W.-M., Cui, J., Cui, Y.-L., Chong, K., Li, Y.-X., Li, Y.-H., Zhao, Z., Shi, T.-L., Yang, Z.-N. (2008). Construction of a chloroplast protein interaction network and functional mining of photosynthetic proteins in Arabidopsis thaliana. *Cell Research*, 18, 1007-1019.

Yuan, L., Xiong, L.-g., Deng, T.-t., Wu, Y., Li, J., Liu, S.-q., Huang, J.-a., Liu, Z.-h. (2015). Comparative profiling of gene expression in Camellia sinensis L. cultivar AnJiBaiCha leaves during periodic albinism. *Gene*, 561, 23-29.

Yuan, N., Wang, J., Zhou, Y., An, D., Xiao, Q., Wang, W., Wu, Y. (2019). EMB-7L is required for embryogenesis and plant development in maize involved in RNA splicing of multiple chloroplast genes. *Plant Science*, 287, 110203.

Zghidi, W., Merendino, L., Cottet, A., Mache, R., Lerbs-Mache, S. (2006). Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the psb N gene in plastids. *Nucleic Acids Research*, 35, 455-464.

Zhang, B., Yang, X., Yang, C., Li, M., Guo, Y. (2016). Exploiting the CRISPR/Cas9 System for Targeted Genome Mutagenesis in Petunia. *Scientific Reports*, 6, 20315.

Zhang, H.-D., Cui, Y.-L., Huang, C., Yin, Q.-Q., Qin, X.-M., Xu, T., He, X.-F., Zhang, Y., Li, Z.-R., Yang, Z.-N. (2015). PPR protein PDM1/SEL1 is involved in RNA editing and splicing of plastid genes in Arabidopsis thaliana. *Photosynthesis Research*, 126, 311-321.

Zhang, J., Xiao, J., Li, Y., Su, B., Xu, H., Shan, X., Song, C., Xie, J., Li, R. (2017). PDM3, a pentatricopeptide repeat-containing protein, affects chloroplast development. *Journal of Experimental Botany*, 68, 5615-5627.

Zhang, J.-H., Zeng, J.-C., Wang, X.-M., Chen, S.-F., Albach, D. C., Li, H.-Q. (2020). A revised classification of leaf variegation types. *Flora*, 272, 151703.

Zhang, L., Zhou, W., Che, L., Rochaix, J.-D., Lu, C., Li, W., Peng, L. (2019). PPR Protein BFA2 Is Essential for the Accumulation of the atpH/F Transcript in Chloroplasts. *Frontiers in Plant Science*, 10.

Zhang, M., Zhao, Y., Meng, Y., Xiao, Y., Zhao, J., Xiao, B., An, C., Gao, Y. (2022). PPR proteins in the tea plant (Camellia sinensis) and their potential roles in the leaf color changes. *Scientia Horticulturae*, 293, 110745.

Zhao, X., Huang, J., Chory, J. (2020). Unraveling the Linkage between Retrograde Signaling and RNA Metabolism in Plants. *Trends in Plant Science*, 25, 141-147.

Zheng, S., Poczai, P., Hyvönen, J., Tang, J., Amiryousefi, A. (2020). Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. *Frontiers in Genetics*, 11.

Zhou, W., Lu, Q., Li, Q., Wang, L., Ding, S., Zhang, A., Wen, X., Zhang, L., Lu, C. (2017). PPR-SMR protein SOT1 has RNA endonuclease activity. *Proceedings of the National Academy of Sciences*, 114, E1554-E1563.

Zimorski, V., Ku, C., Martin, W. F., Gould, S. B. (2014). Endosymbiotic theory for organelle origins. *Current Opinion in Microbiology*, 22, 38-48.

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

Zoschke, R., Kroeger, T., Belcher, S., Schöttler, M. A., Barkan, A., Schmitz-Linneweber, C. (2012). The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast atpB/E mRNA. *The Plant Journal*, 72, 547-558.

Zoschke, R., Qu, Y., Zubo, Y. O., Börner, T., Schmitz-Linneweber, C. (2013). Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in Arabidopsis thaliana. *Journal of Plant Research*, 126, 403-414.

Żur, I., Gajecka, M., Dubas, E., Krzewska, M., Szarejko, I. (2021). Albino Plant Formation in Androgenic Cultures: An Old Problem and New Facts, en: Doubled Haploid

Technology: Volume 2: Hot Topics, Apiaceae, Brassicaceae, Solanaceae, Segui-Simarro, J. M. (Ed.). Springer US, New York, NY, pp. 3-23.