



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE BIOLOGÍA

BIOLOGÍA EVOLUTIVA

IDENTIFICACIÓN DE LOS GENES *PLETHORA* EN *Mammillaria san-angelensis*

(Cactaceae) Y AVANCES EN SU AISLAMIENTO EXPERIMENTAL EN OTRAS

ESPECIES DEL GÉNERO *Mammillaria*

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

LUIS FERNANDO CISNEROS LEMUS

DR. ULISES YUNUÉN ROSAS LÓPEZ

INSTITUTO DE BIOLOGÍA, UNAM

DR. CÉSAR LUIS CUEVAS VELÁZQUEZ

FACULTAD DE QUÍMICA, UNAM

DR. ALEJANDRO FRANCISCO OCEGUERA FIGUEROA

INSTITUTO DE BIOLOGÍA, UNAM

CIUDAD UNIVERSITARIA, CD. MX.

FEBRERO, 2024



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE BIOLOGÍA

BIOLOGÍA EVOLUTIVA

IDENTIFICACIÓN DE LOS GENES *PLETHORA* EN *Mammillaria san-angelensis*

(Cactaceae) Y AVANCES EN SU AISLAMIENTO EXPERIMENTAL EN OTRAS

ESPECIES DEL GÉNERO *Mammillaria*

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

LUIS FERNANDO CISNEROS LEMUS

DR. ULISES YUNUÉN ROSAS LÓPEZ

INSTITUTO DE BIOLOGÍA, UNAM

DR. CÉSAR LUIS CUEVAS VELÁZQUEZ

FACULTAD DE QUÍMICA, UNAM

DR. ALEJANDRO FRANCISCO OCEGUERA FIGUEROA

INSTITUTO DE BIOLOGÍA, UNAM

CIUDAD UNIVERSITARIA, CD. MX.

FEBRERO, 2024

COORDINACIÓN GENERAL DE ESTUDIOS DE POSGRADO
COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
INSTITUTO DE BIOLOGÍA
OFICIO: CGEP/CPCB/ IB /0939/2023
ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **16 de octubre de 2023** se aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **BIOLOGÍA EVOLUTIVA** del estudiante **CISNEROS LEMUS LUIS FERNANDO** con número de cuenta **312246158** con la tesis titulada **“Identificación de los genes PLETHORA en *Mammillaria san-angelensis* (Cactaceae) y avances en su aislamiento experimental en otras especies del género *Mammillaria*”,** realizada bajo la dirección del **DR. ULISES YUNUÉN ROSAS LÓPEZ**, quedando integrado de la siguiente manera:

Presidente: DR. ÁNGEL SALVADOR ARIAS MONTES
Vocal: DRA. SVETLANA SHISHKOVA
Vocal: DRA. DULCE ALEJANDRA ROUGON CARDOSO
Vocal: DR. GUSTAVO RODRÍGUEZ ALONSO
Secretario: DR. CÉSAR LUIS CUEVAS VELÁZQUEZ
Emergente: DR. ALEJANDRO FRANCISCO OCEGUERA FIGUEROA

Sin otro particular, me es grato enviarle un cordial saludo.

A T E N T A M E N T E
“POR MI RAZA HABLARÁ EL ESPÍRITU”
Ciudad Universitaria, Cd. Mx., a 30 de noviembre de 2024

COORDINADOR DEL PROGRAMA



DR. ADOLFO GERARDO NAVARRO SIGÜENZA

c. c. p. Expediente del alumno

AGNS/RGA/EARR/rga



Agradecimientos institucionales

Al Posgrado en Ciencias Biológicas de la UNAM por las aportaciones de esta institución y sus profesores a mi formación profesional, académica y personal.

Al CONAHCYT por la beca de posgrado con el número de apoyo 803737 otorgada durante la realización de este proyecto de investigación.

Al proyecto PAPIIT DGAPA UNAM IN214322 (“Las raíces vegetales como modelo de estudio de la biodiversidad, conservación, fisiología y anatomía”), otorgado al Dr. Ulises Y. Rosas López, con el cual se financió este proyecto.

A mi tutor, el Dr. Ulises Y. Rosas López. Sin su apoyo académico y personal, este trabajo no hubiera sido posible.

A mi comité tutor, el Dr. Alejandro Francisco Ocegüera Figueroa (Instituto de Biología) y el Dr. César Luis Cuevas Velázquez (Facultad de Química) por los atentos comentarios, y el apoyo para lograr la mejor versión de este trabajo.

Agradecimientos a título personal

A mi jurado, el Dr. Salvador Arias, el Dr. César Cuevas, el Dr. Gustavo Rodríguez, la Dra. Alejandra Rougon, y la Dra. Svetlana Shishkova, gracias a su apoyo este trabajo se entrega en las mejores condiciones posibles.

Al Prof. Michael Purugganan de la New York University, y al Dr. Darío Copetti del Arizona Genomics Institute por el apoyo en la secuenciación PacBio.

A la Dra. Svetlana Shishkova, y al Dr. Gustavo Rodríguez Alonso, colaboradores de este proyecto y cuya orientación y apoyo fue vital para llevar a término este trabajo.

Al servicio del Laboratorio de Biología Molecular del Instituto de Biología de la Universidad Nacional Autónoma de México, como parte del Laboratorio Nacional de Biodiversidad, y en particular el apoyo técnico de la M. en C. Andrea Jiménez Marín pues su apoyo y orientación fue vital en todo el trabajo de biología molecular realizado para esta tesis.

Al Dr. Ángel Salvador Arias Montes, Curador de la Colección Nacional de Cactáceas del Jardín Botánico del IBUNAM, por brindar el acceso al germoplasma empleado.

Al Biol. Panuncio Jerónimo Reyes Santiago por proveer el material biológico (semilla y planta viva), necesario para esta investigación, y proveniente del Programa de Propagación de Cactáceas y Suculentas del Jardín Botánico del IBUNAM.

Al Dr. Victor Chávez Ávila y su equipo de trabajo por brindar las plantas vivas provenientes de los trabajos del Dr. Abraham Rubluo, para la secuenciación PacBio.

Al Jardín Botánico y al Instituto de Biología de la UNAM y su personal pues solo con su amable apoyo se pudo llevar a cabo este proyecto.

Al Biól José de Jesús Gonzales Sánchez por su ayuda en la disección de las plantas de *Mammillaria*.

A los compañeros del Laboratorio de Biología de Raíces por sus comentarios sobre el trabajo que ayudaron a cubrir sus flancos expuestos.

A mis padres pues su apoyo en todos los aspectos de mi vida fue indispensable para llevar a cabo este proyecto. Sus consejos, su apoyo emocional, incluso las ocasiones en que me ayudaban a salir a tiempo de casa hicieron toda la diferencia en estos dos años.

A mi novia, por el cariño con el que me escuchó, apoyo y dio ánimos cada vez que algo no salía de la forma esperada y mis ánimos flaqueaban.

A mis amigos por escucharme, darme sus opiniones y apoyarme en este camino.

Índice

Resumen.....	1
Abstract.....	2
1. Introducción.....	4
2. Antecedentes.....	6
2.1 El desarrollo de la raíz.....	6
2.2 Estructura de la raíz.....	8
2.3 El Meristemo Apical de la Raíz (RAM, <i>Root Apical Meristem</i>).....	9
2.4 El crecimiento indeterminado de la raíz.....	10
2.5 Las raíces en la familia Cactaceae.....	12
2.6 Vías que regulan el mantenimiento o agotamiento del meristemo apical de la raíz.....	13
2.7 Los genes <i>PLETHORA</i>	14
2.8 Papel de los genes <i>PLT</i> en el desarrollo vegetal.....	16
2.9 Los genes <i>PLT</i> y el crecimiento de la raíz.....	18
2.10 Avances recientes en el estudio de los genes <i>PLETHORA</i> y su papel en el desarrollo de raíz en especies no modelo.....	19
2.11 El género <i>Mammillaria</i>	20
2.12 Citogenética de <i>Mammillaria</i>	21
3 Justificación.....	22
4 Hipótesis.....	22
5 Objetivos.....	22
5.1 Objetivo general.....	22
5.2 Objetivos particulares.....	23
6 Metodología.....	24
6.1 Secuenciación del genoma completo de <i>M. san-angelensis</i> usando la tecnología PacBio.....	24
6.2 Trabajo bioinformático.....	25
6.3 Análisis filogenético.....	27
6.4 Elaboración de modelos para los genes identificados.....	29
6.5 Trabajo molecular experimental.....	29
7. Resultados.....	34
7.1 Las secuencias genómicas de <i>M. san-angelensis</i>	34

7.2 El análisis filogenético permite reconocer a los ortólogos putativos de <i>AtPLT1-5, 7</i> en el genoma de <i>M. san-angelensis</i>	37
7.3 Amplificación de los productos <i>MsaPLT1/2.1, MsaPLT1/2.2, MsaPLT3/7.1, MsaPLT5</i> y sus ortólogos en cinco especies del género <i>Mammillaria</i>	45
8. Discusión.....	61
Conclusiones.....	70
Referencias bibliográficas	71
Anexos	80

Lista de abreviaturas

PLT: PLETHORA

AP2: APETALA2

EREBP: Etylene Response Element Binding Proteins

RAM: Root Apical Meristem (Meristemo apical de la raíz)

PCR: Polymerase Chain Reaction (reacción en cadena de la polimerasa)

SCN: Stem Cell Niche (Nicho de Células Troncales)

MTO2: METHIONINE OVER-ACCUMULATOR 2

SHR: SHORT-ROOT

SCR: SCARECROW

ANT: AINTEGUMENTA

AIL: AINTEGUMENTA-LIKE

ERF: ETHYLENE RESPONSE FACTOR

BBM: BABY BOOM

QC: Quiescent center (centro quiescente)

CCS: Circular Consensus Sequencing (Lectura de consenso circular)

ORF: Open Reading Frame (Marco de lectura abierto)

Resumen

En *Arabidopsis thaliana* existen seis genes *PLETHORA* (*PLT*) que codifican para factores de transcripción que pertenecen a la familia AP2/EREBP. La familia de genes AP2/EREBP surgió antes de la divergencia de las plantas terrestres y sus miembros están altamente conservados. Las proteínas PLT tienen un papel central en el desarrollo de raíz pues regulan la rizotaxia, la diferenciación celular, y el mantenimiento del Meristemo Apical de la Raíz (conocido como RAM por sus siglas en inglés). La gran mayoría de las plantas analizadas de la subfamilia Cactoideae presentan crecimiento determinado de la raíz. Este fenómeno consiste en el agotamiento del RAM de la raíz primaria lo cual da lugar a la diferenciación de todas las células del ápice y, como consecuencia, a la detención del crecimiento de dicha raíz, así como a la ramificación del sistema radical, pudiendo originar un sistema de raíces de apariencia fibrosa. Los mecanismos genéticos que regulan este proceso no han sido descritos. Estudios filogenómicos de los genes PLT señalan que la presencia de los ortólogos de PLT1 y PLT2 en dicotiledóneas y su ausencia en monocotiledóneas puede estar asociado al desarrollo de un sistema de raíces pivotante en dicotiledóneas y un sistema de raíz fibroso en las monocotiledóneas. Asimismo, estudios con mutantes de pérdida de función de *PLT1* y *PLT2* en *A. thaliana* muestran fenotipos con desarrollo de raíz morfológicamente similar al desarrollo natural de las raíces de algunas cactáceas de la subfamilia Cactoideae. Además, el estudio del transcriptoma del ápice de raíz de *Pachycereus pringlei* (Cactaceae) con RAM activo o exhausto mostró que los programas transcripcionales que operan en el ápice de la raíz son similares en esta especie y en *A. thaliana*. Las plantas del género *Mammillaria* son un modelo interesante para el estudio del crecimiento determinado de la raíz primaria en cactáceas pues presentan un ciclo de vida relativamente corto, alta diversidad y son de fácil reproducción. Sin embargo, la exploración de este fenómeno en *Mammillaria* se ve limitada por la escasez de recursos genómicos para este género. En este estudio llevamos a cabo la identificación de los ortólogos putativos de los genes *PLT* de *A. thaliana* en *Mammillaria san-angelensis* utilizando métodos bioinformáticos y moleculares, con la finalidad de expandir nuestro conocimiento del agotamiento del

RAM en especies de cactáceas. Para ello, en este trabajo se desarrolló un nuevo protocolo de predicción *ab-initio* de genes de interés a partir de la secuenciación PacBio que se obtuvo (con lecturas ~15 kb) de genoma completo no ensamblado de *M. san-angelensis*. Mediante el análisis filogenético se identificaron los ortólogos de los *PLT* en *M. san-angelensis*, y se hallaron discrepancias en cuanto al número de ortólogos que se han reportado en otras cactáceas. Para validar los ensamblajes y las predicciones bioinformáticas se amplificaron por PCR productos que corresponden a algunos de los genes identificados bioinformáticamente, confirmando en algunos casos las predicciones del tamaño de secuencias genómicas y de transcritos, tanto en *M. san-angelensis* como en otras cinco especies de *Mammillaria*. De esta forma se ha sentado una base para el análisis del papel de los genes *PLT* en los diversos procesos que tienen lugar en la raíz de las cactáceas, usando como modelo a *M. san-angelensis*.

Abstract

In *Arabidopsis thaliana* there are six *PLETHORA* genes (*PLT*) encoding transcription factors that belong to the *AP2/EREBP* gene family. The *AP2/EREBP* gene family arose before the divergence of land plants and its members are highly conserved. *PLT* proteins play a central role in root development by regulating processes like rhizotaxy, cell differentiation, and maintenance of the Root Apical Meristem (RAM). The vast majority of plants analyzed from the subfamily Cactoideae show determinate growth of the primary root. This phenomenon consists of the depletion of the RAM of the primary root which leads to the differentiation of all the cells in the root apex and, as a consequence, to the loss of the growth of the primary root. Determinate primary root growth can also promote branching of the root system giving place to the development of a fibrous root system. The genetic mechanisms that regulate this process have not been elucidated. Phylogenomic studies of *PLT* genes indicate that the presence of homologues of *PLT1* and *PLT2* in dicots and their absence in monocots might be associated with the root system architecture difference between monocots and dicots. Likewise, double *plt1 plt2* *A. thaliana* mutants show phenotypes with root development morphologically similar to the

natural development of cactus roots. Furthermore, the study of the root apex transcriptome in *Pachycereus pringlei* (Cactaceae) with an active or depleted RAM showed that the transcriptional programs operating in this plant's root apex are similar to those operating in *A. thaliana*'s. The plants of the genus *Mammillaria* are an interesting model for the study of the determinate growth of the primary root in cacti because they have a relatively short life cycle, high diversity and are easy to reproduce. However, exploration of this phenomenon in *Mammillaria* is limited by our lack of understanding regarding the genes that may be regulating this process. In this study we carry out the identification of putative orthologues of *A. thaliana* *PLT* genes in *Mammillaria san-angelensis* using bioinformatic and molecular methods to expand our knowledge of RAM depletion in cacti species. To this end we developed a new protocol for the *ab-initio* prediction of genes of interest from our whole genome unassembled PacBio sequencing reads (with an average read length ~15 kb) of *M. san-angelensis*. We conducted phylogenetic analyses to identify homologues of *PLTs* in *M. san-angelensis*. Discrepancies were found regarding the number of homologues that have been reported in other cacti. To validate the assemblies and predictions, PCR products that correspond to some of the genes identified bioinformatically were amplified, confirming in some cases the predictions of the size of genomic sequences and transcripts, both in *M. san-angelensis* and in other five species of *Mammillaria*. In this way, the bases have been laid for the analysis of the role of *PLT* genes in the various processes that take place in the root of cacti, using *M. san-angelensis* as a model plant.

1. Introducción

La familia Cactaceae es un grupo de plantas dicotiledóneas perteneciente al orden de las Caryophyllales. Esta familia se originó en Sudamérica, con una diversificación fuerte durante el Mioceno tardío. Su distribución abarca una gran variedad de ecosistemas. Las cactáceas presentan una serie de características particulares y novedades evolutivas que las distinguen de otros grupos (Guerrero et al., 2019).

Entre dichas novedades evolutivas destaca la presencia de aréolas, sinapomorfía que caracteriza a la familia. Las areolas son meristemas axilares especializados de los cuales surgen las espinas y que, en algunos taxa, presentan superficialmente tricomas o glóquidas (Simpson, 2010). Tres linajes amplios (subfamilias) de cactáceas carecen de hojas pecioladas, de hecho, en su madurez, la mayoría de las especies parecen carecer completamente de hojas. Por otro lado, muchas especies presentan tallos ensanchados fotosintéticos que adoptan una gran variedad de formas y tamaños. Sus flores y frutos son otras estructuras notables pues presentan un pericarpelo, estructura aparentemente vegetativa que rodea una porción del receptáculo que contiene el ovario, y un perianto que muestra un gradiente morfológico que va de brácteas a estructuras similares a pétalos (Rosas-Reinhold et al., 2021).

A esta variedad de fenómenos del desarrollo vegetal es importante agregar el particular desarrollo de raíz que observamos en la gran mayoría de las especies analizadas dentro de la subfamilia Cactoideae: el crecimiento determinado de la raíz primaria. Este fenómeno consiste en el agotamiento del meristemo apical de la raíz primaria lo cual da lugar a la diferenciación de todas las células del ápice y, como consecuencia, a la detención del crecimiento de dicha raíz, así como a la ramificación del sistema radical, pudiendo originar un sistema de raíces de apariencia fibrosa (Shishkova et al., 2013).

Se ha descrito ampliamente la morfología de las cactáceas, las características histológicas de sus estructuras y la diversidad taxonómica que existe en ellas. Sin embargo, se conoce muy poco acerca de los mecanismos genéticos que dan lugar

a fenómenos del desarrollo en estas plantas. En estudios previos fue posible hallar algunas pistas sobre los mecanismos que pueden estar regulando el crecimiento determinado de la raíz primaria en cactáceas (Kerstens et al., 2020; Rodríguez-Alonso et al., 2018). Se ha señalado que la presencia de los ortólogos de los genes *PLETHORA1 (PLT1)* y *PLT2* en dicotiledóneas y su ausencia en monocotiledóneas pueden ser lo que determina que las dicotiledóneas desarrollen una raíz pivotante y las monocotiledóneas un sistema de raíces fibroso. Asimismo, estudios con mutantes dobles de pérdida de función de *plt1plt2* en *A. thaliana* describen fenotipos de desarrollo de raíz morfológicamente similar al desarrollo natural de las raíces de cactáceas. Además, el estudio del transcriptoma del ápice de raíz de *Pachycereus pringlei* (Cactaceae), con RAM activo o exhausto, mostró que los programas transcripcionales que operan en el ápice de la raíz son similares en esta especie y en *A. thaliana* ortólogo (Rodríguez-Alonso et al., 2018).

Los seis genes *PLT* que existen en *A. thaliana* codifican para factores de transcripción con un papel central en el desarrollo de raíz. Son parte de la superfamilia AP2/EREBP. Las proteínas PLT regulan la rizotaxia, filotaxia, la diferenciación celular, y el mantenimiento del RAM entre otros procesos (Horstman et al., 2014).

La familia AP2/EREBP se divide en cuatro subfamilias, la subfamilia DREB (Dehydration Responsive Element-Binding), la subfamilia RAV (Related to ABI3/VP), la subfamilia EREBP (Ethylene Responsive Element Binding Protein), y la subfamilia AP2. Los genes de la subfamilia AP2 cuentan con regiones que codifican dos dominios AP2/ERF. Los factores de transcripción de la subfamilia AP2 funcionan como reguladores esenciales en el desarrollo de órganos reproductivos y vegetativos; en esta subfamilia encontramos a los genes *PLETHORA (PLT)*, *APETALA1 (AP1)*, *AINTEGUMENTA (ANT)*, entre otros (Feng et al., 2020).

Los mecanismos genéticos que regulan el agotamiento del RAM en la subfamilia Cactoideae no han sido descritos, y el papel que los genes PLT juegan en este proceso sigue siendo una incógnita. La exploración de este fenómeno se ve limitada por diversos factores entre los que hallamos: la escasez de genes identificados que

pueden estar regulando este proceso en plantas en las que el agotamiento del RAM es un fenómeno constitutivo, la falta de estudios que vinculen lo descubierto en plantas modelo como *A. thaliana* y plantas no modelo, así como las dificultades inherentes al trabajo con cactáceas, las cuales presentan agotamiento del RAM, más son a menudo de difícil cultivo, gran talla, y tiempo generacional que puede ir de los años a las décadas.

Las plantas del género *Mammillaria* pueden ser un modelo interesante para el estudio del crecimiento determinado de la raíz primaria en cactáceas, pues presentan un ciclo de vida relativamente corto, alta diversidad y son de fácil reproducción. Por ello, en este estudio se optó por llevar a cabo la identificación de los ortólogos putativos de los genes *PLT* de *A. thaliana* en *Mammillaria san-angelensis* utilizando métodos bioinformáticos y moleculares con la finalidad de expandir nuestro conocimiento del agotamiento del RAM en especies de cactáceas.

En este escrito reportamos nuestros esfuerzos para la generación de recursos genómicos en *M. san-angelensis*, el desarrollo de herramientas para la búsqueda eficiente de genes en el genoma de esta especie, así como el uso de herramientas moleculares para validar nuestro trabajo bioinformático. Finalmente, los resultados obtenidos se discutieron a la luz de la información disponible sobre otras especies de cactáceas, y de monocotiledóneas/dicotiledóneas, sentando las bases para el estudio funcional de los genes *PLT* en los diversos procesos que tienen lugar en la raíz de las cactáceas, usando como modelo a *M. san-angelensis*.

2. Antecedentes

2.1 El desarrollo de la raíz

El desarrollo de la raíz se puede dividir en dos etapas: a) desarrollo embrionario y b) desarrollo post-embrionario. Durante el desarrollo embrionario el óvulo fertilizado se convierte en un embrión. En el embrión, se establecen los meristemos primarios, los ejes corporales, y las principales capas de tejidos. El desarrollo post-embrionario comienza durante la germinación. Durante este proceso la radícula, o la raíz

embrionaria, sale de la semilla por el proceso del crecimiento celular y así se convierte en la raíz primaria. Enseguida se activan los meristemos apicales del brote y de la raíz. En etapas posteriores la actividad del RAM, la elongación celular, y la formación *de novo* de meristemos dentro de los primordios de raíces laterales, aumentan la ramificación y la funcionalidad del sistema radical (Sánchez-Calderón, 2013).

La actividad de los meristemos, junto con la generación nuevos meristemos a lo largo de la planta, permite la formación de nuevos órganos. Gracias a su capacidad de dar origen a nuevos órganos las plantas poseen un alto grado de plasticidad en el desarrollo, mismo que se usa para responder a factores externos y afecta directamente la adecuación del organismo (Péret et al., 2014). A grandes rasgos el sistema radical se compone de: una raíz primaria establecida durante la embriogénesis a partir de la radícula, raíces laterales iniciadas post-embrionalmente a partir de tejidos vasculares en la raíz, y raíces adventicias que se forman a partir de tejidos del vástago. En conjunto, este sistema permite aumentar el sostén mecánico de la planta, ampliar su rizosfera, etc. (Nibau et al., 2008; Steffens & Rasmussen, 2016).

A la configuración espacial del sistema radical (número y longitud de órganos laterales), se le denomina arquitectura de raíz. Dicha arquitectura es el resultado de la interacción entre mecanismos reguladores endógenos y estímulos externos tanto bióticos como abióticos (Malamy, 2005). Entre los mecanismos internos se encuentra una red de mecanismos que involucran diversas moléculas de señalización como fitohormonas y factores de transcripción. Esta red es la encargada de regular el destino celular y la diferenciación de las células iniciales (Dirsch & Stahl, 2015). Conocemos como células iniciales a cuatro tipos de células troncales que dan origen a las filas de células que formarán los tejidos de la raíz. Las células iniciales se dividen para dar origen a células hijas cuya diferenciación da origen a patrones de arquitectura complejos (Scheres et al., 2002).

2.2 Estructura de la raíz

Generalmente se usa como planta modelo a *A. thaliana* para hablar de desarrollo debido a la cantidad de información existente acerca del control genético de su desarrollo. La raíz de las plantas de esta especie se estructura de la siguiente forma: transversalmente iniciando desde el centro podemos distinguir al haz vascular, periciclo, endodermis, córtex, y epidermis (Figura 1). En el ápice de la raíz podemos observar el meristemo apical de la raíz; cubierto por la cofia lateral y la columela (Taiz & Zeiger, 2003).

Longitudinalmente podemos distinguir 3 zonas en el ápice de la raíz primaria. Desde la punta hacia la base del vástago encontramos: La zona meristemática, la zona de elongación y la zona de diferenciación (Ivanov & Dubrovsky, 2013; Verbelen et al., 2006). En la zona meristemática se producen las nuevas células que permiten el crecimiento de la raíz. La zona meristemática está compuesta por: El nicho de células troncales (SCN), el dominio de proliferación y el dominio de transición. El SCN a su vez está conformado por el centro quiescente (QC), una región de baja división celular vital para el mantenimiento de la raíz; y las células iniciales, una población de células que se dividen para renovarse y dar lugar a las células que originan los tejidos de la raíz (Ivanov & Dubrovsky, 2013; Verbelen et al., 2006; Zluhan-Martínez et al., 2021). La zona de elongación está caracterizada por células que se alargan rápidamente, exhiben endoreduplicación y debilitamiento de la pared celular (Scheres et al., 2002). Finalmente, en la zona de diferenciación las células adquieren las características finales de su tipo celular, como la presencia de pelos radicales epidérmicos (Scheres et al., 2002).

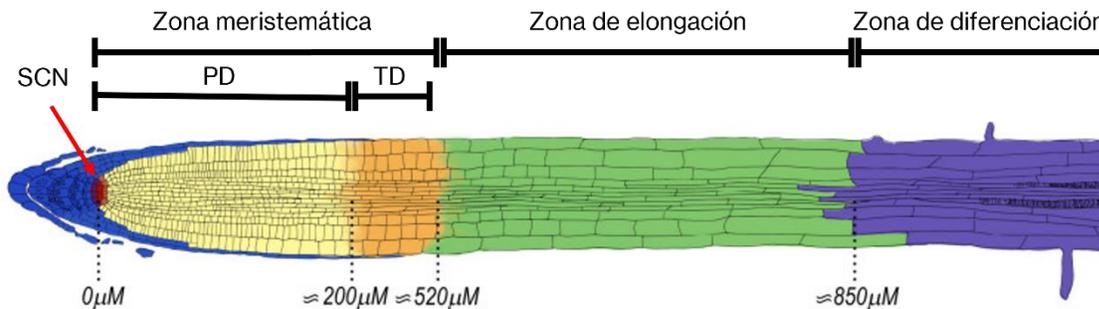


Figura 1. Zonas y dominios en el ápice de la raíz de *A. thaliana*. Se marca la cofia (azul), el nicho de células troncales (rojo), el dominio de proliferación (PD, amarillo), el dominio de transición (TD, naranja), la zona de elongación (verde) y la zona de diferenciación (morado). Se indica la distancia aproximada desde el QC en la raíz primaria 5 días después de la germinación según lo reportado por Verbelen et al., 2006. Adaptado de (Zluhan-Martínez et al., 2021).

2.3 El Meristemo Apical de la Raíz (RAM, *Root Apical Meristem*)

En la ápice de la gran mayoría de las raíces de plantas angiospermas se encuentra la cofia, estructura que da protección al RAM (Jiang & Feldman, 2005) Al igual que los meristemos que se encuentran en el vástago, el RAM alberga células indiferenciadas que generan nuevas células. En *A. thaliana* cuatro a seis células suelen formar el QC, alrededor de las cuales se encuentran las células iniciales, o troncales. Las células del QC mantienen a las células iniciales en un estado indiferenciado; estas últimas al dividirse dan lugar a una célula hija inicial y la otra célula hija pasará al dominio de proliferación (Figura 2, van den Berg et al., 1997). Así las células del RAM producen los tres sistemas de tejidos de la raíz: dérmico, vascular y fundamental. Los nuevos RAM surgen dentro de los primordios de raíces laterales, o los primordios de raíces adventicias en el vástago de la planta (Gilbert et al., 2001)

La posición dentro de la fila celular que las células adoptan al ser desplazadas del SCN dentro del RAM generalmente sigue siendo la misma durante el crecimiento longitudinal de la raíz y definirá el destino celular de sus descendientes. Así, se generan capas celulares con linajes específicos organizadas concéntricamente que representan un gradiente de desarrollo (Drisch & Stahl, 2015)

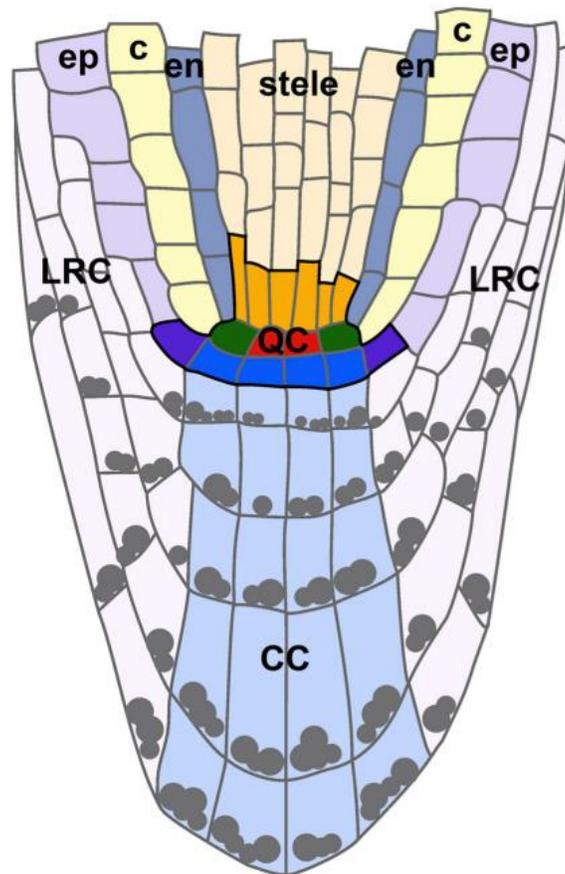


Figura 2. Representación esquemática del ápice de la raíz en *A. thaliana*. El nicho de células troncales (delineado en negro) contiene el centro quiescente QC (rojo), las células iniciales del córtex/ endodermis (verde), las células iniciales del estele (naranja), las células iniciales de la cofia lateral y epidermis (púrpura) y las células iniciales de la columela (azul). ep: epidermis; c: córtex; en: endodermis; LRC: cofia lateral de la raíz; CC: células de la columela (Adaptado de Drisch & Stahl, 2015).

2.4 El crecimiento indeterminado de la raíz

En angiospermas, el crecimiento longitudinal de la raíz se mantiene debido a la actividad del RAM. Sin embargo, en ciertos casos el RAM está programado para dejar de producir células nuevas en un estadio del desarrollo (Sablowski, 2007). La presencia de un meristemo activo por tiempos largos, o determinado, o un

meristemo activo por tiempos cortos y posteriormente consumido (agotado), es decir, determinado, es la diferencia fundamental entre un órgano que se considera determinado o indeterminado (Shishkova et al., 2008). En general, la raíz se considera un órgano con crecimiento indeterminado (Bitonti & Chiappetta, 2011; Kenrick & Strullu-Derrien, 2014; Sena et al., 2009), pues a diferencia de órganos como, por ejemplo, una flor que tendrá un tamaño y forma predecible, la raíz seguirá creciendo a través de la vida de la planta llegando a tener un tamaño y arquitectura variable, dependiente del ambiente local (Sablowski, 2007). Sin embargo, si una raíz individual es un órgano determinado o indeterminado es un tema más complejo. El concepto general aceptado considera que generalmente la raíz es un órgano indeterminado, aunque existen varios ejemplos de raíces de diferentes tipos con crecimiento determinado.

El tiempo de vida de las raíces individuales es altamente variable, va desde unos días hasta algunos años (Eissenstat et al., 2000; Shishkova et al., 2008). Ya que una raíz puede permanecer viva y funcional aun después de que acaba su crecimiento, el criterio para considerarla determinada o indeterminada es la presencia del meristemo apical de raíz, ya sea activo o inactivo (Rodríguez-Alonso et al., 2018). Estudios en *A. thaliana*, *Gossypium hirsutum*, y representantes de cinco familias de plantas (*Clarkia unguiculata*, *Oxalis laxus*, *Oxalis corniculata*, *Dianthus caryophyllus*, *Blumenbachia hieronymi*, y *Salvia farinaceae*), proveen evidencia robusta para inferir que la raíz primaria finalmente deja de crecer en todas las dicotiledóneas (Chapman et al., 2002; Reinhardt & Rost, 1995). La determinación de las raíces relacionada al envejecimiento del organismo suele ser acompañada por diversos cambios en el RAM (Shishkova et al., 2008). En el proceso de envejecimiento la organización del RAM cambia, se vuelve más pequeño, las células dentro del mismo se aíslan simplásticamente, y finalmente el RAM deja de funcionar.

El crecimiento de la mayoría de las raíces individuales puede ser dividido en dos fases: fase de crecimiento, y terminación del crecimiento. En la mayoría de las angiospermas durante la fase de crecimiento, cuando el RAM producirá nuevas

células, persiste por tiempos prolongados. Por otro lado, la fase de terminación del crecimiento se refiere al momento en que el crecimiento de la raíz se detiene debido a edad, estadio de desarrollo, tamaño o algún factor ambiental. En este caso un RAM organizado aún se encuentra presente y sus células mantienen potencial de dividirse, entonces, como se describió arriba, a este tipo de crecimiento se refiere como “indeterminado”. Por otro lado, cuando la raíz presenta “crecimiento determinado”, se ejecuta un programa de desarrollo que culmina con el agotamiento completo del RAM y la diferenciación de todas las células en el ápice de la raíz. Decimos que dicho crecimiento determinado es constitutivo cuando ocurre bajo cualquier condición ambiental (Shishkova et al., 2008).

En *A. thaliana* el mantenimiento del RAM está regulado por diversos mecanismos entre los cuales destacan aquellos en que participan los factores de transcripción como WUSCEL-LIKE HOMEBOX 5 (WOX5) (Sarkar et al., 2007), PLT1, PLT2, PLT3, y PLT4 (Galinha et al., 2007), SHORT-ROOT (SHR) (Nakajima et al., 2001), SCARECROW (SCR) (Sabatini et al., 2003), SPATULA (SPT) (Makkena & Lamb, 2013), TARGET OF MP 7 (TMO7) (Schlereth et al., 2010), UPBEAT1 (UPB1) (Tsukagoshi et al., 2010), NO TRANSMITTING TRACT (NTT) (Crawford et al., 2015), BREVIS RADIX (BRX) (Mouchel et al., 2004), FEZ y SOMBRERO (SMB) (Bennett et al., 2014).

2.5 Las raíces en la familia Cactaceae

En la subfamilia Cactoideae perteneciente a la familia Cactaceae se ha documentado el crecimiento determinado de la raíz primaria en la mayoría de las especies analizadas. Este fenómeno representa un programa de desarrollo estable con un valor ecológico (Shishkova et al., 2013). Una característica común es la duración relativamente corta del crecimiento de la raíz primaria como consecuencia de lagotamiento temprano del RAM. Lo anterior junto con la diferenciación de todas las células previamente meristemáticas, constituye un programa de desarrollo altamente estable, con variaciones temporales menores dentro de la misma especie dependientes de factores ambientales (Shishkova et al., 2008).

Un programa de desarrollo establecido y presente en la gran mayoría de miembros de una familia a menudo es una novedad evolutiva que posee una relevancia ecológica (Shishkova et al., 2013). El establecimiento rápido de plántulas de cactáceas en ambientes con cortos periodos de humedad óptima es un reto. Se ha propuesto que el crecimiento determinado de la raíz es una adaptación en esta subfamilia (Dubrovsky et al., 1998). El inicio del agotamiento del RAM en algunas especies está relacionado con el desarrollo de raíces laterales. Es decir, podemos interpretar este proceso como una decapitación fisiológica del RAM que promueve la ramificación de la raíz. Este fenómeno resulta en un sistema de raíces altamente ramificado y compacto que permite hacer más eficiente la absorción de nutrientes y agua, y facilita la acumulación de biomasa en el tallo, un factor crítico para la supervivencia de estas plantas en las condiciones desérticas en que se desarrollan. Por lo tanto, la determinación de la raíz primaria en la subfamilia Cactoideae podría incrementar la adecuación de las especies (Shishkova et al., 2008); sin embargo, cómo se regula este proceso no es conocido.

2.6 Vías que regulan el mantenimiento o agotamiento del meristemo apical de la raíz

Se han identificado una gran variedad de mecanismos que regulan la actividad del RAM. En algunos de ellos la tasa de división celular puede ser significativamente disminuida mediante limitaciones nutricionales (Xiong et al., 2013), modificaciones al estado redox celular (Jiang et al., 2016), o mediante modificación a las concentraciones celulares de aminoácidos (Wang et al., 2015).

Algunos de los mecanismos que modulan la actividad del RAM también están involucrados en su mantenimiento. Estos mecanismos suelen estar mediados por factores de transcripción como son los de la familia GRAS: SHR (Nakajma et al., 2001) y SCR (Sabatini et al., 2003). También podemos agregar a la lista a WOX 5 perteneciente a la familia WUSCHEL HOMEBOX (WOX) (Sarkar et al., 2007), a SPT, TMO7, y UPB1 pertenecientes a la familia bHLH (Makkena & Lamb, 2013; Schlereth et al., 2010; Tsukagoshi et al., 2010) y a los factores con dominio NAC (FEZ y SMB)(Bennett et al., 2014). Destaquemos a los miembros de la familia AP2

de factores de transcripción. Esta familia es importante en la regulación del RAM, particularmente las proteínas PLT1, PLT2, PLT3, y PLT4 que han demostrado ser importantes para el mantenimiento del crecimiento indeterminado de la raíz en la planta modelo *A. thaliana* (Aida et al., 2004; Galinha et al., 2007). En esta especie, se ha observado que el meristemo se consume y la raíz pierde su indeterminación bajo las siguientes condiciones: en el mutante de pérdida de función *shr* (Benfey et al., 1993), en *scr* (Sabatini et al., 2003), en el mutante doble *plt1 plt2* (Aida et al., 2004), y cuando *PLT2* es regulado a la baja de manera inducible en la mutante triple *plt1 plt3 plt4* (Mähönen et al., 2014). Estos resultados, entre otros, nos llevan a concluir que la participación de estos factores genéticos es esencial para el mantenimiento del RAM y el crecimiento indeterminado de la raíz.

2.7 Los genes *PLETHORA*

Durante el desarrollo, los organismos animales se valen de moléculas que generan respuestas celulares de diferenciación y crecimiento, de forma dosis-dependiente, a lo largo de una distribución en gradiente. Sistemas similares se han observado recientemente en plantas. Un ejemplo de esto es el efecto de los factores de transcripción *PLETHORA*. En el ápice de la raíz, la alta expresión de *PLT1* y *PLT2* se correlaciona fuertemente con concentraciones máximas de la hormona vegetal auxina. Se ha demostrado que este máximo de auxina tiene una profunda actividad organizadora, una propiedad a menudo asociada con fuentes de gradientes instructivos en animales (Galinha et al., 2007).

Los genes *PLT* pertenecen a la familia multigénica *AP2 (APETALA2)/EREBP*. Dicho grupo está compuesto por genes altamente divergentes, que presentan regiones que codifican para al menos un dominio *AP2/ERF*.

La familia *AP2/EREBP* se divide en cuatro subfamilias: Las subfamilias *DREB* (Dehydration Responsive Element-Binding) y *EREBP* que codifican para productos que presentan un único dominio *AP2/ERF*, la subfamilia *RAV* (Related to *ABI3/VP*) cuyos miembros codifican para productos que presentan un dominio *AP2/ERF* y un dominio *B3*, y finalmente, la subfamilia *AP2* en la cual los genes codifican para

productos que cuenta con dos dominios AP2/ERF; los genes en esta subfamilia funcionan como reguladores esenciales en el desarrollo de órganos reproductivos y vegetativos (Feng et al., 2020; Sakuma et al., 2002; Shigyo et al., 2006).

La subfamilia AP2 se divide a su vez en el linaje EuAP2, que se caracteriza por un motivo de unión miR172 en el extremo C-terminal; y el linaje AINTEGUMENTA (ANT), que se caracteriza por inserciones de aminoácidos característicos en el primer y segundo dominio AP2 / ERF. El linaje ANT se divide en los linajes basalANT y euANT/AIL debido a una serie de inserciones de aminoácidos conservados en el primer dominio AP2/ERF (euANT1) y la región N-terminal (euANT2–4) de las proteínas euANT/AIL (Figura 3). Estas características están altamente conservadas en los AIL, tanto inter como intra-específicamente. Cabe señalar que las proteínas AIL funcionan exclusivamente en vías relacionadas con el desarrollo, no en respuesta al estrés (Horstman et al., 2014).

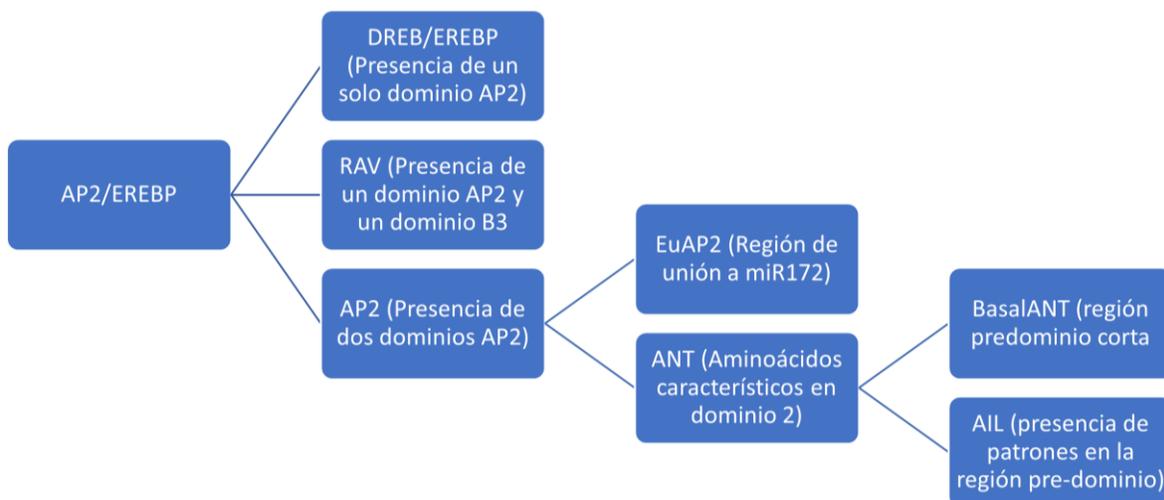


Figura 3. Características estructurales y clasificación de las proteínas AIL en plantas (Adaptado de Kim et al., 2006, Feng et al., 2020)

Los ocho genes *AINTEGUMENTA-LIKE* (AIL) que codifican para factores de transcripción dentro del genoma de *A. thaliana* incluyen *AINTEGUMENTA* (ANT), *AINTEGUMENTA-LIKE* (AIL1) y seis genes *PLT*: *AIL3/PLT1*, *AIL4/PLT2*, *AIL6/PLT3*, *AIL2/BBM/PLT4*, *AIL5/PLT5*, y *AIL7/PLT7*. Estos genes tienen papeles redundantes en el establecimiento y mantenimiento de los meristemas, así como en

la organogénesis. Se ha demostrado que las proteínas AIL son reguladores clave de estos procesos de desarrollo. El papel central de esta familia se refleja en su conservación en un grupo filogenéticamente amplio de organismos fotosintéticos, incluyendo musgos, algas, gimnospermas y angiospermas (Horstman et al., 2014).

En las proteínas codificadas por los genes de este grupo, cada uno de los dominios AP2/ERF se compone de una secuencia repetida de entre 60 y 70 aminoácidos, esencial para las funciones de las proteínas de la familia AP2/EREBP y confiere actividad de unión al DNA de manera secuencia específica. Aunque aún no se han identificado los residuos exactos de aminoácidos dentro del dominio AP2 necesarios para la unión al DNA, las comparaciones de secuencias han revelado dos motivos (secuencias de aminoácidos) altamente conservados dentro del dominio AP2 (Feng et al., 2020; Okamuro et al., 1997; Shigyo et al., 2006)

Dentro de cada dominio AP2 se encuentran dos bloques que caracterizan la familia génica. El primer bloque, conocido como el elemento YRG, consta de 19-22 aminoácidos, es altamente básico (está cargado positivamente) y contiene un motivo de aminoácidos YRG conservados. El segundo bloque, conocido como el elemento RAYD, tiene una longitud de 42-43 aminoácidos y contiene una región central de 18 aminoácidos altamente conservada que se predice que forma una hélice anfipática en los dominios AP2 (Feng et al., 2020; Okamuro et al., 1997; Shigyo et al., 2006).

2.8 Papel de los genes *PLT* en el desarrollo vegetal

Las proteínas PLT forman gradientes desde una zona de máxima transcripción en el SCN y menor tasa de transcripción en la parte distal del dominio de proliferación; varios procesos, como el movimiento intercelular a corta distancia y la dilución ligada a la división y elongación celular, contribuyen en la formación de este gradiente (Mähönen et al., 2014). La expresión de genes PLT, se restringe a tejidos en desarrollo en embriones, raíces y brotes; los factores de transcripción PLT participan en la regulación del mantenimiento de las zonas meristemáticas, la proliferación celular, embriogénesis, crecimiento celular, la filotaxia, la rizotaxia, etc.(Aida et al.,

2004; Galinha et al., 2007; Hofhuis et al., 2013; Krizek & Eaddy, 2012; Prasad et al., 2011).

En la raíz de *A. thaliana*, un clado de cuatro ortólogos de *PLT* es necesario para el desarrollo adecuado de la raíz. Las actividades de las proteínas *PLT* son aditivas y dosis-dependientes. Altos niveles de actividad de *PLT1*, *PLT2*, *PLT3*, y *PLT4* promueven el mantenimiento del SCN, mientras que en zonas con niveles menores de la misma se promueven la actividad mitótica, en zonas con niveles aún más bajos se da la elongación celular y con niveles casi nulos se da la diferenciación (Galinha et al., 2007).

PLT1 y *PLT2* controlan diversos aspectos del desarrollo de raíz de forma dosis-dependiente a través de gradientes de expresión de *PLT* que culminan en el SCN. Los genes *PLT1* y *PLT2* de *A. thaliana* codifican factores de transcripción esenciales para definir el SCN en la raíz. Mutantes dobles *plt1plt2* muestran pérdida de células troncales, pérdida de células amplificadoras de tránsito y expansión celular reducida. La expresión de *PLT1* y *PLT2* está fuertemente correlacionada con una respuesta transcripcional máxima de la fitohormona auxina en el ápice de la raíz; dicho máximo ha sido demostrado como provocador de actividad organizadora, una propiedad relacionada a fuentes de gradientes instructivos (Galinha et al., 2007).

En *A. thaliana* se sabe que *PLT3*, *PLT5* y *PLT7* son factores clave para la iniciación y primeros etapas de desarrollo de primordios de raíces laterales (Hofhuis et al., 2013). Du y Scheres, (2017) demostraron que la actividad “temprana” de estos tres *PLT*, que actúan de forma redundante, es necesaria para las divisiones formativas de las capas celulares interna y externa en las primeras etapas de desarrollo de los primordios de raíz. Además, descubrieron que estos factores son necesarios para la activación posterior de los genes *PLT1*, *PLT2* y *PLT4* necesarios para el establecimiento *de novo* de los RAM.

Los factores codificados por *PLT3*, *PLT5* y *PLT7* también regulan la regeneración *de novo* de brotes en *A. thaliana*. Estos genes actúan de forma redundante para controlar: los reguladores de pluripotencia *PLT1* y *PLT2*, así como factores de promoción de formación de brotes (como *CUC2*), permitiendo así la regeneración

de brotes(Kareem et al., 2015). En resumen, las proteínas PLT tienen un papel central en el mantenimiento de la raíz.

2.9 Los genes *PLT* y el crecimiento de la raíz

La participación de los genes *PLT* en el mantenimiento del RAM y en su determinación, está ampliamente documentada en *A. thaliana*: dentro del RAM la acción coordinada de *PLT1*, *PLT2*, *PLT3* y *PLT4* promueve la expresión de genes implicados en la proliferación celular, y los genes implicados en la diferenciación son reprimidos (Santuari et al., 2016). Por otro lado, cuando la expresión de *PLT2* está reprimida por RNAi inducible en la mutante triple *plt1 plt3 plt4*, el RAM se agota y la raíz pierde el crecimiento indeterminado (Mähönen et al., 2014). Tanto los mutantes individuales *plt1* como *plt2* muestran una ligera pero notable reducción en la tasa de crecimiento de la raíz y en el número de células meristemáticas debido a una producción reducida de células. En los mutantes dobles *plt1 plt2*, el tamaño del RAM disminuye rápidamente y desaparece de forma que todas las células en el ápice de la raíz se diferencian poco después de la germinación. Numerosas raíces laterales de los dobles mutantes *plt1 plt2* también terminan su crecimiento poco tiempo después de su surgimiento. Usando marcadores específicos QC, se ha demostrado que en los mutantes dobles se cambia la identidad de las células del QC (Aida et al., 2004).

Por otro lado, se ha reportado la inducción de estructuras con características radicales de manera ectópica como respuesta a la expresión constitutiva de los genes *PLT* (Aida et al., 2004), y después de la inducción de la expresión de *PLT4* (Srinivasan et al., 2007). Además, experimentos de inducción de la actividad de *PLT2* bajo un promotor constitutivo insertada en una doble mutante *plt1plt2* han mostrado el crecimiento de raíces ectópicas en el ápice de plantas de *A. thaliana* (Figura 4) como respuesta a la activación de *PLT2* (Galinha et al., 2007). Estos experimentos indican que los genes *PLT* son reguladores maestros para el desarrollo de la raíz. En conjunto, la información revisada hasta ahora sugiere que los genes *PLT* son necesarios para mantener la identidad del QC y la actividad del RAM, tanto en la raíz primaria, como en las raíces laterales (Aida et al., 2004). Sin

embargo, no se sabe si cambios en el número de ortólogos de *PLTs* en el genoma, cambios en su patrón de expresión, o cambios en su estructura controlan el programa de determinación de las raíces en otras especies.



Figura 4. Ápice del brote de una planta con la construcción 35S-PLT2-GR 9 días después de la germinación, 6 días después de la inducción de PLT2 con dexametasona. Se observa el crecimiento ectópico de raíz en el ápice de la planta (Imagen adaptada de Galinha et al., 2007)

2.10 Avances recientes en el estudio de los genes *PLETHORA* y su papel en el desarrollo de raíz en especies no modelo

Desde hace ya más de una década, la búsqueda y predicción bioinformática de genes se han vuelto herramientas invaluable para el estudio comparativo de familias de genes en diversas especies (Do & Choi, 2006). En casos particulares las herramientas bioinformáticas de predicción de genes han sido utilizadas para la búsqueda a gran escala de genes particulares en genomas completos (Li et al., 2016). Métodos bioinformáticos, filogenómicos y análisis de transcriptoma han sido utilizados previamente para identificar a los ortólogos putativos de los genes *PLT* en los genomas de más de 100 especies de angiospermas (Rodríguez-Alonso et al., 2018; Albarrán, 2020; Franssen et al., 2015; Kerstens et al., 2020).

El uso de datos de diversas especies permite llevar a cabo una diversidad de estudios, entre ellos los de naturaleza filogenómica. En el caso de la familia AP2, los estudios de esta naturaleza han permitido llegar a diversas conclusiones acerca del papel de estos factores de transcripción en la evolución del desarrollo de raíz y en los mecanismos de desarrollo de este órgano (Kerstens et al., 2020; Kim et al.,

2006; Shigyo et al., 2006). Por ejemplo, se ha encontrado que la presencia o ausencia de miembros de la familia PLT podría explicar la diferencia en arquitectura de la raíz en monocotiledóneas y dicotiledóneas (Kerstens et al., 2020). Particularmente en cactáceas, el estudio de Rodríguez-Alonso y colaboradores publicado en 2018 sobre el transcriptoma de *Pachycereus pringlei* sugiere que programas transcripcionales específicos tienen lugar en el ápice de raíz de esta especie en momentos particulares del desarrollo, siendo uno de los componentes principales de estas redes de regulación el bloque de las proteínas PLT de forma similar a como ocurre en *A. thaliana* (Figura 5).

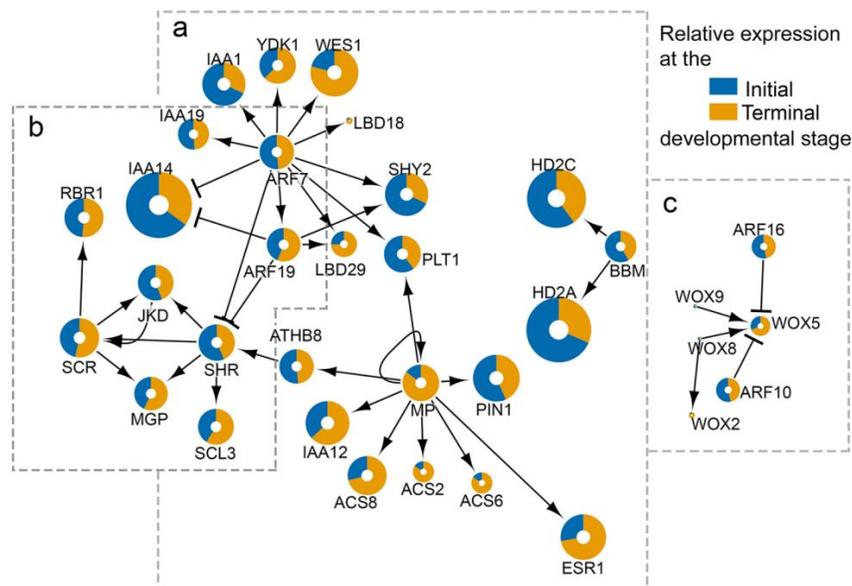


Figura 5. Módulos que controlan el desarrollo de raíz en la red regulatoria inferida para *P. pringlei*. En la figura se observan tres módulos: a) PpPLT, b) PpSHR-PpSCR y c) PPWOX. El tamaño de los nodos representa el nivel de expresión promedio ($\log_2(\text{RPKM})$) del contig en el transcriptoma de *P. pringlei* (Imagen adaptada de Rodríguez-Alonso et al., 2018)

2.11 El género *Mammillaria*

Mammillaria es un género de cactáceas pequeñas usualmente de tallos globosos con tuberculación distintiva. Los tubérculos suelen ser cónicos sin surco y areolas dimórficas en las cuales las flores surgen de la base de los tubérculos en distribución subapical (Anderson, 2001). La mayoría de las especies en este género se encuentran en climas áridos o semiáridos, con una distribución desde el Sur de Estados Unidos hasta Centroamérica. En México podemos encontrar la gran

mayoría de las especies, una fracción importante de estas son endémicas (Hernández & Gómez-Hinostrosa, 2015).

Mammillaria es uno de los géneros más diversos de la familia Cactaceae (Butterworth & Wallace, 2004; Guerrero et al., 2019). Actualmente el número aceptado de especies oscila entre las 164 y 200 (Hunt, Taylor, Nigel, Charles, Graham., International Cactaceae Systematics Group., 2006; Hunt, 2016). Este género se encuentra dentro de la tribu Cactaeae conformada por alrededor de 27 géneros entre los cuales se encuentran *Mammillaria*, *Astrophytum*, *Echinocactus* y *Ferocactus* (Anderson, 2001). Se ha propuesto que *Mammillaria* s.l. no es un grupo monofilético (Butterworth & Wallace, 2004). Los estudios más recientes, de Breslin et al. (2021), proponen el clado Mammilloide; este circunscribe tres géneros monofiléticos: *Mammillaria* s.s., *Coryphantha* y *Cochemiea* s.l.. Se estima que el clado Mammilloide ha divergido durante los últimos 8.62 millones de años (Hernández-Hernández, et al., 2014). Este grupo joven y morfológicamente diverso conforma un marco adecuado para estudiar la evolución, estructura, función y regulación de los genes *PLETHORA* en una escala distinta a las analizadas anteriormente en la familia *AP2/EREBP*.

2.12 Citogenética de *Mammillaria*

Las especies que hasta ahora han sido estudiadas en el género *Mammillaria* son diploides (Del Angel, et al., 2006). Dentro de la serie supertextae, a la cual pertenece *Mammillaria san-angelensis*, los genomas tienen un tamaño que va de los 1489 Mpb (*M. flavicentra*) hasta los 1600 Mpb (*M. san-angelensis*) (Del Angel et al., 2006; Palomino et al., 1999). En cuanto a *M. san-angelensis*, la especie posee 11 pares de cromosomas, y tiene un tamaño de genoma, similar al de otras especies dentro de la serie (Palomino et al., 1999). Es decir, el tamaño del genoma en especies de la serie Supertextae es relativamente estable.

3 Justificación

Las plantas del género *Mammillaria* son un modelo interesante para el estudio del crecimiento determinado de la raíz primaria en cactáceas pues presentan un ciclo de vida relativamente corto, alta diversidad y son de fácil reproducción. Sin embargo, la exploración de este fenómeno en *Mammillaria* se ve limitada por factores relacionados al desarrollo de raíz, la falta de estudios sobre el desarrollo de su raíz primaria la falta de estudios centrados en la morfoanatomía del ápice de raíz, y la escasez de genes identificados que pueden estar regulando este proceso. En este estudio se propone identificar los ortólogos putativos de los genes *PLT* de *A. thaliana* en *Mammillaria san-angelensis* utilizando métodos bioinformáticos y moleculares con el objetivo de generar conocimiento sobre genes que pueden estar relacionados al agotamiento del RAM en especies de cactáceas. Esto con la finalidad de sentar una base para el análisis del papel de los genes *PLT* en los diversos procesos que tienen lugar en la raíz de las cactáceas, usando como modelo a *M. san-angelensis*.

4 Hipótesis

Si el proceso de secuenciación de genoma completo de *Mammillaria san-angelensis* y la posterior identificación informática de genes son un protocolo adecuado para la identificación de ORF presentes en el genoma de dicha especie, entonces será posible respaldar parcialmente los resultados usando métodos filogenéticos y moleculares.

5 Objetivos

5.1 Objetivo general

Identificar los genes ortólogos putativos de los genes *PLT* de *A. thaliana* presentes en el genoma *M. san-angelensis*, y compararlos con secuencias reportadas previamente en cactáceas y otras angiospermas.

5.2 Objetivos particulares

- Diseñar un protocolo para la identificación de genes de interés en conjuntos de datos genómicos de lecturas largas no ensambladas.
- Investigar las relaciones evolutivas entre los genes *PLT* en *M. san-angelensis* con respecto a sus ortólogos putativos en *A. thaliana*.
- Validar la existencia los genes *PLT* identificados bioinformaticamente en *M. san-angelensis*, a través de estudios moleculares experimentales en la misma especie y otras especies del género.

6 Metodología

6.1 Secuenciación del genoma completo de *M. san-angelensis* usando la tecnología PacBio

1.1 Colecta de tejido

Con la ayuda del Biól. José de Jesús González se recolectó la epidermis de 17 individuos de *M. san-angelensis*. Estos individuos fueron plantas adultas que se propagaron *in-vitro* (Método de propagación clonal de plantas en recipientes cerrados bajo condiciones asépticas) a partir de 15 semillas de la especie, según lo detallado en “*In-vitro* mass propagation of the near-extinct *Mammillaria haageana* ssp. *san-angelensis* Sánchez-Mejorada” (Martínez-Vázquez & Rubluo, 1989). El tejido recolectado fue almacenado en ultracongelación a -80°C por no más de 15 días antes de su procesamiento.

1.2 Obtención de gDNA y secuenciación

En el laboratorio del Arizona Genomics Institute (AGI), de la Universidad de Arizona, Arizona, E.U.A., se llevó a cabo la extracción de DNA de alto peso molecular. Posteriormente se llevó a cabo la construcción de biblioteca compatible con el sistema PacBioSequel II. Para la construcción de biblioteca se usó el kit de preparación “SMRTbell Express Template”. Finalmente se llevó a cabo la secuenciación de genoma completo de *M. san-angelensis* con 4 SMRT cells en un secuenciador PacBio Sequel II.

El AGI llevó a cabo el análisis de CCS por cada celda. La empresa Pacific Biosciences define una secuencia CCS de la siguiente forma: “Una secuencia CCS combina múltiples sublecturas de la misma molécula SMRTbell utilizando un modelo estadístico para producir una secuencia de consenso altamente precisa, también llamada lectura HiFi, junto con valores de calidad de las bases.” (Pacific Biosciences. “CCS how to. Disponible en <https://ccs.how/>). La información de CCS así como los *subreads* del proceso fueron puestos a disposición a través de un servidor en línea en formato bam y fastq.

Se llevó a cabo la descarga de datos utilizando la herramienta irods (iRODS Consortium, 2023) disponible para sistemas operativos Linux. Se realizó una suma de chequeo checksum para verificar la integridad de los datos. La descarga fue exitosa y los datos se hallaron íntegros.

6.2 Trabajo bioinformático

Análisis de bases de datos de A. thaliana

Usando las bases de datos Araport version 1.11.6 (Pasha et al., 2020) y TAIR versión 10.1 (Berardini et al., 2015) se hizo la recopilación de datos sobre los genes de interés. Estas bases recolectan y curan información genética y molecular de la especie modelo *A. thaliana*. Para los 6 genes *PLT* de *A. thaliana* se recopilaron nombres alternativos, su función reportada, su longitud, orientación en el genoma, modelo con características del gen, comentario de su función predicha con base a similitud curado de la base de datos UniProt, y la información del tejido, región y organelos en que se expresa, dicha información fue de utilidad para el desarrollo del proyecto más no se presenta en este escrito.

Identificación de ortólogos putativos en Beta vulgaris y Selenicereus undatus

Se utilizó la plataforma PLAZA versión 5.0 (Van Bel et al., 2018) para identificar los posibles ortólogos de los genes *PLT* de *A. thaliana* en el genoma de *Beta vulgaris* (betabel) y *S. undatus* (pitahaya). La plataforma PLAZA es un punto de acceso para estudios genómicos comparativos de grupos vegetales y conjunta datos de diversos proyectos de secuenciación.

El flujo de trabajo comenzó con la selección de un gen de interés para ser usado como sonda. En este trabajo se realizaron ensayos de minado usando como sondas a *PLT1*, *PLT2*, *PLT3*, *PLT4*, y *PLT7* de *A. thaliana*. Usando BLASTP/DIAMOND, la plataforma buscó secuencias similares las cuales mediante TribeMCL (usando los parámetros: mclblastline, l = 2, scheme = \$ y demás parámetros en su condición por default) y Orthofinder (usando los parámetros en su condición por default) se usaron para armar grupos de genes ortólogos mediante agrupamiento de proteínas basado

en secuencias. Posteriormente usando el algoritmo MUSCLE (Multiple Sequence Comparison by Log- Expectation) (Edgar, 2004) y MSA editing PLAZA realizó alineamientos múltiples de secuencias proteicas, los cuales también están disponibles para su análisis. Los árboles filogenéticos se construyeron usando Fastree (bajo un algoritmo de aproximadamente máxima verosimilitud), y Notung 2.6 se usa para enraizar los árboles. PhyD3 se usó para mostrar la representación gráfica de estos resultados.

Finalmente, un análisis de dominios proteicos usando la plataforma Interpro permitió descartar las secuencias que no pertenecían al grupo de genes del linaje EuAP2. De esta forma se seleccionaron ortólogos putativos en *S. undatus* correspondientes a los clados PLT4 y PLT3/7. Las 3 secuencias que se encontraron para *B. vulgaris* fueron conservadas sin ser asignadas previamente a un clado.

Construcción de una base de datos en formato adecuado para BLAST

Se concatenaron los archivos que contenían la información de CCS para así trabajar con una única base de datos. Usando BLAST para línea de comando (NCBI, 2008), se construyó la base de datos necesaria para poder llevar a cabo búsquedas de BLAST contra el conjunto de las lecturas.

Indexado del genoma

Para poder acceder a la información de las secuencias marcadas como hits por BLAST, fue necesario indexar el genoma utilizando la suite de programas samtools (Danecek et al., 2021). Finalmente se utilizó la herramienta faidx para extraer las secuencias de interés desde la base de datos.

BLAST contra las lecturas CCS concatenadas del genoma de *M. san-angelensis*

Se llevó a cabo un tBLASTn en su configuración por default con una matriz de sustitución BLOSUM62, usando como sonda las secuencias de aminoácidos de los ortólogos putativos de *PLT 3*, *PLT4* y *PLT7* previamente minados del genoma de *S. undatus*.

Ensamble de las lecturas

Se utilizó el programa Hifiasm, un ensamblador de-novo para lecturas Hi-Fi de PacBio (Cheng et al., 2021). Para obtener contigs a partir de las lecturas seleccionadas en el proceso anterior fue necesario instalar el programa Anaconda, una distribución de Python (versión 3.10.1) y R (versión 4.1.3) para cómputo científico. Usando esta distribución, se creó un ambiente local que alojaría el programa Hifiasm-0.16.1 (r375) y sus dependencias.

Anotado de las secuencias

Se identificaron los marcos de lectura abierto (ORF) dentro de cada uno de los contigs ensamblados. Para ello se seleccionó la herramienta Augustus, un programa diseñado para la predicción ab-initio de genes en secuencias genómicas de organismos eucariotas (Stanke et al., 2008). Augustus es uno de los mejores programas para la predicción ab-initio y tiene una exactitud de predicción a nivel de exón cercana al 90%. Se ingresaron las secuencias, se utilizaron las condiciones por default y se seleccionó como genoma de referencia el genoma de *A. thaliana*.

Selección de ortólogos putativos de AtPLT

Se usaron las características estructurales (presencia de dominios y patrones de aminoácidos), previamente encontradas en las proteínas PLT de *A. thaliana* y *S. undatus*, para identificar los *Open Reading Frames* (ORF) propuestos por Augustus que son candidatos a ser ortólogos putativos de los genes encontrados en *A. thaliana* (*AtPLT*). Finalmente, un análisis de dominios proteicos utilizando la plataforma en línea InterPro se utilizó para analizar los dominios presentes en las proteínas codificadas por las secuencias minadas.

6.3 Análisis filogenético

Construcción de alineamientos múltiples de secuencias

Se construyeron alineamientos múltiples de secuencias de aminoácidos usando el algoritmo MUSCLE dentro del programa MEGAX. Para estos se utilizaron las secuencias de aminoácidos de las 19 proteínas teóricas de *M. san-angelensis*

producto de los procesos ya mencionados, las secuencias de los seis proteínas AtPLT, las cuales pertenecen al clado euANT, y AtADAP, la cual pertenece al clado basalANT, por lo que se utilizó como grupo externo (una proteína que presenta los dos dominios AP2, pero no presenta los patrones de aminoácidos característicos para los miembros del clado euANT); 3 secuencias correspondientes a homólogos putativos de los *AtPLT* minadas del genoma de *B. vulgaris*, las 3 secuencias minadas del genoma de *S. undatus*; y 5 secuencias del transcriptoma ensamblado *de novo* del ápice de la raíz primaria de *P. pringlei*. Las secuencias de transcritos de *P. pringlei* fueron provistas por la Dra. Svetlana Shishkova y el Dr. Gustavo Rodríguez.

Se generaron tres alineamientos alternativos. En el alineamiento uno se conservaron las secuencias de *S. undatus*, *A. thaliana* y *B. vulgaris* como fueron obtenidas de PLAZA. Las secuencias de aminoácidos de *M. san-angelensis* se usaron tal y como fueron propuestas por el predictor Augustus, y las secuencias de *P. pringlei* fueron usadas tal y como fueron entregadas, producto de un trabajo de análisis de transcriptoma. En el alineamiento dos se utilizó únicamente la región dominio-interdominio-dominio tal como ha sido trabajado anteriormente en la familia AP2/ERF (Kim et al., 2006). El alineamiento tres conserva todos los sitios que no presentaron gap en al menos 50% de las secuencias utilizadas usando como base el alineamiento original.

Predicción de modelo evolutivo

Para los alineamientos obtenidos se llevó a cabo una prueba de selección de modelo evolutivo usando JModelTest (Posada, 2008). Este programa basado en "Phyml", permite la selección estadística de modelos de sustitución de nucleótidos y aminoácidos. El modelo de sustitución de nucleótidos más adecuado para los alineamientos fue el de Jones-Taylor-Thornton con un estimador de la proporción de sitios invariables, optimización de las tasas de sustitución, modelo GAMMA de heterogeneidad de tasas de sustitución (con estimación del parámetro alfa) y uso de frecuencia empírica de bases.

Análisis filogenético

Se realizaron análisis filogenéticos para cada alineamiento usando el método de máxima verosimilitud en la plataforma RAxML y la información del mejor modelo de sustitución previamente obtenida. El soporte de nodos se obtuvo mediante un análisis de Bootstrap de 500 réplicas.

6.4 Elaboración de modelos para los genes identificados

Nomenclatura para los ORF predichos de *M. san-angelensis*

Siguiendo los estándares de nomenclatura sugeridos por TAIR (Plant Gene Nomenclature (arabidopsis.org)) se etiquetó a los ORF predichos usando métodos bioinformáticos de la siguiente forma: MsaPLT**A/B.C** en donde los primeros 3 caracteres se refieren al género y la especie vegetal, los caracteres en mayúscula que les siguen se refieren al gen o en este caso grupo de genes que son sus ortólogos putativos en *A. thaliana*. Finalmente, los dígitos en A y B se utilizan para denotar que este gen pertenece a un clado en el cual se agrupan dos genes, y no pueden ser distinguido como ortólogo putativo de un gen en específico dentro del clado, el dígito C se usa para distinguir entre los distintos genes agrupados en dicho clado. Por ejemplo, uno de los ORF predichos se etiquetó como *MsaPLT1/2.1*

Construcción de modelos genéticos

La construcción de modelos genéticos para cada ORF predicho se llevó a cabo en Geneious versión 2023.0 (<https://www.geneious.com>). Para construir los modelos se incorporaron los datos de las lecturas base, los datos de la predicción de Augustus que incluyeron posibles intrones y exones, así como la traducción de la secuencia a aminoácidos; y la anotación manual de la posición de los dominios AP2.

6.5 Trabajo molecular experimental

Diseño de *primers*

El diseño de *primers* se llevó a cabo utilizando el software Geneious versión 2023.0. Los *primers* diseñados se con del nombre del ORF predicho sobre el cuál se

diseñaron. Al final de nombre de cada juego de *primers* se agregó una D si este juego fue planeado para amplificar únicamente la región dominio-interdominio-dominio. Los juegos donde no se agregó la D buscaban amplificar un producto lo más cercano posible al ORF completo. Las especificaciones de cada uno de los *primers* utilizados se encuentran en el anexo 1. Los pares de *primers* se sintetizaron por la empresa IDT (Integrated DNA Technologies) (<https://www.idtdna.com/>). Los *primers* se recibieron liofilizados, fueron rehidratados y llevados a una concentración de stock 100µM.

Después de tomar las alícuotas necesarias, las soluciones *stock* de *primers* fueron depositados a resguardo en el Laboratorio de Biología de Raíces del Jardín Botánico del Instituto de Biología de la UNAM. Para esto se les dio un número en la base de datos además del nombre con que fueron etiquetados. La base de datos contiene los siguientes detalles: Nombre del ORF, nombre original del primer, T_m, orientación, secuencia, especie para la cual fue diseñado, autor y año. En esta tesis, se hace referencia a los *primers* por el nombre bajo el que fueron ordenados para síntesis.

Aislamiento en gDNA de *M. san-angelensis*

Un individuo maduro (reconocido como tal por haber presentado floración previamente) de *M. san-angelensis* se obtuvo del Jardín Botánico del Instituto de Biología de la UNAM, proporcionado por el programa de cultivo y propagación de cactáceas y suculentas en apoyo a la conservación de especies bajo alguna categoría de riesgo, a cargo del Biól Jerónimo Reyes Santiago. Se retiraron 20 tubérculos de esta planta, de estos se retiró la epidermis con cutícula misma que se dejó en gel de sílica durante una semana.

Para la extracción de gDNA de *M. san-angelensis* se utilizó el producto DNeasy® Plant Pro Kit de la marca Qiagen. Se utilizaron aproximadamente 10mg de tejido epidérmico secado durante una semana en gel de sílica. Las extracciones se llevaron a cabo según lo especificado en el manual provisto por la empresa de manufactura (DNeasy® Plant Pro Kit Handbook QIAGEN). Posteriormente se llevó

a cabo un proceso conocido como “Concentración de DNA eluido” según lo especificado en el manual anteriormente mencionado.

Para las reacciones de PCR se usó un termociclador Veriti™ de 96 pozos de la marca Applied Biosystems™. Las reacciones se llevaron a cabo en un volumen de 10 µL con 0.25U de polimerasa “Platinum II Taq Hot-Start” (Invitrogen), 0.25 µL de dNTPs 10mM, 0.25 µL de cada oligonucleótido 10 µM (Integrated DNA Technologies), 1.5 µL de buffer de reacción 10x (provisto por Invitrogen), 0.6 µL de BSA, 0.4 µL de MgSO4 50mM, y 4 µL del gDNA de *M. san-angelensis*.

El protocolo de termociclador 1 consistió en una desnaturalización inicial a 94 °C por 2 minutos, seguido de 35 ciclos de 15 s de desnaturalización a 94°C, 30 s a 52°C, 4 minutos a 68°C; y una extensión final de 4 minutos a 72°C. Este se utilizó para amplificar productos con los juegos de *primers*: MhPLT1.2.1D, MhPLT1.2.1, MhPLT1.2.2D, MhPLT1.2.2, MhPLT3.7.1D y MhPLT3.7.2D (Tabla 1).

El protocolo de termociclador 2 consistió en una desnaturalización inicial a 94 °C por 2 minutos, seguido de 30 ciclos de 15 s de desnaturalización a 94°C, 30 s a 52°C, 7 minutos a 68°C; y una extensión final de 4 minutos a 72°C. Este se utilizó para amplificar productos con los juegos de *primers*: MhPLT3.7.1, MhPLT 3.7.2, MhPLT5D y MhPLT5 (Tabla 1). El resultado de las reacciones de PCR fue evaluado mediante electroforesis en gel.

Aislamiento de RNA y síntesis de cDNA de especies de *Mammillaria*

Se obtuvo del Jardín Botánico del Instituto de Biología de la UNAM, proporcionadas por el programa de cultivo y propagación de cactáceas y suculentas en apoyo a la conservación de especies bajo alguna categoría de riesgo, a cargo del Biól. Jerónimo Reyes Santiago, y de la colección de cactáceas cuyo responsable es el Dr. Salvador Arias semillas de las siguientes especies: *M. san-angelensis*, *M. coahuilensis*, *M. duwei*, *M. hernandezii*, *M. lasiacantha* y *M. pectinifera*. 50 semillas de cada especie se colocaron en una caja de Petri sobre una servilleta humedecida con 15 mL de solución de captan 1g/L. Las cajas Petri se mantuvieron en posición horizontal para la germinación de semillas en cámara de crecimiento (Percival

Scientific modelo CU22L) con fotoperiodo de 12 horas de luz y 12 horas de oscuridad a 28°C. Se realizaron visitas semanales para observar el avance de la germinación y monitorear posibles contaminaciones. Cuando al menos 10 semillas de cada individuo germinaron se llevó a cabo la extracción de RNA total.

La extracción de RNA se realizó usando el “Spectrum™ Plant Total RNA Kit”. Se colectaron las plántulas de la caja de Petri y se sumergieron inmediatamente en nitrógeno líquido. Se molieron aproximadamente 100mg de plántulas de cada especie usando un mortero previamente esterilizado. Se utilizó nitrógeno líquido para mantener el material congelado en todo momento. A partir del material vegetal molido se realizaron las extracciones según lo especificado en el manual provisto por el fabricante. Las muestras de RNA total aislado fueron almacenadas inmediatamente a -80°C.

La síntesis de cDNA se realizó usando el “ProtoScript® II First Strand cDNA Synthesis Kit”. Este método fue utilizado para sintetizar cDNA de primera hebra a partir del RNA total extraído de plantas del género *Mammillaria*. Las reacciones de síntesis se llevaron a cabo según lo especificado en el manual provisto por el fabricante. Las muestras de cDNA fueron almacenadas a -4°C

Las PCR se llevaron a cabo utilizando un termociclador Veriti™ de 96 pozos de la marca Applied Biosystems™. Las reacciones se prepararon en un volumen de 10 µL con 0.25U de polimerasa “Platinum II Taq Hot-Start” (Invitrogen), 0.25 µL de dNTPs 10mM, 0.25 µL de cada oligonucleótido 10 µM (Integrated DNA Technologies), 1.25 µL de buffer de reacción 10x (provisto por Invitrogen), 0.6 µL de BSA, 0.4 µL de MgSO4 50mM, y 1 µL del cDNA de la especie de *Mammillaria* cuyo material se quería amplificar.

El protocolo de termociclador 3 consistió en una desnaturalización inicial a 94 °C por 2 minutos, seguido de 35 ciclos de 15 s de desnaturalización a 94°C, 30 s a 52°C, 1 minuto a 68°C; y una extensión final de 4 minutos a 72°C. Este se utilizó para amplificar productos con los juegos de *primers*: MhPLT1.2.1D, MhPLT1.2.2D, MhPLT3.7.1D y MhPLT5D (Tabla 1).

El protocolo de termociclador 4 consistió en una desnaturalización inicial a 94 °C por 2 minutos, seguido de 35 ciclos de 15 s de desnaturalización a 94°C, 30 s a 52°C, 2:30 minutos a 68°C; y una extensión final de 4 minutos a 72°C. Este se utilizó para amplificar productos con los juegos de *primers*: MhPLT1.2.1 y MhPLT5 (Tabla 1).

El protocolo de termociclador 5 consistió en una desnaturalización inicial a 94 °C por 2 minutos, seguido de 40 ciclos de 15 s de desnaturalización a 94°C, 30 s a 50°C, 1 minuto a 68°C; y una extensión final de 4 minutos a 72°C. Este se utilizó para amplificar productos con los juegos de *primers*: MhPLT1.2.1D, MhPLT3.7.1D y MhPLT5D (Tabla 1).

El protocolo de termociclador 6 consistió en una desnaturalización inicial a 94 °C por 2 minutos, seguido de 40 ciclos de 15 s de desnaturalización a 94°C, 30 s a 52°C, 2:30 minutos a 68°C; y una extensión final de 4 minutos a 72°C. Este se utilizó para amplificar productos con los juegos de *primers*: MhPLT1.2.1 (Tabla 1).

El protocolo de termociclador 7 consistió en una desnaturalización inicial a 94 °C por 2 minutos, seguido de 40 ciclos de 15 s de desnaturalización a 94°C, 30 s a 54°C, 2:30 minutos a 68°C; y una extensión final de 4 minutos a 72°C. Este se utilizó para amplificar productos con los juegos de *primers*: MhPLT1.2.2D y MhPLT5 (Tabla 1).

Tabla 1. Protocolos de termociclador utilizados

Protocolo	Desnaturalización inicial	Desnaturalización	Alineamiento	Extensión	Ciclos	Extensión final
1	2min a 94°C	15seg a 94°C	30seg a 52°C	4 min a 68°C	35	4 min a 72°C
2	2min a 94°C	15seg a 94°C	30seg a 52°C	7 min a 68°C	30	4 min a 72°C
3	2min a 94°C	15seg a 94°C	30seg a 52°C	1 min a 68°C	35	4 min a 72°C
4	2min a 94°C	15seg a 94°C	30seg a 52°C	2.5 min a 68°C	35	4 min a 72°C
5	2min a 94°C	15seg a 94°C	30seg a 50°C	1 min a 68°C	40	4 min a 72°C
6	2min a 94°C	15seg a 94°C	30seg a 52°C	2.5 min a 68°C	40	4 min a 72°C
7	2min a 94°C	15seg a 94°C	30seg a 54°C	2.5 min a 68°C	40	4 min a 72°C

Electroforesis en gel de agarosa

Los resultados de las reacciones de PCR fueron evaluados mediante electroforesis en gel de agarosa al 1%. Para preparar el gel se utilizó buffer TAE 1x (0.04M tris-acetato y 0.01M EDTA, agarosa, y 1 µL del tinte de gel SYBR™ Safe (Invitrogen) por cada 10mL de solución. Los productos de PCR fueron mezclados con 1 µL de buffer de carga BlueJuice™ (Invitrogen) 10x, y agua hasta llevar la muestra a 10 µL. El ladder PCRBIOLadder II (PCRBIOSYSTEMS) fue corrido junto con las muestras

7. Resultados

7.1 Las secuencias genómicas de *M. san-angelensis*

1. Construcción de la biblioteca de secuencias sin ensamblar

Una biblioteca de alta fidelidad (Hi-Fi) con lecturas de entre 10 kb y 25 kb se construyó en el Arizona Genomics Institute (AGI). Usando ésta se llevó a cabo la secuenciación de genoma completo de *M. san-angelensis*.

Se obtuvieron 7,968,978 lecturas CCS con una longitud promedio de 15,306 pares de bases (Tabla 2). Estas presentaron un *quality value* (QV) promedio de alrededor

de 34, valor relacionado a lecturas de alta calidad (Jiao et al., 2013). Dado el tamaño del genoma 1C en *Mammillaria san-angelensis* de 1.6 Gpb (Palomino et al., 1999), se estimó que la cobertura de la secuenciación fue de 86x.

Tabla 2. Tamaño y calidad de las lecturas PacBio

Código de muestra	Datos brutos (Gb)	Lecturas en bruto	Datos CCS (Gb)	Lecturas CCS	QV promedio	Longitud promedio
C01	416.93	5,624,116	24.48	1,602,389	Q33	15280
D01	381.54	5,084,107	22.54	1,447,992	Q34	15568
E01	433.28	5,623,172	25.32	1,649,327	Q34	15352
F01	418.56	5,453,722	25.16	1,643,722	Q34	15308
B01	427.96	5,207,552	24.42	1,625,548	Q34	15022

2. Búsqueda de ortólogos putativos con BLAST

Como resultado del tBLASTn realizado usando como sonda secuencias de aminoácidos de los ortólogos putativos de *PLT3*, *PLT4* y *PLT7* previamente minados del genoma de *S. undatus* se obtuvieron 9275 hits, de los cuales se seleccionaron 903 usando como criterio un e-value de 0 y un bitscore >50 (Figura 6).

3. Ensamblado y anotación de regiones (intrones, exones y dominios)

Usando el programa Hifiasm se logró construir 39 contigs con una profundidad promedio de 10 y una longitud promedio de 45 kb.

Usando la herramienta Augustus encontramos en los contigs ensamblados 209 posibles marcos de lectura abiertos, interrumpidos o no por los intrones, con longitud de entre 177 y 7419 pares de bases (Figura 6).

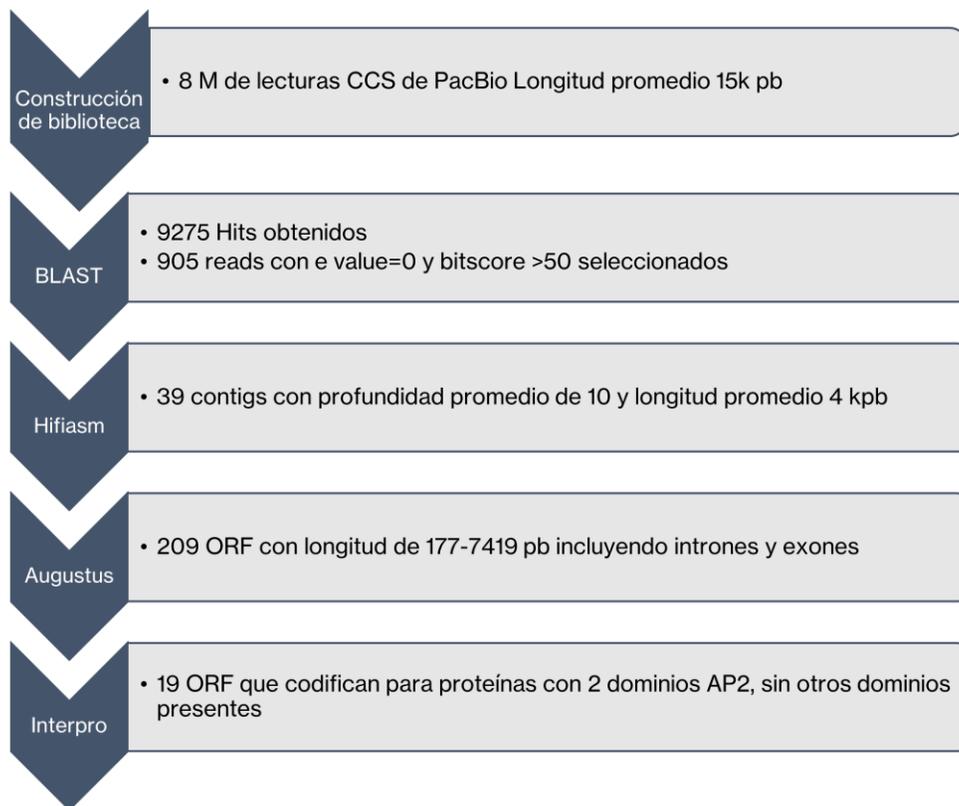


Figura 6. Diagrama de flujo. Resume el tratamiento que se le dio a los datos hasta llegar a las lecturas candidatas a codificar para los ortólogos putativos de AtPLT en el genoma de *M. san-angelensis*. En las flechas se anotan los procedimientos/herramientas que se usaron. Los cuadros denotan el resultado de cada procedimiento.

4. Análisis de las proteínas predichas bioinformáticamente

Se usaron las características estructurales, previamente encontradas en las proteínas PLT (Figura 7), para determinar cuáles de los ORF propuestos por Augustus son candidatos a ser ortólogos putativos de las proteínas PLT de *A. thaliana* (AtPLT). De los 209 posibles ORF, 19 contenían información que codificaba las características estructurales necesarias: la presencia de dos dominios AP2, 4 patrones de aminoácidos (euANT1: NSC[K/R][K/R]EGQ[T/S]R, euANT2, 3 y 4: WLGFSLS, PKLEDFLG, y TFGQR, respectivamente) y una inserción de 1 aminoácido en el segundo dominio AP2. Finalmente, un análisis de dominios utilizando la plataforma en línea InterPro permitió verificar su identidad ya que las 19 proteínas teóricas traducidas de los ORF únicamente presentaban en su estructura dos dominios AP2 de unión al DNA, no otros dominios.

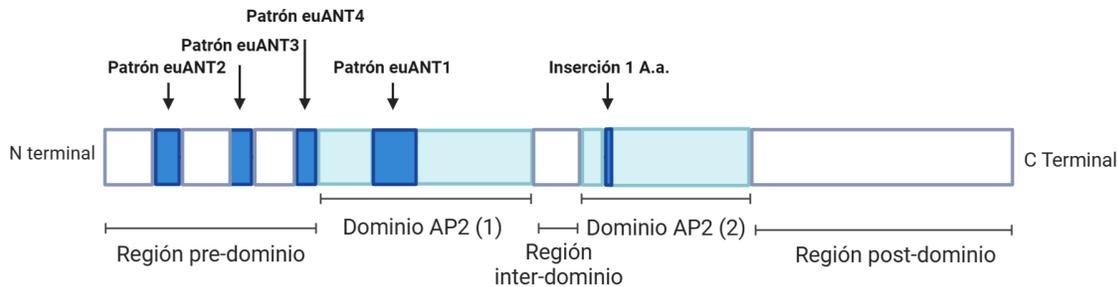


Figura 7. Estructura de las proteínas PLT en *A. thaliana*.

7.2 El análisis filogenético permite reconocer a los ortólogos putativos de *AtPLT1-5*, 7 en el genoma de *M. san-angelensis*

1. Construcción de alineamientos múltiples de secuencias

Se construyeron alineamientos múltiples de secuencias de aminoácidos usando el algoritmo MUSCLE (Edgar, 2004) dentro del programa MEGAX. Para estos alineamientos se utilizaron: las 19 secuencias de aminoácidos de *M. san-angelensis* resultado de los procesos ya mencionados, las secuencias de las seis proteínas PLT y ADAP de *A. thaliana*, 3 secuencias correspondientes a ortólogos putativos de los *AtPLT* minadas del genoma de *B. vulgaris*, 3 secuencias minadas del genoma de *S. undatus*, y 5 secuencias del transcriptoma de *P. pringlei*.

El alineamiento inicial permitió reconocer 13 secuencias redundantes de *M. san-angelensis*. Esto redujo el número de secuencias únicas de este organismo incluidas en los alineamientos múltiples a 6. Se generaron 3 alineamientos alternativos. En el alineamiento uno se conservaron las secuencias de *S. undatus*, *A. thaliana* y *B. vulgaris* como fueron obtenidas de PLAZA. Las secuencias de aminoácidos de *M. san-angelensis* se usaron tal y como fueron propuestas por el predictor Augustus (Figura 8). En el alineamiento dos se utilizó únicamente la región dominio-interdominio-dominio (Figura 9). El alineamiento tres se conservan todos los sitios que no presentaron gap en al menos 50% de las secuencias utilizadas (Figura 10).

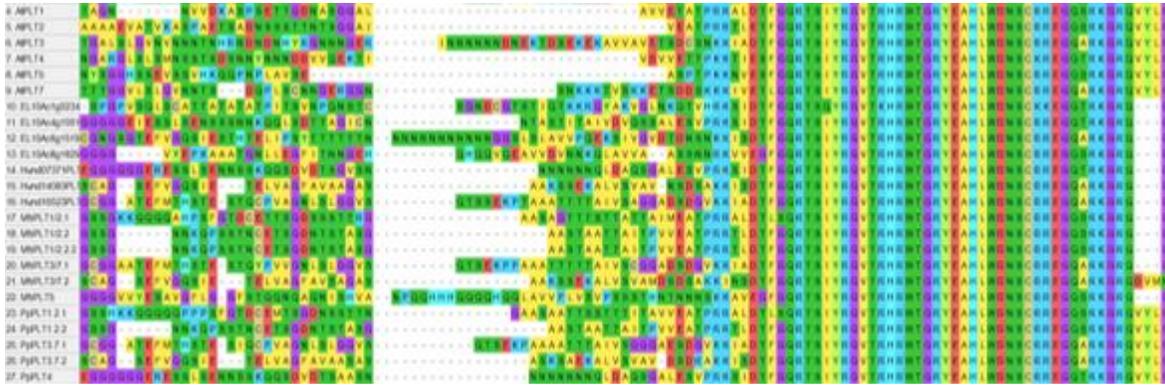


Figura 8. Sección del alineamiento uno. Del lado derecho se muestra la sección amino terminal del primer dominio AP2. En este alineamiento se conservaron las secuencias de *A. thaliana*, *B. vulgaris*, *S. undatus*, y *P. pringlei* como fueron obtenidas de PLAZA. Las secuencias de aminoácidos de *M. san-angelensis* se usaron tal y como fueron propuestas por el predictor Augustus. Los colores indican la polaridad de los aminoácidos.

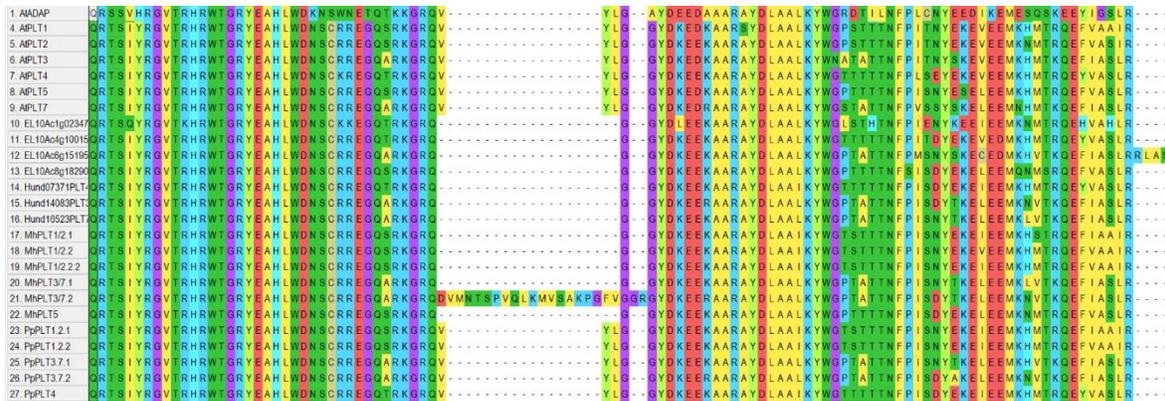


Figura 9. Sección del alineamiento dos. Se muestra el primer dominio AP2 de las secuencias alineadas. En este alineamiento se utilizó únicamente la región dominio-interdominio-domino. Los colores indican la polaridad de los aminoácidos.

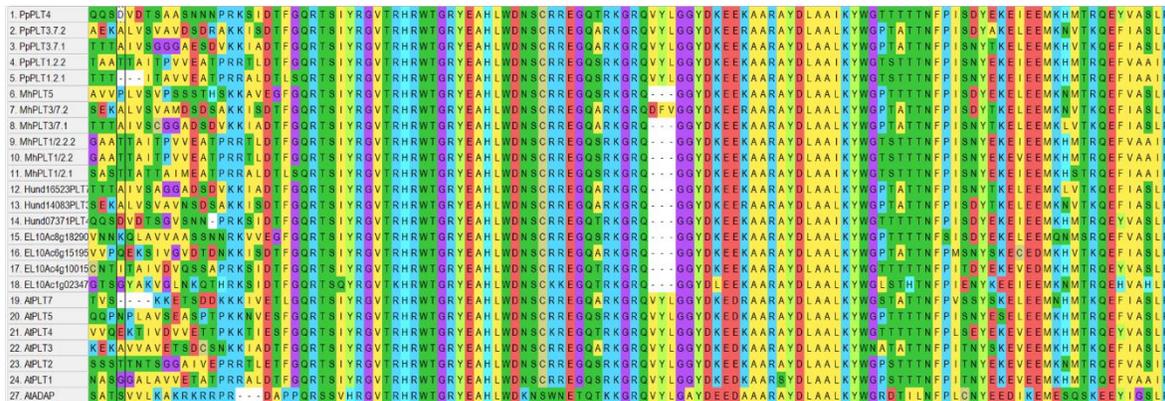


Figura 10. Sección del alineamiento tres. Se muestra una sección del primer dominio AP2 de las secuencias alineadas. En este alineamiento se conservaron todos los sitios que no presentaron gap en al menos 50% de las secuencias utilizadas. Los colores indican la polaridad de los aminoácidos.

2. Análisis filogenético

Se llevó a cabo la reconstrucción de la filogenia molecular de las proteínas PLT, para evaluar la relación que presentaban los ortólogos putativos minados del genoma de *M. san-angelensis* con las proteínas conocidos AtPLT1, AtPLT2, AtPLT3, AtPLT4, AtPLT5, y AtPLT7. En la filogenia reconstruida a partir del alineamiento uno se recupera un clado en el que se encuentra *AtPLT1* y *AtPLT2*, junto con sus ortólogos putativos en el transcriptoma de *P. pringlei*, y el genoma de *M. san-angelensis* (3 genes). Un segundo clado agrupa al gen *AtPLT5* y a su ortólogo putativo en *M. san-angelensis* (1 ORF), así como a un gen de *B. vulgaris*. Un tercer clado contiene a *AtPLT4* y sus ortólogos putativos en *P. pringlei*, *S. undatus* y *B. vulgaris*. Se recuperan de forma dispersa *AtPLT3*, *AtPLT7* y sus ortólogos putativos en cactáceas. Los nodos terminales presentan altos valores de soporte, sin embargo, hacia la base los valores son bajos (figura 11).

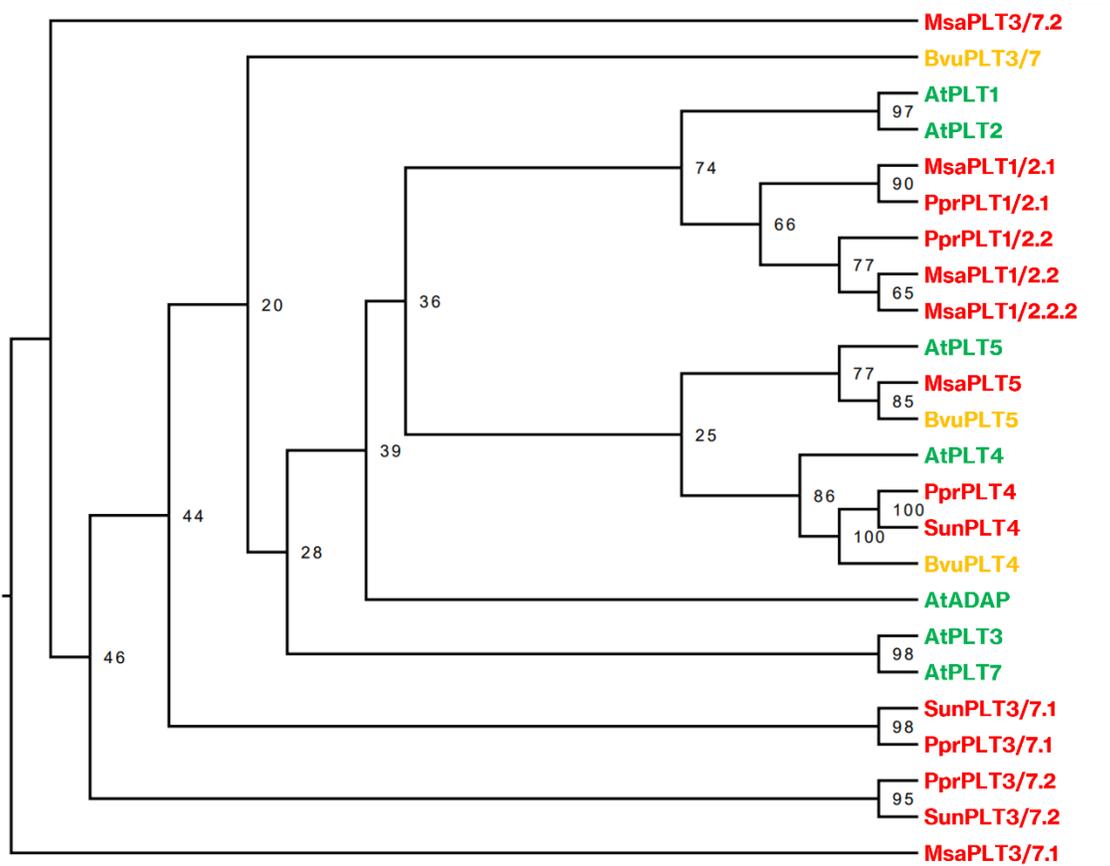


Figura 11. Reconstrucción de la filogenia molecular de las proteínas PLT a partir del alineamiento uno y máxima verosimilitud, enraizada en punto medio. En cada nodo se pueden observar los valores de Bootstrap que le dan soporte al clado. Los nombres de las secuencias pertenecientes a plantas de la familia Cactaceae aparecen en letras de color rojo, los pertenecientes a *B. vulgaris* (familia Amaranthaceae) aparecen en letras amarillas, los pertenecientes a *A. thaliana* (familia Brassicaceae) aparecen en letras verdes. At: *A. thaliana*, Msa: *M. san-angelensis*, Ppr: *P. pringlei*, Sun: *S. undatus* y Bvu: *B. vulgaris*.

En la filogenia molecular de las proteínas PLT reconstruida a partir del alineamiento dos, la menor cantidad de datos utilizados da lugar a diferencias cruciales. *AtPLT5* se encuentra en una terminal aislada, sin formar grupos con otros genes. Se recupera un clado PLT1/2 que contiene a *AtPLT1*, *AtPLT2* y sus ortólogos en *M. san-angelensis* (3 ORF) y *P. pringlei* (2 genes). También se recupera un clado PLT3/7 donde se encuentran *AtPLT3*, *AtPLT7* y sus ortólogos en *M. san-angelensis* (2 genes), *P. pringlei* (2 genes), *S. undatus* (2 genes) y *B. vulgaris* (1 gen) (Figura 12). Se recupera un clado PLT4 con los mismos miembros que en el análisis anterior y cambios en su topología. Los valores de soporte se vieron afectados por la modificación al alineamiento. La mayoría de los clados observados presentaron un menor valor de soporte que al usar el alineamiento uno para la construcción.

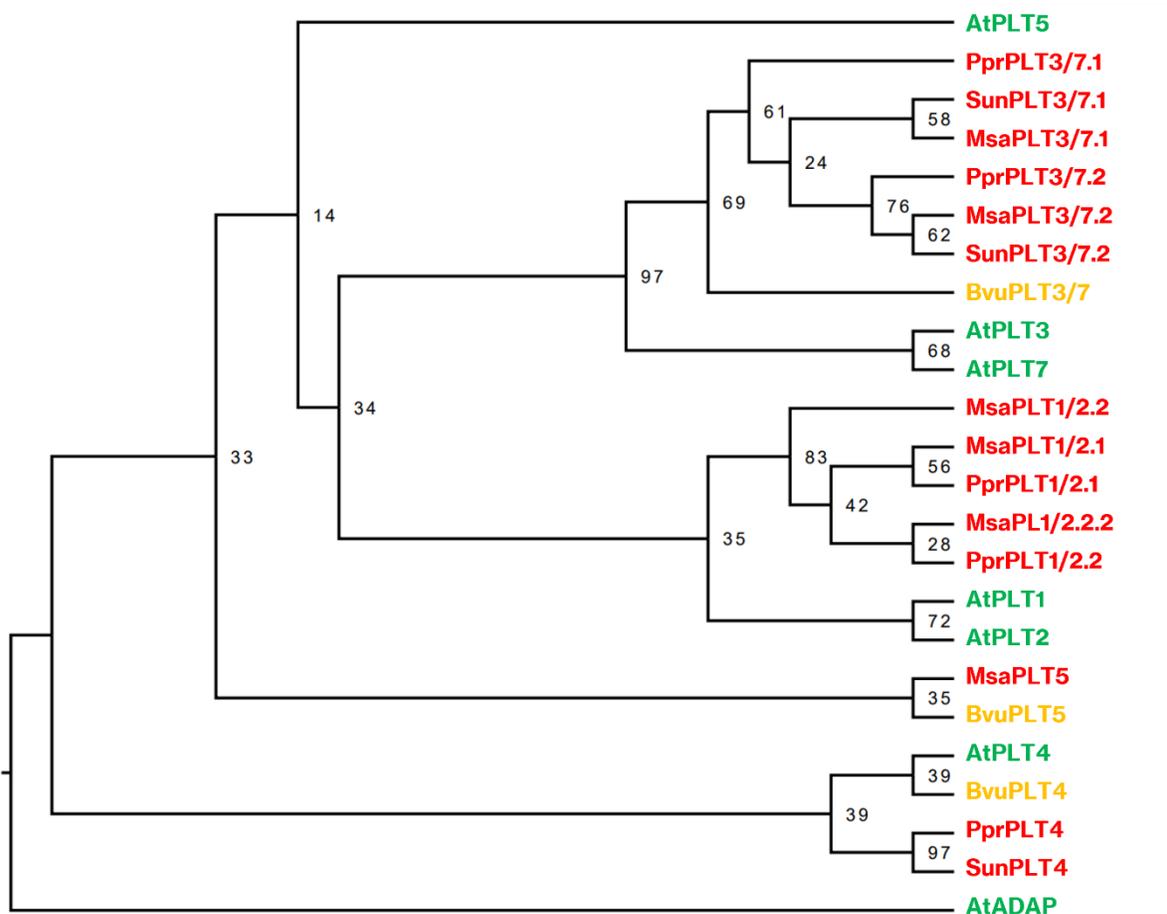


Figura 12. Reconstrucción de la filogenia molecular de las proteínas PLT a partir del alineamiento dos y máxima verosimilitud, enraizada en punto medio. En cada nodo se pueden observar los valores de Bootstrap que le dan soporte al clado. Los nombres de las secuencias pertenecientes a plantas de la familia Cactaceae aparecen en letras de color rojo, los pertenecientes a *B. vulgaris* (familia Amaranthaceae) aparecen en letras amarillas, los pertenecientes a *A. thaliana* (familia Brassicaceae) aparecen en letras verdes. At: *A. thaliana*, Msa: *M. san-angelensis*, Ppr: *P. pringlei*, Sun: *S. undatus* y Bvu: *B. vulgaris*.

En la filogenia molecular de las proteínas PLT reconstruida a partir del alineamiento tres se conservan la mayoría de los clados observados en el análisis anterior. La topología del clado PLT1/2 se mantiene sin diferencias. Los clados PLT3/7 y PLT4 presentan diferencias menores, pero conservan a sus miembros. Únicamente el ortólogo *AtPLT5* cambia completamente de posición quedando en un clado junto con sus ortólogos putativos hallado en el genoma de *M. san-angelensis* y *B. vulgaris*. El árbol presentó valores de soporte consistentes en nodos terminales y basales (figura 13).

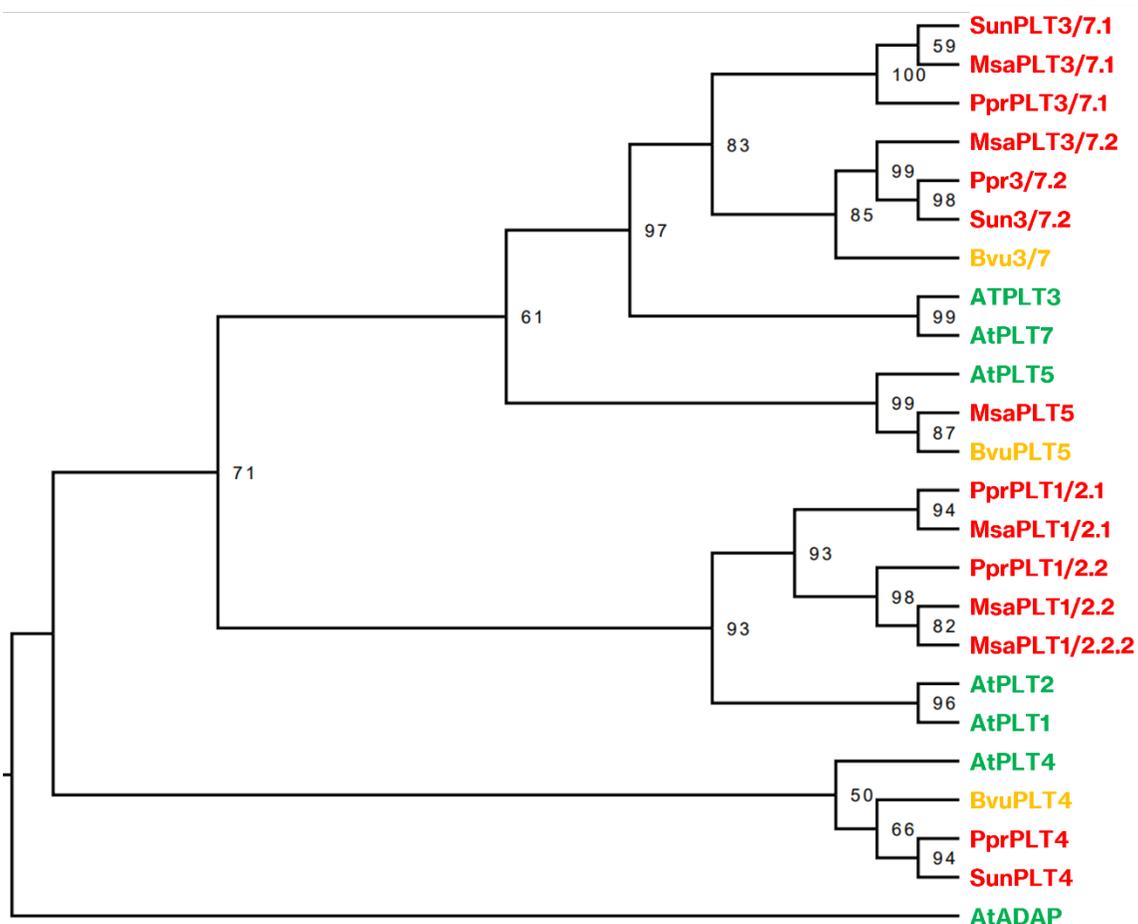


Figura 13. Reconstrucción de la filogenia molecular de las proteínas PLT a partir del alineamiento tres y máxima verosimilitud, enraizada en punto medio. En cada nodo se pueden observar los valores de Bootstrap que le dan soporte al clado. Los nombres de las secuencias pertenecientes a plantas de la familia Cactaceae aparecen en letras de color rojo, los pertenecientes a *B. vulgaris* (familia Amaranthaceae) aparecen en letras amarillas, los pertenecientes a *A. thaliana* (familia Brassicaceae) aparecen en letras verdes. At: *A. thaliana*, Msa: *M. san-angelensis*, Ppr: *P. pringlei*, Sun: *S. undatus* y Bvu: *B. vulgaris*.

1. Los ortólogos putativos de *AtPLT1* y *AtPLT2* en *M. san-angelensis*

Después de analizar los árboles de genes se seleccionaron dos ortólogos putativos de *AtPLT1* y *AtPLT2* en el genoma de *M. san-angelensis*. El análisis filogenético no fue capaz de discernir entre el ortólogo putativo de *AtPLT1* y el de *AtPLT2*. Por ello, los ORF seleccionados fueron etiquetados de la siguiente forma: *MsaPLT1/2.1* y *MsaPLT1/2.2*. *MsaPLT1/2.1* presenta en su estructura teórica 8 exones y 6 intrones. Tiene una longitud de 3355 pb (Figura 14A). *MsaPLT1/2.2* presenta en su estructura teórica 8 exones y 6 intrones, y tiene una longitud de 3880 pb (Figura 14B).

2. Los ortólogos putativos de *AtPLT3* y *AtPLT7* en *M. san-angelensis*

De forma similar se seleccionaron dos ortólogos putativos de *AtPLT3* y *AtPLT7* en el genoma de *M. san-angelensis*. Como ocurre con los ortólogos putativos de *AtPLT1/2*, el análisis filogenético no fue capaz de discernir entre el ortólogo putativo de *AtPLT3* y el de *AtPLT7*. Por ello, los ORF seleccionados fueron etiquetados de la siguiente forma: *MsaPLT3/7.1* y *MsaPLT3/7.2*. *MsaPLT3/7.1* presenta en su estructura teórica 8 exones y 7 intrones. Tiene una longitud de 5485 pb (Fig. 14C), mientras que *MsaPLT3/7.2* presenta en su estructura teórica 9 exones y 8 intrones, y tiene una longitud teórica de 4577 pb (Fig. 14D).

3. El ortólogo putativo de *AtPLT5* en *M. san-angelensis*

Después de analizar los árboles de genes se seleccionó un ORF ortólogo putativo de *AtPLT5* en el genoma de *M. san-angelensis* que fue etiquetado como *MsaPLT5*. *MsaPLT5* presenta en su estructura teórica 8 exones y 7 intrones y tiene una longitud teórica de 6867pb (Figura 14E).

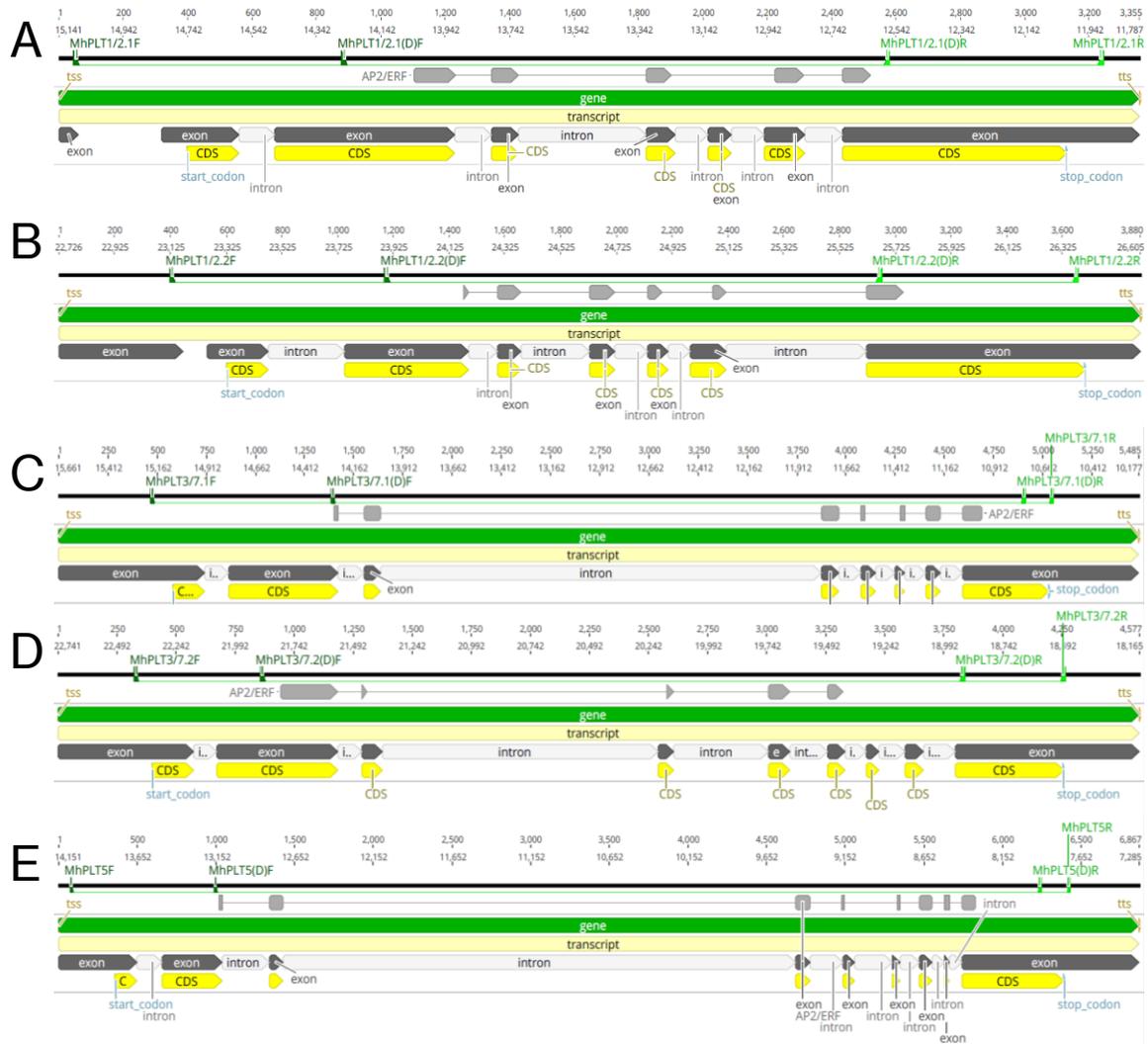


Figura 14. Modelos genéticos de los ortólogos putativo de los genes *AtPLT* en el genoma de *M. san-angelensis*. A) *MsaPLT1/2.1* B) *MsaPLT1/2.2* C) *MsaPLT3/7.1* D) *MsaPLT3/7.2* E) *MsaPLT5* Los números superiores corresponden a los nucleótidos dentro del transcrito primario del gen, los números que se encuentran debajo de estos enumeran los nucleótidos con referencia a su posición en el contig del que fue obtenido. Se marcan con líneas verdes verticales los sitios que fueron usados como referencia para el diseño de primers en esta secuencia. La figura gris sobre la barra “gen” señala la región correspondiente al dominio-interdominio-dominio. “transcript” se refiere al transcrito inmaduro. “tss” señala el sitio de inicio transcripcional (Por las siglas en inglés transcriptional start site). “tts” señala el sitio de término transcripcional (Por las siglas en inglés transcriptional termination sites). Se observan regiones de transcrito no etiquetadas como intrón o exón, se dejan así pues la plataforma no fue capaz la naturaleza de dichas regiones. Todos estos datos fueron resultado del análisis de Augustus.

7.3 Amplificación de los productos *MsaPLT1/2.1*, *MsaPLT1/2.2*, *MsaPLT3/7.1*, *MsaPLT5* y sus ortólogos en cinco especies del género *Mammillaria*

1. Amplificación usando DNA genómico (gDNA) de *M. san-angelensis*

Para respaldar experimentalmente la identificación de ORF obtenidos a través de los procedimientos bioinformáticos, se diseñaron *primers* para dos productos distintos de cada gen ortólogo putativo reconocido: el primer juego buscando amplificar la mayor parte de la secuencia que se encuentra entre el primer y último exón, y el segundo juego buscando amplificar las secuencias abarcando la mayor región posible entre el inicio y el final del CDS. Ambos juegos de *primers* se diseñaron en secuencias cuya predicción en el transcrito no fueran sujetas de splicing, de tal manera que fueran útiles para amplificar en gDNA y cDNA. Los sitios de alineamiento de los *primers* están indicados en los modelos de genes descritos en la sección anterior, y denotados por una F o una R al final del elemento (Fig. 14). Estos fueron utilizados para amplificar las secuencias de los genes *PLT* de *M. san-angelensis* (MSA). Los pares de *primers* cuyo blanco eran los ORF *MsaPLT1/2.1*, *MsaPLT1/2.2*, *MsaPLT3/7.1*, y *MsaPLT5* fueron capaces de amplificar (Figura 15) bajo las condiciones de termociclador descritas en el protocolo 1, 1, 1, y 2, respectivamente.

El producto amplificado por los *primers* *MsaPLT1/2.1(D)* tuvo una longitud de alrededor de 1500 pb; su longitud teórica era de 1705 pb. El producto amplificado por los *primers* *MsaPLT1/2.1* tuvo una longitud de alrededor de 3000 pb; su longitud teórica era de 3201 pb. El producto amplificado por los *primers* *MsaPLT1/2.2* tuvo una longitud de alrededor de 1500 pb; su longitud teórica era de 1785 pb. El producto amplificado por los *primers* *MsaPLT3/7.1* tuvo una longitud aproximada de 2000 pares de bases; su longitud teórica era de 3526 pb. Finalmente, el producto amplificado por los *primers* *MsaPLT5(D)* tuvo una longitud de aproximadamente 5000 pares de bases; su longitud teórica era de 5256 pb. Los dos *primers* diseñados para *Msa3/7.2* no amplificaron (Figura 15).

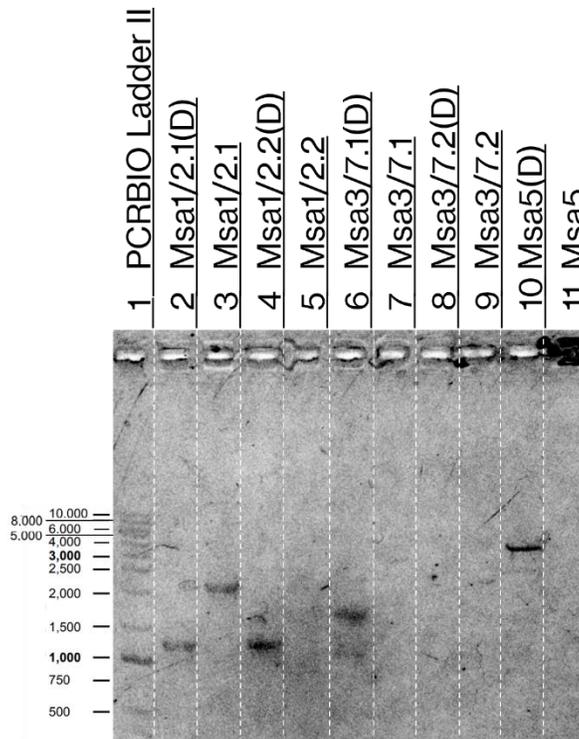


Figura 15. Productos de PCR de primers diseñados sobre las secuencias teóricas minadas del genoma de *M. san-angelensis*. gDNA proveniente de *M. san-angelensis*. Electroforesis en gel de agarosa al 1%.

2. Amplificación de los productos correspondientes a *MsaPLT1/2.1*, *MsaPLT1/2.2*, *MsaPLT3/7.1*, *MsaPLT5* y sus ortólogos en cinco especies del género *Mammillaria*

Usando como templado cDNA se buscó amplificar mediante PCR los productos que habían sido amplificados previamente usando gDNA de *M. san-angelensis* usando los mismos pares de *primers*. Los pares de *primers* diseñados sobre el ORF *MsaPLT1/2.1* demostraron ser capaces de amplificar a partir de material genético de *M. hernandezii* (HER) y *M. lasiacantha* (LAS). *primers* Ambos productos tuvieron una longitud de alrededor de 1500 pb, su longitud esperada era de 874 pb. No fue posible amplificar *Msa1/2.1* ni sus ortólogos putativos en *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. duwei* (DUW), ni en *M. san-angelensis* (MSA) (Figura 16).

También fue posible amplificar *MsaPLT1/2.2* a partir de cDNA de *M. san-angelensis* (MSA). Este amplicón tuvo una longitud de alrededor de 1500 pb, y una longitud teórica de 1785 pb. No hubo amplificación en las reacciones llevadas a cabo con

cDNA de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 16).

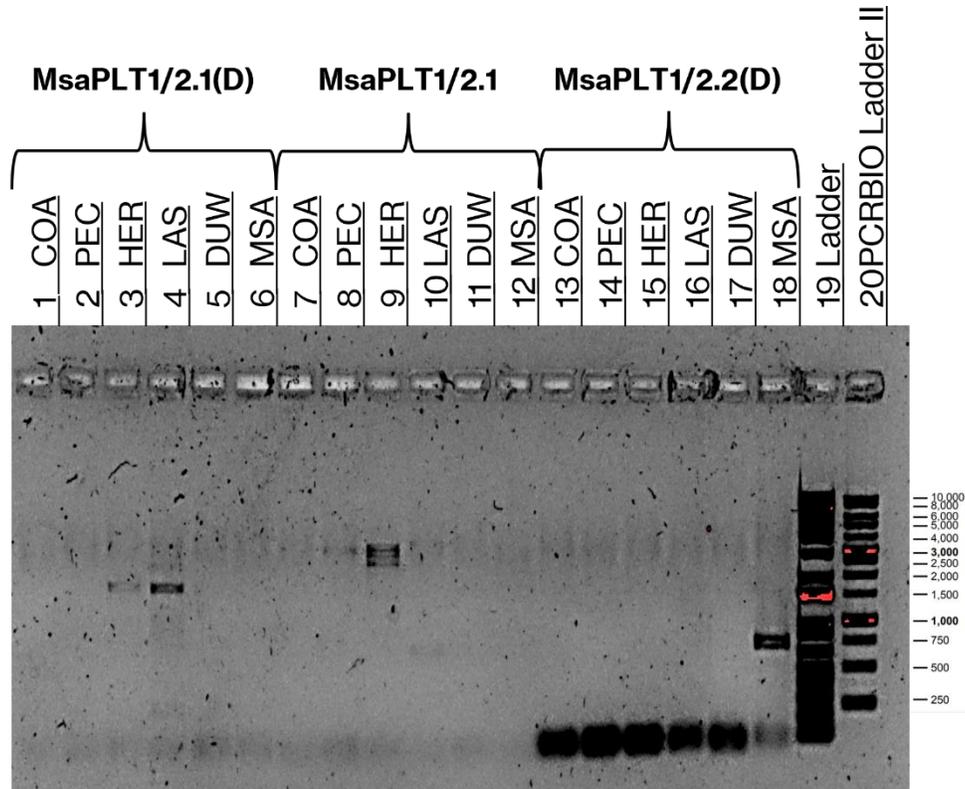


Figura 16. Productos de PCR de los juegos de primers MsaPLT1/2.1(D), MsaPLT1/2.1, y MsaPLT1/2.2(D) diseñados sobre las secuencias teóricas minadas del genoma de *M. san-angelensis*. cDNA proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), *M. duwei* (DUW) y *M. san-angelensis* (MSA). Electroforesis en gel de agarosa al 1%.

Fue posible amplificar utilizando el juego de primers MsaPLT3/7.1 (D) en cDNA de *M. pectinifera* (PEC), *M. hernandezii* (HER) y *M. san-angelensis* (MSA) estos amplicones tuvieron un tamaño observado de alrededor de 500 pb, cuando el tamaño esperado de este producto era de 729 pb. La amplificación con este juego de primers no fue exitosa en el cDNA proveniente de *M. lasiacantha* (LAS) ni *M. duwei* (DUW) (Figura 17). Usando el juego de primers MsaPLT5 (D) fue posible obtener un producto de amplificación con cDNA de *M. pectinifera* (PEC); este producto tuvo un tamaño de alrededor de 750 pares de bases, su tamaño esperado era de 729 pb. La amplificación con este juego de primers no fue exitosa en el cDNA

proveniente de *M. coahuilensis* (COA), *M. hernandezii* (HER), *M. lasiacantha* (LAS), *M. san-angelensis* (MSA), ni de *M. duwei* (DUW) (Figura 17).

Fue posible obtener un producto de amplificación utilizando el juego de *primers* MsaPLT5 y cDNA de *M. hernandezii* (HER), y *M. lasiacantha* (LAS). En ambos casos se observó la amplificación de dos productos. Para ambas especies se esperaba un producto de amplificación de 1901 pb. Los productos observados tuvieron en ambos casos un tamaño aproximado de 1500 pb y 3000 pb, respectivamente. La amplificación con este juego de *primers* no fue exitosa en el cDNA proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. san-angelensis* (MSA), ni de *M. duwei* (DUW) (Figura 17).

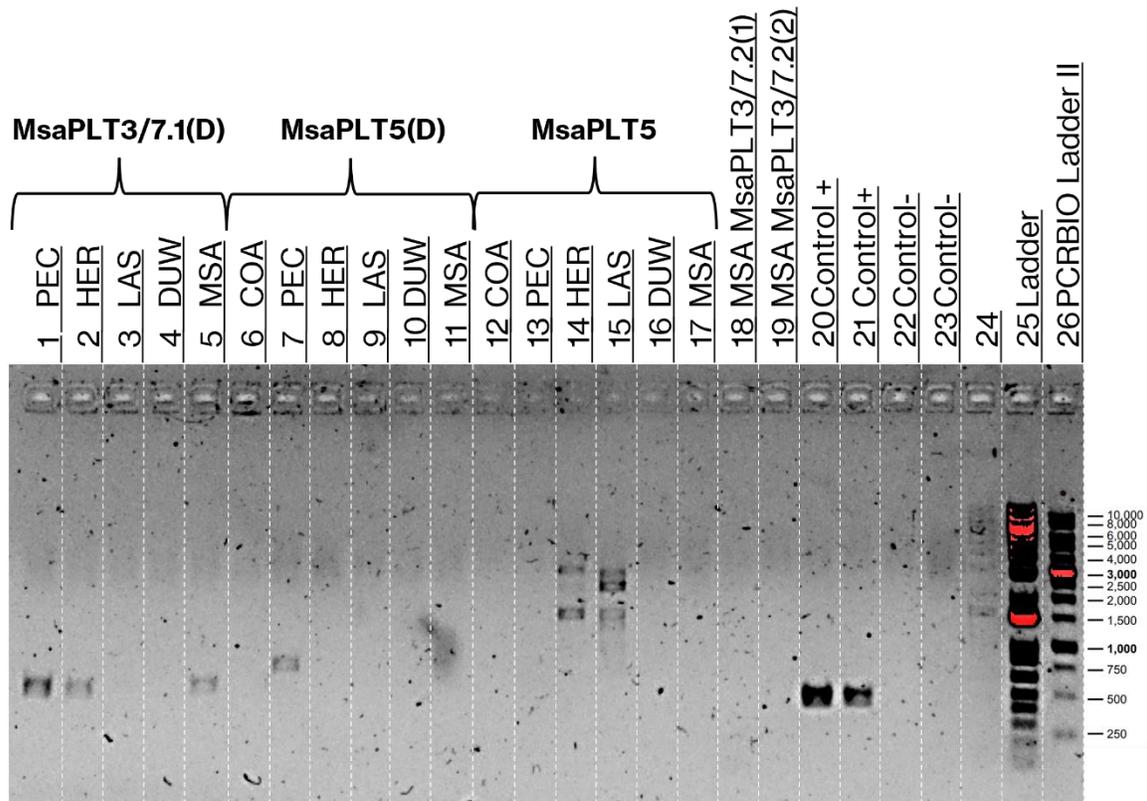


Figura 17. Productos de PCR de los juegos de *primers* MsaPLT3/7.1(D), MsaPLT5, y MsaPLT5(D) diseñados sobre las secuencias teóricas minadas del genoma de *M. san-angelensis*. cDNA proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), *M. duwei* (DUW) y *M. san-angelensis* (MSA). Electroforesis en gel de agarosa al 1%.

3. Ensayos con diferentes condiciones de termociclado.

Después de evaluar los resultados anteriores, se optó por modificar las condiciones de termociclador utilizadas anteriormente. En el caso del juego de *primers* MsaPLT1/2.1(D) se utilizó el protocolo de termociclador 5. Esto no resultó en amplificación en el cDNA proveniente de *M. san-angelensis* (MSA), *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 18).

Utilizar el juego de *primers* MsaPLT1/2.2(D) con las condiciones de termociclador descritas en el protocolo 7 no resultó en amplificación en el cDNA proveniente de *M. san-angelensis* (MSA), *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 18).

En el caso del juego de *primers* MsaPLT3/7.1 el uso del protocolo de termociclador 5 resultó en bandas más claras en la electroforesis de *M. pectinifera* (PEC), *M. hernandezii* (HER), y *M. san-angelensis* (MSA), así como en la amplificación en *M. coahuilensis* (COA) de un producto con un tamaño de alrededor de 1000 pb; el tamaño de producto esperado en este caso era de 729 pb (Figura 18, carriles 15-18). La amplificación con este juego de *primers* y condiciones no fue exitosa en el cDNA proveniente de *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 18).

En el caso del juego de *primers* MsaPLT5(D) el uso del protocolo de termociclador 5 únicamente resultó en una banda más clara para *M. pectinifera* (Figura 18, carril 23). La amplificación con este juego de *primers* no fue exitosa en el cDNA proveniente de *M. coahuilensis* (COA), *M. hernandezii* (HER), *M. lasiacantha* (LAS), *M. san-angelensis* (MSA), ni de *M. duwei* (DUW) (Figura 18).

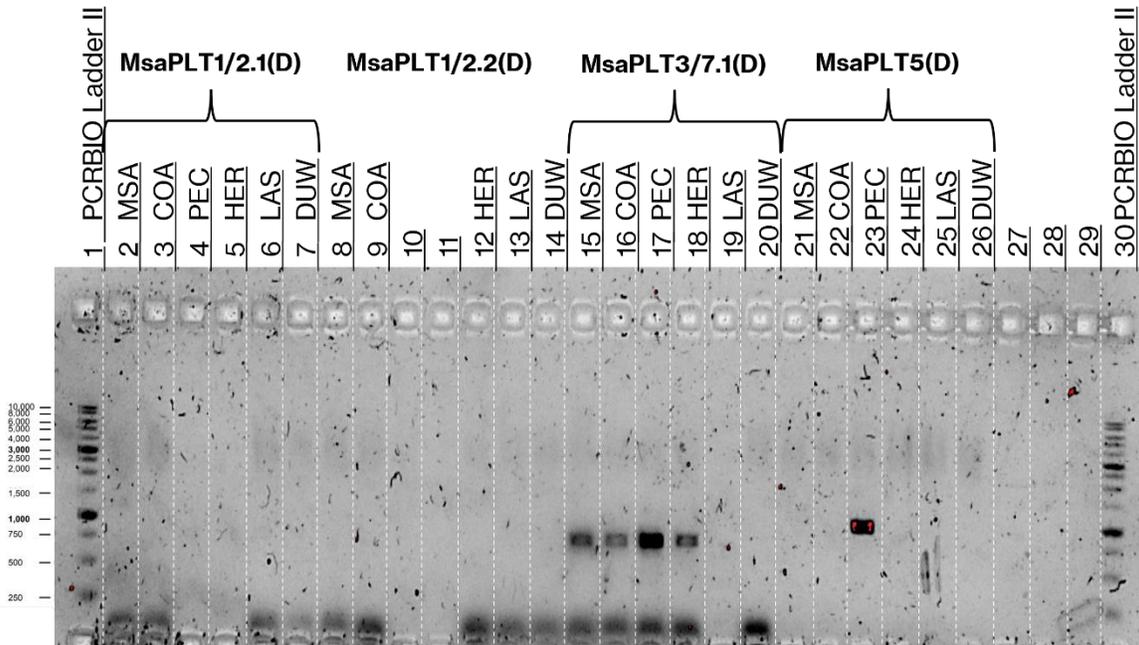


Figura 18. Productos de PCR de los juegos de primers *MsaPLT1/2.1(D)*, *MsaPLT1/2.2(D)*, y *MsaPLT3/7.1(D)* y *MsaPLT5(D)* diseñados sobre las secuencias teóricas minadas del genoma de *M. san-angelensis*. cDNA sintetizado usando OligodT con material proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), *M. duwei* (DUW) y *M. san-angelensis* (MSA). Electroforesis en gel de agarosa al 1%.

Usando las mismas condiciones modificadas de termociclado, se utilizó cDNA sintetizado usando random primer mix, en lugar de oligodT. Para el juego de *primers* *MsaPLT1/2.1(D)* las condiciones del protocolo 5 no resultaron en amplificación en el cDNA (sintetizado usando RandomPrimerMix) proveniente de *M. san-angelensis* (MSA), *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 19).

Usar el juego de *primers* *MsaPLT1/2.2(D)* bajo las condiciones descritas en el protocolo 7 no resultó en amplificación en el cDNA (sintetizado usando RandomPrimerMix) proveniente de *M. san-angelensis* (MSA), *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 19).

En el caso del juego de *primers* *MsaPLT3/7.1* las condiciones del protocolo 5 resultaron en amplificación junto con el cDNA (Sintetizado usando RandomPrimerMix) de *M. pectinifera* (PEC) y *M. san-angelensis* (MSA). En ambos casos se amplificó un producto con un tamaño de alrededor de 1000 pb, el tamaño

de producto esperado en este caso era de 729 pb (Figura 19, carriles 14 y 16) La amplificación con este juego de *primers* y condiciones no fue exitosa en el cDNA proveniente de *M. coahuilensis* (COA), *M. hernandezii* (HER), *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 19).

En el caso del juego de *primers* MsaPLT5(D) el uso del protocolo de termociclador 5 resultó en la amplificación de un producto junto con el cDNA de *M. pectinifera* (PEC), el producto esperado en este caso tenía una longitud de 963 pb, el producto observado tuvo una longitud de 1000 pb (Figura 19, carril 22). Así como en la amplificación de dos productos usando el cDNA de *M. hernandezii* (HER), el producto esperado en este caso tenía una longitud de 963 pb, los productos observados tuvieron una longitud de alrededor de 1000 y 1500 pb, respectivamente (Figura 19, carril 23). La amplificación con este juego de *primers* no fue exitosa en el cDNA proveniente de *M. coahuilensis* (COA), *M. lasiacantha* (LAS), *M. san-angelensis* (MSA), ni de *M. duwei* (DUW) (Figura 19).

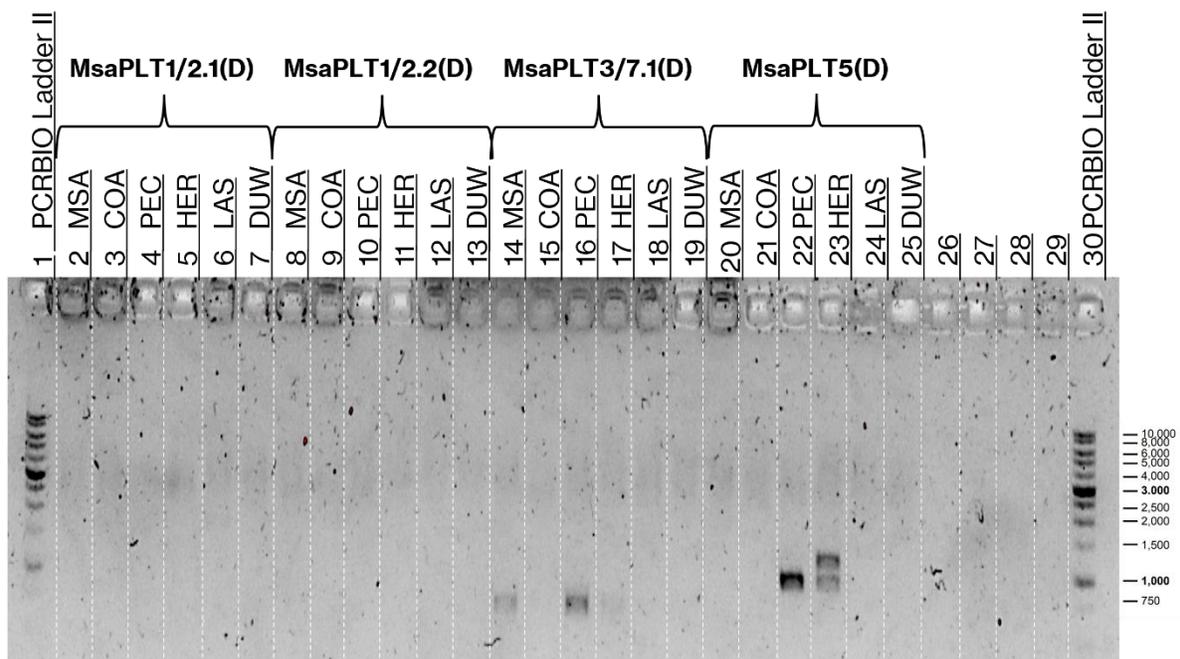


Figura 19. Productos de PCR de los juegos de *primers* MsaPLT1/2.1(D), MsaPLT1/2.2(D), y MsaPLT3/7.1(D) y MsaPLT5(D) diseñados sobre las secuencias teóricas minadas del genoma de *M. san-angelensis*. cDNA sintetizado usando RandomPrimerMix con material proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), *M. duwei* (DUW) y *M. san-angelensis* (MSA). Electroforesis en gel de agarosa al 1%.

En el caso del juego de *primers* MsaPLT1/2.1 el cambio al protocolo 5 resultó en una amplificación similar a lo observado anteriormente para *M. hernandezii* (HER), este producto tuvo una longitud de alrededor de 2000 pb, su longitud esperada era de 2206 pb (Figura 20, carril 16). La amplificación con este juego de *primers* y condiciones no fue exitosa en el cDNA proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. san-angelensis* (MSA), *M. lasiacantha* (LAS), ni de *M. duwei* (DUW) (Figura 20).

Finalmente, al usar el juego de *primers* MsaPLT5 el uso del protocolo de termociclador 7 resultó en una amplificación exitosa en *M. duwei* (DUW). Este producto tuvo una longitud de alrededor de 2000 pb, su longitud esperada era de 1901 pb (Figura 20, carril 20). La amplificación con este juego de *primers* y condiciones no fue exitosa en el cDNA proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), ni de *M. san-angelensis* (MSA) (Figura 20).

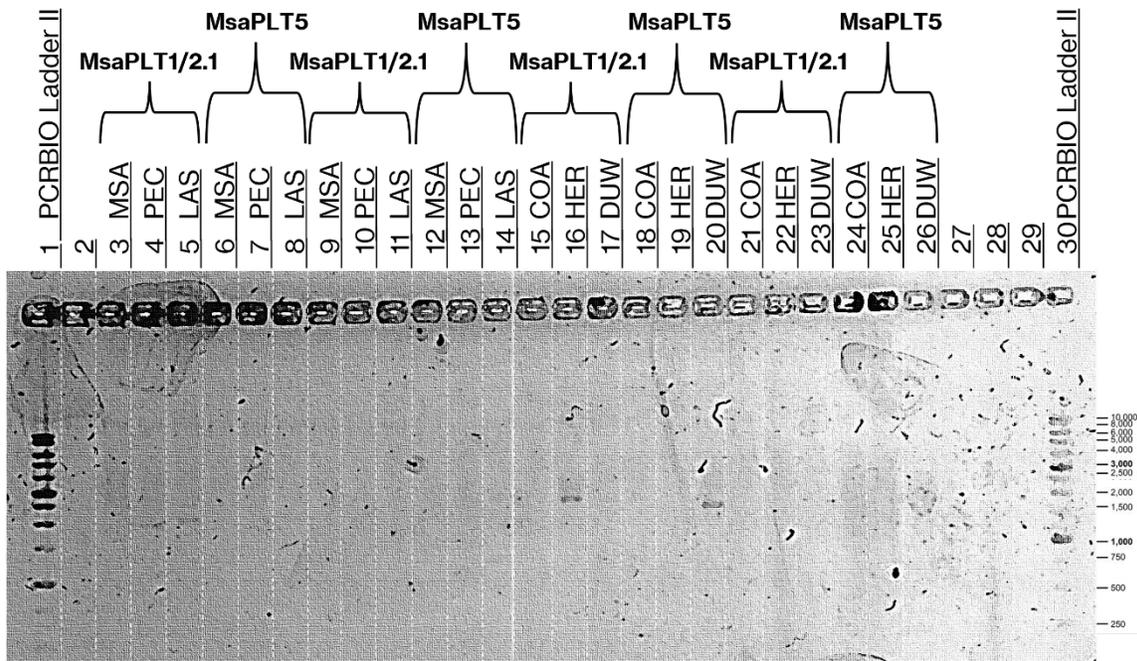


Figura 20. Productos de PCR de los juegos de *primers* MsaPLT1/2.1 y MsaPLT5 diseñados sobre las secuencias teóricas minadas del genoma de *M. san-angelensis*. cDNA sintetizado usando RandomPrimerMix con material proveniente de *M. coahuilensis* (COA), *M. pectinifera* (PEC), *M. hernandezii* (HER), *M. lasiacantha* (LAS), *M. duwei* (DUW) y *M. san-angelensis* (MSA). Electroforesis en gel de agarosa al 1%.

4. Resumen de la amplificación de productos candidatos a ser *MsaPLT1/2.1*, *MsaPLT1/2.2*, *MsaPLT3/7.1* y *MsaPLT5* en *M. san-angelensis*

En *M. san-angelensis* fue posible amplificar un producto utilizando gDNA junto con los juegos de *primers* *MsaPLT1/2.1(D)* y *MsaPLT1/2.1* usando el protocolo de termociclador 1. También fue posible amplificar *MsaPLT1/2.2* utilizando el juego de *primers* *MsaPLT1/2.2 (D)* junto con gDNA y el protocolo de termociclador 1; así como usando cDNA (sintetizado usando oligodT) y el protocolo de termociclador 3. Se amplificó *MsaPLT3/7.1* utilizando el juego de *primers* *MsaPLT3/7.1(D)* junto con gDNA y el protocolo de termociclador 1; así como usando cDNA (sintetizado usando oligodT) usando el protocolo 3 y 5, y usando cDNA (sintetizado usando RandomPrimerMix) junto con el protocolo de termociclador número 5. Finalmente, fue posible amplificar *MsaPLT5* utilizando gDNA junto con el juego de *primers* *MsaPLT5(D)* y el protocolo de termociclador 2. No fue posible amplificar un producto correspondiente al ORF *MsaPLT3/7.2* con las diferentes condiciones evaluadas (Tabla 3).

Tabla 3. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. san-angelensis*

Especie	Material	Juego de primers usado	Tamaño de producto esperado (bases)	Tamaño de producto observado pb
MSA	gDNA	MsaPLT1/2.1(D)	1705	~1500
MSA	gDNA	MsaPLT1/2.1	3201	~3000
MSA	gDNA	MsaPLT1/2.2 (D)	1785	~1500
MSA	gDNA	MsaPLT1/2.2	3260	-
MSA	gDNA	MsaPLT3/7.1(D)	3526	~2000
MSA	gDNA	MsaPLT3/7.1	4587	-
MSA	gDNA	MsaPLT3/7.2(D)	2984	-
MSA	gDNA	MsaPLT3/7.2	3945	-
MSA	gDNA	MsaPLT5(D)	5256	~5000
MSA	gDNA	MsaPLT5	6357	-
MSA	dTcDNA	MsaPLT1/2.1(D)	874	-
MSA	dTcDNA	MsaPLT1/2.1	2206	-
MSA	dTcDNA	MsaPLT1/2.2 (D)	731	~750
MSA	dTcDNA	MsaPLT3/7.1(D)	729	~500
MSA	dTcDNA	MsaPLT5(D)	963	-
MSA	dTcDNA	MsaPLT5	1901	-
MSA	dTcDNA	MsaPLT3/7.2(D)	~1344	-
MSA	dTcDNA	MsaPLT3/7.2	~1344	-
MSA	dTcDNA	MsaPLT1/2.1(D)	874	-
MSA	dTcDNA	MsaPLT1/2.2 (D)	731	-
MSA	dTcDNA	MsaPLT3/7.1(D)	729	~1000
MSA	dTcDNA	MsaPLT5(D)	963	-
MSA	rPcDNA	MsaPLT1/2.1(D)	874	-
MSA	rPcDNA	MsaPLT1/2.2 (D)	731	-
MSA	rPcDNA	MsaPLT3/7.1(D)	729	~750
MSA	rPcDNA	MsaPLT5(D)	963	-
MSA	dTcDNA	MsaPLT1/2.1	2206	-
MSA	dTcDNA	MsaPLT5	1901	-
MSA	rPcDNA	MsaPLT1/2.1	2206	-
MSA	rPcDNA	MsaPLT5	1901	-

5. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. coahuilensis*

En *M. coahuilensis* fue posible amplificar un producto utilizando los *primers* diseñados sobre el ORF *MsaPLT3/7.1*. Esto se logró utilizando el juego de *primers* *MsaPLT3/7.1(D)* junto con cDNA (sintetizado usando oligodT) y usando el protocolo de termociclador 5. No fue posible amplificar productos utilizando los *primers* diseñados sobre los ORF *MsaPLT1/2.1*, *MsaPLT1/2.2* ni *MsaPLT5* (Tabla 4).

Tabla 4. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. coahuilensis*

Especie	Material	Juego de <i>primers</i> usado	Tamaño de producto esperado (bases)	Tamaño de producto observado pb
COA	dTcDNA	<i>MsaPLT1/2.1(D)</i>	874	-
COA	dTcDNA	<i>MsaPLT1/2.1</i>	2206	-
COA	dTcDNA	<i>MsaPLT1/2.2 (D)</i>	731	-
COA	dTcDNA	<i>MsaPLT5(D)</i>	963	-
COA	dTcDNA	<i>MsaPLT5</i>	1901	-
COA	dTcDNA	<i>MsaPLT1/2.1(D)</i>	874	-
COA	dTcDNA	<i>MsaPLT1/2.2 (D)</i>	731	-
COA	dTcDNA	<i>MsaPLT3/7.1(D)</i>	729	~1000
COA	dTcDNA	<i>MsaPLT5(D)</i>	963	-
COA	rPcDNA	<i>MsaPLT1/2.1(D)</i>	874	-
COA	rPcDNA	<i>MsaPLT1/2.2 (D)</i>	731	-
COA	rPcDNA	<i>MsaPLT3/7.1(D)</i>	729	-
COA	rPcDNA	<i>MsaPLT5(D)</i>	963	-
COA	dTcDNA	<i>MsaPLT1/2.1</i>	2206	-
COA	dTcDNA	<i>MsaPLT5</i>	1901	-
COA	rPcDNA	<i>MsaPLT1/2.1</i>	2206	-
COA	rPcDNA	<i>MsaPLT5</i>	1901	-

6. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. duwei*

En *M. duwei* fue posible amplificar un producto utilizando los *primers* diseñados sobre el ORF *MsaPLT5*. Esto se logró utilizando el juego de *primers* *MsaPLT5* junto con cDNA (sintetizado usando oligodT) y usando el protocolo de termociclador 7. No fue posible amplificar productos utilizando los *primers* diseñados sobre los ORF *MsaPLT1/2.1*, *MsaPLT1/2.2* ni de *MsaPLT3/7.1* (Tabla 5).

Tabla 5. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. duwei*

Especie	Material	Juego de primers usado	Tamaño de producto esperado (bases)	Tamaño de producto observado pb
DUW	dTcDNA	MsaPLT1/2.1(D)	874	-
DUW	dTcDNA	MsaPLT1/2.1	2206	-
DUW	dTcDNA	MsaPLT1/2.2 (D)	731	-
DUW	dTcDNA	MsaPLT3/7.1(D)	729	-
DUW	dTcDNA	MsaPLT5(D)	963	-
DUW	dTcDNA	MsaPLT5	1901	-
DUW	dTcDNA	MsaPLT1/2.1(D)	874	-
DUW	dTcDNA	MsaPLT1/2.2 (D)	731	-
DUW	dTcDNA	MsaPLT3/7.1(D)	729	-
DUW	dTcDNA	MsaPLT5(D)	963	-
DUW	rPcDNA	MsaPLT1/2.1(D)	874	-
DUW	rPcDNA	MsaPLT1/2.2 (D)	731	-
DUW	rPcDNA	MsaPLT3/7.1(D)	729	-
DUW	rPcDNA	MsaPLT5(D)	963	-
DUW	dTcDNA	MsaPLT1/2.1	2206	-
DUW	dTcDNA	MsaPLT5	1901	~2000
DUW	rPcDNA	MsaPLT1/2.1	2206	-
DUW	rPcDNA	MsaPLT5	1901	-

7. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. hernandezii*

En *M. hernandezii* fue posible amplificar un producto utilizando los *primers* diseñados sobre el ORF *MsaPLT1/2.1*. Para esto se uso cDNA (sintetizado usando oligodT) junto con los juegos de *primers* *MsaPLT1/2.1(D)*, y el protocolo 3, y el juego de *primers* *MsaPLT1/2.1* usando los protocolos de termociclador 4 y 6. Se amplificó utilizando los *primers* diseñados sobre el ORF *MsaPLT3/7.1* utilizando el juego de *primers* *MsaPLT3/7.1(D)* junto con cDNA (sintetizado usando oligodT) y los protocolos de termociclador 3 y 5. Finalmente, fue posible amplificar utilizando los *primers* diseñados sobre el ORF *MsaPLT5* utilizando cDNA (sintetizado usando oligodT) junto con el juego de *primers* *MsaPLT5* y el protocolo de termociclador 4, así como utilizando cDNA (sintetizado usando RandomPrimerMix) junto con el juego de *primers* *MsaPLT5(D)* y el protocolo de termociclador 5. No fue posible amplificar productos usando los *primers* diseñados sobre el ORF *MsaPLT1/2.2* con las diferentes condiciones evaluadas (Tabla 6).

Tabla 6. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. hernandezii*

Especie	Material	Juego de primers usado	Tamaño de producto esperado (bases)	Tamaño de producto observado
HER	dTcDNA	MsaPLT1/2.1(D)	874	~1500
HER	dTcDNA	MsaPLT1/2.1	2206	~3000
HER	dTcDNA	MsaPLT1/2.2 (D)	731	-
HER	dTcDNA	MsaPLT3/7.1(D)	729	~500
HER	dTcDNA	MsaPLT5(D)	963	-
HER	dTcDNA	MsaPLT5	1901	~1500, ~3000
HER	dTcDNA	MsaPLT1/2.1(D)	874	-
HER	dTcDNA	MsaPLT1/2.2 (D)	731	-
HER	dTcDNA	MsaPLT3/7.1(D)	729	~1000
HER	dTcDNA	MsaPLT5(D)	963	-
HER	rPcDNA	MsaPLT1/2.1(D)	874	-
HER	rPcDNA	MsaPLT1/2.2 (D)	731	-
HER	rPcDNA	MsaPLT3/7.1(D)	729	-
HER	rPcDNA	MsaPLT5(D)	963	~1000, ~1500
HER	dTcDNA	MsaPLT1/2.1	2206	~2000
HER	dTcDNA	MsaPLT5	1901	-
HER	rPcDNA	MsaPLT1/2.1	2206	-
HER	rPcDNA	MsaPLT5	1901	-

8. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. lasiacantha*

En *M. lasiacantha* fue posible amplificar un producto utilizando los *primers* diseñados sobre el ORF *MsaPLT1/2.1*. Esto se logró utilizando el juego de *primers* MsaPLT1/2.1(D) junto con cDNA (sintetizado usando oligodT). Y usando el protocolo de termociclador 3. También se logró amplificar utilizando los *primers* diseñados sobre el ORF *MsaPLT5* utilizando el juego de *primers* MsaPLT5 junto con cDNA (sintetizado usando oligodT). Y usando el protocolo de termociclador 4. No fue posible amplificar utilizando los *primers* diseñados sobre los ORF *MsaPLT1/2.2* ni *MsaPLT3/7.1* (Tabla 7).

Tabla 7. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. lasiacantha*

Especie	Material	Juego de primers usado	Tamaño de producto esperado (bases)	Tamaño de producto observado
LAS	dTcDNA	MsaPLT1/2.1(D)	874	~1500
LAS	dTcDNA	MsaPLT1/2.1	2206	-
LAS	dTcDNA	MsaPLT1/2.2 (D)	731	-
LAS	dTcDNA	MsaPLT3/7.1(D)	729	-
LAS	dTcDNA	MsaPLT5(D)	963	-
LAS	dTcDNA	MsaPLT5	1901	~1500, ~3000
LAS	dTcDNA	MsaPLT1/2.1(D)	874	-
LAS	dTcDNA	MsaPLT1/2.2 (D)	731	-
LAS	dTcDNA	MsaPLT5(D)	963	-
LAS	rPcDNA	MsaPLT1/2.1(D)	874	-
LAS	rPcDNA	MsaPLT1/2.2 (D)	731	-
LAS	rPcDNA	MsaPLT3/7.1(D)	729	-
LAS	rPcDNA	MsaPLT5(D)	963	-
LAS	dTcDNA	MsaPLT1/2.1	2206	-
LAS	dTcDNA	MsaPLT5	1901	-
LAS	rPcDNA	MsaPLT1/2.1	2206	-
LAS	rPcDNA	MsaPLT5	1901	-

9. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. pectinifera*

En *M. pectinifera* fue posible amplificar el ortólogo de *MsaPLT3/7.1* utilizando el juego de primers *MsaPLT3/7.1(D)* junto con cDNA (sintetizado usando oligodT) y los protocolos de termociclador 3 y 4, así como utilizando cDNA (sintetizado usando RandomPrimerMix) junto con el juego de primers *MsaPLT3/7.1(D)* y el protocolo de termociclador 5. También, fue posible amplificar al ortólogo de *MsaPLT5* utilizando cDNA (sintetizado usando oligodT) junto con el juego de primers *MsaPLT5(D)* y los protocolos de termociclador 3 y 5, así como utilizando cDNA (sintetizado usando RandomPrimerMix) junto con el juego de primers *MsaPLT5(D)* y el protocolo de termociclador 5. No fue posible amplificar el producto ortólogo del aislado para el

ORF *MsaPLT1/2.1* ni el de *MsaPLT1/2.2* con las diferentes condiciones evaluadas (Tabla 8).

Tabla 8. Resumen de las reacciones de PCR llevadas a cabo usando material genético de *M. pectinifera*

Especie	Material	Juego de primers usado	Tamaño de producto esperado (bases)	Tamaño de producto observado
PEC	dTcDNA	MsaPLT1/2.1(D)	874	-
PEC	dTcDNA	MsaPLT1/2.1	2206	-
PEC	dTcDNA	MsaPLT1/2.2 (D)	731	-
PEC	dTcDNA	MsaPLT3/7.1(D)	729	~500
PEC	dTcDNA	MsaPLT5(D)	963	~750
PEC	dTcDNA	MsaPLT5	1901	-
PEC	dTcDNA	MsaPLT1/2.1(D)	874	-
PEC	dTcDNA	MsaPLT3/7.1(D)	729	~1000
PEC	dTcDNA	MsaPLT5(D)	963	~1000
PEC	rPcDNA	MsaPLT1/2.1(D)	874	-
PEC	rPcDNA	MsaPLT1/2.2 (D)	731	-
PEC	rPcDNA	MsaPLT3/7.1(D)	729	~750
PEC	rPcDNA	MsaPLT5(D)	963	~1000
PEC	dTcDNA	MsaPLT1/2.1	2206	-
PEC	dTcDNA	MsaPLT5	1901	-
PEC	rPcDNA	MsaPLT1/2.1	2206	-
PEC	rPcDNA	MsaPLT5	1901	-

8. Discusión

En el presente trabajo se reporta la secuenciación del genoma de *M. san-angelensis* usando la plataforma PacBio Sequel II. Usando las lecturas de dicha secuenciación como base de datos se identificaron 19 candidatos a ser los ortólogos putativos de los seis genes PLT presentes en el genoma de *A. thaliana*. El análisis filogenético permitió identificar 5 de estas 19 secuencias como candidatos a ser ortólogos putativos no redundantes de los genes PLT de *A. thaliana*: MsaPLT1/2.1, MsaPLT1/2.2, MsaPLT3/7.1, MsaPLT3/7.2 y MsaPLT5. Usando la información de estas secuencias se diseñaron *primers* sobre los ORF etiquetados como MsaPLT1/2.1, MsaPLT1/2.2, MsaPLT3/7.1 y MsaPLT5 que permitieron amplificar productos a partir de gDNA de *M. san-angelensis*. También fue posible, utilizando cDNA, amplificar productos usando los *primers* diseñados sobre el ORF MsaPLT1/2.1 en *M. lasiacantha* y *hernandezii*; MsaPLT3/7.1 en *M. coahuilensis*, *M. pectinifera* y *M. hernandezii*; y MsaPLT5 en *M. pectinifera*, *M. duwei*, *M. lasiacantha* y *M. hernandezii*, sentando las bases para futuros estudios de la evolución de genes *PLT* en *Mammillaria* y en cactáceas.

Predicción de genes de interés *ab-initio* en datos genómicos no ensamblados

Reportamos el diseño e implementación de un sistema para la identificación de ORFS, y predicción de sus características, a partir de datos genómicos de lecturas largas de alta calidad. La identificación de genes en datos de secuencias de *scaffolds* genómicos, mediante la búsqueda de secuencias similares en bases de datos, es un proceso bien conocido y reportado desde hace años (Do & Choi, 2006). De la misma forma, existen diversos programas que realizan la predicción *ab initio* de genes a partir secuencias de DNA, utilizando información de la presencia y estructura de genes conocidos (Allen & Salzberg, 2005; Stanke et al., 2008). No obstante, creemos que el pipeline desarrollado presenta diversas ventajas sobre el uso aislado de otros métodos, las cuales se discutirán a continuación.

El método que reportamos permitió el trabajo con un conjunto de datos de CCS producto de la secuenciación PacBio, partiendo de lecturas sin ensamblar de *M.*

san-angelensis. En el presente trabajo, el ensamblaje del genoma de *M. san-angelensis* presenta dificultades técnicas, computacionales y de recursos bioinformáticos, por lo cual en estos momentos no se tiene disponible. Es por eso que el diseño de métodos que permitan trabajar con secuencias no ensambladas es de suma importancia. En años recientes, diversos genomas de especies vegetales han sido ensamblados hasta nivel de cromosoma, llegando a más de 300 en el año 2022 (Zhou et al., 2022). Los proyectos de secuenciación de genomas generan un gran número de sets de datos que se hacen públicos a través del servicio SRA del NCBI (SRA, 2009). A inicios del año 2023 el número de sets de datos pertenecientes a plantas vasculares, obtenidos mediante la tecnología PacBio, era de 18461 pertenecientes a 866 especies NCBI (SRA, 2023). Es en la brecha entre el número de genomas secuenciados, y los ensamblados que se resalta la importancia de los métodos como el nuestro que trabajan con lecturas no ensambladas. Será importante a futuro analizar la precisión y exactitud de nuestro método llevando a cabo la búsqueda de diversos conjuntos de datos. Una fuerte limitación del método aquí presentado es que su efectividad varía dependiendo de la profundidad de las lecturas en relación con el genoma de interés.

Por otra parte, el programa Augustus permite la predicción *ab initio* de genes codificantes en situaciones en las que no se cuenta con evidencia transcripcional de la misma especie, como es el caso de *M. san-angelensis*. La estrategia usa evidencia transcripcional inter-específica empleando observaciones de especies relacionadas. Basándose en pruebas que llevaron a cabo en *Oryza sativa*, Liang y sus colaboradores (2009) estiman que podrían realizar en *Sorghum bicolor* una predicción génica con una precisión del 50 al 70% a nivel de CDS. El análisis filogenético junto con la amplificación de secuencias mediante PCR, utilizando *primers* diseñados para los ORF predichos, comprobó parcialmente que el uso de datos de cDNA de *A. thaliana* es un método viable para la predicción de genes en el genoma de *M. san-angelensis*. Un estudio más profundo es necesario para evaluar la precisión y exactitud de estas predicciones utilizando datos transcriptómicos de *M. san angelensis*. El método que utilizamos para identificar los ortólogos putativos de los genes *PLT* de *A. thaliana* en el genoma de *M. san-*

angelensis utiliza una serie de herramientas disyuntas en un proceso de búsqueda y curación. Este es un acercamiento distinto a otros métodos tradicionales como el de la plataforma Gramene que anota predicciones sobre genomas ensamblados utilizando datos de ESTs (Expressed Sequence Tags) y cDNA (Liang et al., 2009). El método aquí planteado es similar a métodos altamente automatizados como RGAugury (Li et al., 2016), que se especializa en la detección de RGAs (Resistance Gene Analogs). Sin embargo, una fortaleza central de nuestro método es que exige requerimientos mínimos en cuanto a la naturaleza de los datos, y puede adaptarse a la búsqueda de diversos genes de interés. A futuro será importante el esfuerzo de automatización de estas herramientas, la cual es altamente factible ya que todas ellas pueden ejecutarse desde un sistema operativo Linux.

En resumen, el método aquí reportado permitió trabajar con un conjunto de lecturas largas de PacBio sin necesidad de llevar a cabo un ensamble de genoma completo, generó predicciones que pudieron ser respaldadas, y tiene el potencial de ser adaptable a la búsqueda de diversos genes de interés. No obstante, será importante llevar a cabo una evaluación profunda de la precisión y exactitud en los resultados obtenidos tanto con el grupo de genes aquí reportado como con diversos genes de interés.

Importancia de curar los datos en el análisis filogenético

Los árboles reconstruidos a partir de alineamientos curados a distintos grados reconstruyeron filogenias dispares. Las secuencias no presentaban datos que pudieran compararse directamente al haber sido obtenidas de diversas fuentes y con técnicas distintas (e.g. datos de origen transcriptómica y genómico). De ahí la importancia de llevar a cabo diversos ejercicios buscando un alineamiento que permitiera una reconstrucción más confiable. Esto, así como la importancia de escoger un grupo externo adecuado ha sido señalado previamente en guías sobre la reconstrucción filogenética (Hall, 2004). En este estudio se llevó a cabo el ejercicio de reconstruir las relaciones filogenéticas de los genes partiendo de información 1. Como fue recibida, 2. Acotando a la región conservada, y 3. Utilizando solo los sitios que no presentaran gap en al menos 50% de las

secuencias. Este último método resulto en un árbol reconstruido con valores de Bootstrap más elevados.

Solo el uso de datos adecuadamente curados en conjunto con una selección adecuada de grupo externo permitió llegar a una reconstrucción filogenética con valores de soporte consistentes en la mayoría de los nodos.

La presencia de ortólogos putativos de los genes *PLT* en *M. san-angelensis*: implicaciones en el desarrollo

Apartir de los árboles filogenéticos se infiere que las proteínas PLT junto con las AIL de *A. thaliana* se agrupan en cinco subclados. Dichos subclados corresponden a PLT1/2, PLT3/7, PLT4, PLT5, y al grupo externo ADAP. Resultados similares han sido reportados previamente para los factores de transcripción de la subfamilia AIL pertenecientes a la familia AP2 (Albarrán, 2020; Horstman et al., 2014; Kerstens et al., 2020).

A partir de este análisis se identificó un par de ortólogos putativos de los genes *PLT1* y *PLT2* de *A. thaliana* en el genoma de *M. san-angelensis*. Ortólogos putativos de *PLT1* y *PLT2* han sido reconocidos previamente en decenas de especies vegetales incluyendo cactáceas como *P. pringlei*, *Carnegiea gigantea* y *Lophocereus schottii* (Albarrán, 2020; Kerstens et al., 2020) De acuerdo a lo esperado, el análisis filogenético no fue capaz de discernir entre el ortólogo putativo de *AtPLT1* y el de *AtPLT2*. La presencia de estos genes concuerda con su importante papel pues los factores de transcripción PLT1 y PLT2 son reguladores cruciales del desarrollo de la raíz, por ejemplo participan en la especificación y mantenimiento del nicho de células troncales (Aida et al., 2004; Galinha et al., 2007).

Kerstens (2020) señala en su análisis filogenómico de los factores de transcripción de la familia AP2 que el subclado compuesto por PLT1 y PLT2 presenta poca variación entre las angiospermas, con la excepción de que las monocotiledóneas probablemente no presentan ortólogos de este clado. Dicho autor sugiere que esta ausencia puede tener una cercana relación con el proceso de formación de raíz en las monocotiledóneas, en las cuales se induce la formación de raíces adventicias

dando lugar a un sistema de raíces fibroso. En contraste, las eudicotiledóneas, incluso las de divergencia más antigua como *Amborella trichopoda*, presentan los ortólogos de PLT1 y PLT2 y cuentan con un sistema de raíz pivotante que se desarrolla a partir de la raíz embrionaria (Trueba et al., 2016).

También encontramos dos ortólogos putativos de los genes *PLT3* y *PLT7* de *A. thaliana* en el genoma de *M. san-angelensis*. Ortólogos putativos de *PLT3* y *PLT7* han sido reconocidos previamente en decenas de especies vegetales incluyendo cactáceas como *P. pringlei* (Albarrán, 2020; Kerstens et al., 2020). En este trabajo, además, identificamos los ortólogos putativos de estos genes en el genoma de *S. undatus*. Como ocurre con los ortólogos putativos de *AtPLT1/2*, el análisis filogenético no fue capaz de discernir entre el ortólogo putativo de *AtPLT3* y el de *AtPLT7*. Estos genes están relacionados con el mantenimiento del meristemo apical del brote, la filotaxia y la rizotaxia (Hofhuis et al., 2013; Mudunkothge & Krizek, 2012; Prasad et al., 2011). A futuro, estudiar el número de copias de estos genes presentes en el genoma de *M. san-angelensis*, así como los genes a los cuales se encuentran asociados por sintenia, será de importancia por su relación con la filotaxia en especies vegetales

Se identificó un ortólogo putativo de *PLT5* de *A. thaliana* en *M. san-angelensis*. En el estudio previo se identificaron transcritos de cinco genes *PLT* que se expresan en el ápice de la raíz primaria de *P. pringlei* los cuales no incluían al posible *PLT5* (Rodríguez-Alonso et al., 2018, ver la introducción). Este factor de transcripción, junto con *PLT3* y *PLT7*, en la *A. thaliana* contribuye en la filotaxia y rizotaxia (Prasad et al., 2011; Hofhuis et al., 2013). Se ha sugerido que *PLT5* participa como componente ancestral en el surgimiento y posicionamiento de órganos laterales (Kerstens, 2020). En el presente estudio se identificó *PLT5* en *M. san-angelensis*, hasta ahora no reportada en las cactáceas. Será importante realizar un estudio comparativo para ahondar en la presencia / ausencia de *PLT5* en linajes de cactáceas, así como en los estudios funcionales de los genes *PLT* de cactáceas.

Finalmente, en los árboles de genes construidos podemos observar un clado en el que se encuentra *PLT4* de *A. thaliana* y sus ortólogos putativos en *P. pringlei*, *S.*

undatus y *B. vulgaris*. Sin embargo, en el trabajo presente no se identificó un ortólogo putativo de *PLT4* en el genoma de *M. san-angelensis*. Por otro lado, el ortólogo de este gen ya ha sido identificado en la cactácea *P. pringlei* previamente (Rodríguez-Alonso et al., 2018, Albarrán, 2020). Además, este gen ha sido identificado con anterioridad utilizando *primers* degenerados en *Capsicum annum* (Irikova et al., 2012), y en el genoma de *Medicago truncatula* y utilizando métodos bioinformáticos (Franssen et al., 2015), así como en varias especies más. Existe la posibilidad de que las condiciones de la búsqueda informática no hayan sido las adecuadas para su identificación, que errores en el procesamiento bioinformático hayan impedido su identificación, o que en el genoma de *M. san-angelensis* no exista un ortólogo de *PLT4*. Será necesaria mayor investigación a futuro para determinar la verdadera causa de esta ausencia en el genoma de *M. san-angelensis*, y de confirmarse, entender las implicaciones biológicas del desarrollo que esto conlleva en *Mammillaria*.

En resumen, en este trabajo se identificó un par de genes correspondientes al subclado PLT1/2 en el genoma de *M. san-angelensis*. También se identificaron un par de genes correspondientes al subclado PLT3/7. Se identificó un ortólogo putativo de *PLT5* en *M. san-angelensis*, lo cual no había sido reportado anteriormente en cactáceas. No se determinó un ortólogo putativo para *PLT4* en el genoma de *M. san-angelensis*. Se recomienda cautela al proponer la presencia o ausencia de genes a partir de datos provenientes de genomas que no han sido trabajados a profundidad con diversas herramientas y tecnologías. No obstante, la obtención de genomas con secuenciación de lecturas largas de alta profundidad da mayor certeza de las predicciones bioinformáticas.

La amplificación por PCR como primera aproximación a la confirmación de predicciones bioinformáticas

Fue posible amplificar mediante PCR productos de tamaño similar a lo esperado de los ORF identificados como *MsaPLT1/2.1*, *MsaPLT1/2.2*, *MsaPLT3/7.1*, y *MsaPLT5* a partir de gDNA de *M. san-angelensis* (Tabla 3). Las diferencias en tamaños podrían deber a baja resolución de los fragmentos en los geles. Esto sugiere que el

tratamiento bioinformático de los datos fue robusto y que las secuencias de estos genes existen en el genoma de *M. san-angelensis*. Sin embargo, para confirmar estos hallazgos, se tendrá que secuenciar los fragmentos amplificados. En pocos estudios se han aislado los ortólogos putativos de los genes *PLT* de *A. thaliana* de forma experimental a partir de gDNA, un ejemplo de estos es el aislamiento de *PLT4* en *C. annuum* (Irikova et al., 2012). Los estudios que se centran en ortólogos de los *PLT* o en genes de la familia AP2 suelen emplear datos genómicos, o bases de datos que integran datos de genoma y transcriptoma (Kim et al., 2006). Los ortólogos putativos de las proteínas PLT han sido identificados bioinformáticamente por Kerstens (2020), en el proteoma de 106 especies de angiospermas; y por Albarrán (2020), en un conjunto de datos mixtos pertenecientes a 61 especies de 32 familias de embriofitas incluyendo 4 especies de cactáceas. Con todo lo anterior en consideración, nuestro método bioinformático ha resultado adecuado para la identificación de genes de interés. Sin embargo, como discutimos más adelante, su eficacia depende de las secuencias que se usen como referencia para el diseño de *primers*.

Por otro lado, con los dos *primers* diseñados para *Msa3/7.2* no se obtuvo amplificación de productos de PCR a partir de gDNA. Es por esto que se plantea la posibilidad de que esta secuencia identificada bioinformáticamente sea de naturaleza quimérica. Sin embargo, antes de confirmar o descartar la existencia de esta secuencia se debe de llevar a cabo mayores esfuerzos de estandarización de estas reacciones de PCR, empezando por probar distintos juegos de *primers* como abordaremos más adelante. Otro tipo de análisis como southern blot permitirá tener mayor evidencia sobre la existencia de esta secuencia en el genoma de *M. san-angelensis*. A esta lista se agregan otro tipo de procedimientos como la amplificación y secuenciación de fragmentos más cortos.

En contraste, no fue posible amplificar *Msa1/2.1*, ni *MsaPLT5* al utilizar cDNA de *M. san-angelensis* junto con los juegos de *primers* que previamente habían servido para amplificar en gDNA y el tamaño del fragmento amplificado de *Msa3/7.2* (~500 pb), similar al caso de amplificación a partir de gDNA, fue menor del tamaño

esperado (750 bp). A continuación, planteo diversas explicaciones de estos resultados. Por un lado, la predicción de exones por parte del programa Augustus puede carecer de precisión o exactitud. Stanke y sus colaboradores (2003) han hecho reportes acerca de la sensibilidad y especificidad con las que el programa Augustus realiza predicciones en organismos modelo; un ejemplo de estos son las predicciones en *A. thaliana*. Dichas predicciones cuentan con una sensibilidad a nivel de exón del 89% y especificidad de 70%. Otra evidencia a favor de esta posibilidad es la discordancia entre la longitud teórica y observada en algunos productos amplificados a partir de cDNA. A futuro, existe la posibilidad de utilizar los datos de las secuencias amplificadas mediante PCR en cDNA de *M. san-angelensis* para enriquecer las predicciones el programa y diseñar juegos de *primers* con mayores posibilidades de amplificación.

Por otro lado, existe el escenario de que no haya sido posible amplificar estos productos debido a que no se hallaban presentes en la muestra de cDNA. Ha sido reportado por diversos autores a través de los años (Aida et al., 2004; Du & Scheres, 2017; Durgaprasad et al., 2019; Galinha et al., 2007; Horstman et al., 2014) que la expresión de los genes *PLT* es altamente limitada tanto espacial como temporalmente.. Por ello es probable que no se hallaran transcritos en la muestra al momento de la colecta del tejido vegetal.

El análisis de especificidad de primers como un procedimiento crucial

Un punto importante que tener en cuenta al considerar el alcance de este trabajo es que no se llevó a cabo un análisis usando los *primers* y las lecturas genómicas que permitiera predecir la especificidad de los juegos de *primers*. Si bien el resultado de una PCR puede verse influenciado por muchas condiciones, como el material genético usado y las condiciones de reacción, un punto crítico es la especificidad de los *primers*. Dado que las diferentes partes de los cromosomas o transcritos pueden compartir cierta similitud de nucleótidos es posible que un juego de *primers* destinado a un objetivo también se una a otro dando lugar a amplificación no específica. Varios estudios han investigado los efectos de los desajustes entre los objetivos y *primers* y han demostrado que un producto puede amplificarse incluso

si tiene algunos desajustes con los *primers* (Ghedira et al., 2009; Sipos et al., 2007). Por ello se propone hacer una búsqueda con diferentes herramientas de regiones que puedan dar lugar a amplificación no específica. A este respecto, vale la pena señalar que el programa BLAST no es una herramienta adecuada para este propósito pues su algoritmo de alineamiento local no siempre produce información completa del ajuste entre el primer y una región (Altschul et al., 1997; Ye et al., 2012). Otras herramientas como PRIMEGENS Sequence Specific Primer Design (PRIMEGENS; http://primegens.org/w3/w32/w3_html/SSPD.html) también se puede utilizar para el diseño de cebadores específicos, pero solo para un número limitado de organismos. En futuros trabajos recomendamos el uso de la herramienta Primer-BLAST (Ye et al., 2012), una herramienta que combina BLAST con un algoritmo de alineamiento global, esto aunado a un aumento en la sensibilidad del algoritmo, permite predecir con mayor certeza y precisión amplificaciones inespecíficas.

Perspectivas metodológicas respecto al aislamiento molecular de los genes *MsaPLT* y sus ortólogos

Proponemos finalmente las siguientes modificaciones al protocolo pues la finalidad del trabajo científico debe de ser siempre producir resultados reproducibles y certeros. En las reacciones utilizando cDNA, usando como referencia las pautas señaladas por Udvardi et al., 2008, se debe de considerar: 1) Recolectar material biológico por triplicado 2) Analizar la calidad el RNA total utilizando una electroforesis en gel de agarosa o un análisis un Bioanalyzer 3) Digerir el RNA purificado utilizando una DNAsa para remover la contaminación con gDNA 4) Analizar la cantidad y calidad de cDNA a utilizarse 5) Establecer controles negativos y positivos que puedan ser utilizados para *M. san-angelensis*.

Conclusiones

En este trabajo se llevó a cabo:

- La construcción de una base de datos de secuencias genómicas de *M. san-angelensis* constituida por lecturas largas de alta fidelidad producto de la secuenciación PacBio en la plataforma Sequel II.
- Establecimiento de un protocolo para la predicción *ab-initio* de genes de interés utilizando lecturas largas de alta fidelidad de *M. san-angelensis*.
- La identificación de 5 ortólogos putativos de los genes *PLETHORA* de *A. thaliana* en el genoma de *M. san-angelensis* (*MsaPLT1/2.1*, *MsaPLT1/2.2*, *MsaPLT3/7.1*, *MsaPLT3/7.2* y *MsaPLT5*).
- El diseño de *primers* que permitieron diseñados sobre los ORF *MsaPLT1/2.1*, *MsaPLT1/2.2*, *MsaPLT3/7.1* y *MsaPLT5* que permitieron amplificar productos usando material genético de *M. san-angelensis*; y amplificar productos ortólogos a lo amplificados usando *primers* diseñados sobre los ORF *MsaPLT1/2.1* en *M. lasiacantha* y *M. hernandezii*; a *MsaPLT3/7.1* en *M. coahuilensis*, *M. pectinifera* y *M. hernandezii*; y a *MsaPLT5* en *M. pectinifera*, *M duwei*, *M. lasiacantha* y *M. hernandezii*.

Referencias bibliográficas

- Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nussaume, L., Noh, Y. S., Amasino, R., & Scheres, B. (2004). The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell*, *119*(1), 109–120. <https://doi.org/10.1016/j.cell.2004.09.018>
- Albarrán, R. (2020). *Participación de los genes Plethora en el desarrollo de la raíz primaria de las cactáceas*. Universidad Autónoma del Estado de Morelos.
- Allen, J. E., & Salzberg, S. L. (2005). JIGSAW: integration of multiple sources of evidence for gene prediction. *Bioinformatics (Oxford, England)*, *21*(18), 3596–3603. <https://doi.org/10.1093/bioinformatics/bti609>
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Anderson, E. F. (2001). *The Cactus Family*. Timber Press.
<https://books.google.com.mx/books?id=vYXQHL2IsZ4C>
- Benfey, P. N., Linstead, P. J., Roberts, K., Schiefelbein, J. W., Hauser, M. T., & Aeschbacher, R. A. (1993). Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. *Development (Cambridge, England)*, *119*(1), 57–70. <https://doi.org/10.1242/dev.119.Supplement.57>
- Bennett, T., van den Toorn, A., Willemsen, V., & Scheres, B. (2014). Precise control of plant stem cell activity through parallel regulatory inputs. *Development (Cambridge, England)*, *141*(21), 4055–4064. <https://doi.org/10.1242/dev.110148>
- Berardini, T., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis Information Resource: Making and Mining the ‘Gold Standard’ Annotated Reference Plant Genome. *Genesis (New York, N.Y. : 2000)*, *53*. <https://doi.org/10.1002/dvg.22877>
- Bitonti, M., & Chiappetta, A. (2011). *Root Apical Meristem Pattern: Hormone Circuitry and Transcriptional Networks* (Vol. 72, pp. 37–71). https://doi.org/10.1007/978-3-642-13145-5_2
- Boke, N. H. (1979). Root Glochids and Root Spurs of *Opuntia arenaria* (Cactaceae). *American Journal of Botany*, *66*(9), 1085–1092. <https://doi.org/10.2307/2442575>
- Breslin, P., Wojciechowski, M., & Majure, L. (2021). Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of Mammillaria. *TAXON*, *70*. <https://doi.org/10.1002/tax.12451>
- Butterworth, C. A., & Wallace, R. S. (2004). Phylogenetic studies of Mammillaria (Cactaceae) - Insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *American Journal of Botany*, *91*(7),

1086–1098. <https://doi.org/10.3732/ajb.91.7.1086>

- Chapman, K., Groot, E. P., Nichol, S. A., & Rost, T. L. (2002). Primary root growth and the pattern of root apical meristem organization are coupled. *Journal of Plant Growth Regulation*, 21(4), 287–295. <https://doi.org/10.1007/s00344-002-0036-x>
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18(2), 170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Crawford, B. C. W., Sewell, J., Golembeski, G., Roshan, C., Long, J. A., & Yanofsky, M. F. (2015). Plant development. Genetic control of distal stem cell fate within root and embryonic meristems. *Science (New York, N. Y.)*, 347(6222), 655–659. <https://doi.org/10.1126/science.aaa0196>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). <https://doi.org/10.1093/gigascience/giab008>
- Del Angel, C., Palomino, G., García, A., & Méndez, I. (2006). Nuclear genome size and karyotype analysis in *Mammillaria* species (Cactaceae). *Caryologia*, 59(2), 177–186. <https://doi.org/10.1080/00087114.2006.10797914>
- Dirsch, R., & Stahl, Y. (2015). Function and regulation of transcription factors involved in root apical meristem and stem cell maintenance. *Frontiers in Plant Science*, 6. <https://doi.org/10.3389/fpls.2015.00505>
- Do, J. H., & Choi, D.-K. (2006). Computational approaches to gene prediction. *Journal of Microbiology (Seoul, Korea)*, 44(2), 137–144.
- Drisch, R. C., & Stahl, Y. (2015). Function and regulation of transcription factors involved in root apical meristem and stem cell maintenance . In *Frontiers in Plant Science* (Vol. 6). <https://www.frontiersin.org/articles/10.3389/fpls.2015.00505>
- Du, Y., & Scheres, B. (2017). PLETHORA transcription factors orchestrate de novo organ patterning during Arabidopsis lateral root outgrowth. *Proceedings of the National Academy of Sciences of the United States of America*, 114(44), 11709–11714. <https://doi.org/10.1073/pnas.1714410114>
- Dubrovsky, J. G., North, G. B., & Nobel, P. S. (1998). Root growth, developmental changes in the apex, and hydraulic conductivity for *Opuntia ficus-indica* during drought. *New Phytologist*, 138(1), 75–82. <https://doi.org/10.1046/j.1469-8137.1998.00884.x>
- Durgaprasad, K., Roy, M. V., Venugopal M., A., Kareem, A., Raj, K., Willemsen, V., Mähönen, A. P., Scheres, B., & Prasad, K. (2019). Gradient Expression of Transcription Factor Imposes a Boundary on Organ Regeneration Potential in Plants. *Cell Reports*, 29(2), 453-463.e3.

<https://doi.org/10.1016/j.celrep.2019.08.099>

- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Eissenstat, D. M., Wells, C. E., Yanai, R. D., & Whitbeck, J. L. (2000). Building roots in a changing environment: implications for root longevity. *New Phytologist*, *147*(1), 33–42. [https://doi.org/DOI: 10.1046/j.1469-8137.2000.00686.x](https://doi.org/DOI:10.1046/j.1469-8137.2000.00686.x)
- Feng, K., Hou, X.-L., Xing, G.-M., Liu, J.-X., Duan, A.-Q., Xu, Z.-S., Li, M.-Y., Zhuang, J., & Xiong, A.-S. (2020). Advances in AP2/ERF super-family transcription factors in plant. *Critical Reviews in Biotechnology*, *40*(6), 750–776. <https://doi.org/10.1080/07388551.2020.1768509>
- Franssen, H. J., Xiao, T. T., Kulikova, O., Wan, X., Bisseling, T., Scheres, B., & Heidstra, R. (2015). Root developmental programs shape the *Medicago truncatula* nodule meristem. *Development*, *142*(17), 2941–2950. <https://doi.org/10.1242/dev.120774>
- Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., & Scheres, B. (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature*, *449*(7165), 1053–1057. <https://doi.org/10.1038/nature06206>
- Ghedira, R., Papazova, N., Vuylsteke, M., Ruttink, T., Taverniers, I., & De Loose, M. (2009). Assessment of Primer/Template Mismatch Effects on Real-Time PCR Amplification of Target Taxa for GMO Quantification. *Journal of Agricultural and Food Chemistry*, *57*(20), 9370–9377. <https://doi.org/10.1021/jf901976a>
- Gilbert, S., Mauch, T. J., & Schoenwolf, G. C. (2001). Developmental Biology. Sixth Edition. *American Journal of Medical Genetics*, *99*(2), 170–171. [https://doi.org/https://doi.org/10.1002/1096-8628\(2000\)9999:999<00::AID-AJMG1133>3.0.CO;2-G](https://doi.org/https://doi.org/10.1002/1096-8628(2000)9999:999<00::AID-AJMG1133>3.0.CO;2-G)
- Guerrero, P. C., Majure, L. C., Cornejo-Romero, A., & Hernández-Hernández, T. (2019). Phylogenetic Relationships and Evolutionary Trends in the Cactus Family. *Journal of Heredity*, *110*(1), 4–21. <https://doi.org/10.1093/jhered/esy064>
- Hall, B. G. (2004). *Phylogenetic trees made easy*. WH Freeman.
- Hernández-Hernández, T., Brown, J. W., Schlumpberger, B. O., Eguiarte, L. E., & Magallón, S. (2014). Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytologist*, *202*(4), 1382–1397. <https://doi.org/https://doi.org/10.1111/nph.12752>
- Hernández, H., & Gómez-Hinostrosa, C. (2015). *Mapping the cacti of Mexico. II*.

Mammillaria.

- Hofhuis, H., Laskowski, M., Du, Y., Prasad, K., Grigg, S., Pinon, V., & Scheres, B. (2013). Phyllotaxis and rhizotaxis in *Arabidopsis* are modified by three plethora transcription factors. *Current Biology*, 23(11), 956–962. <https://doi.org/10.1016/j.cub.2013.04.048>
- Horstman, A., Willemsen, V., Boutilier, K., & Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: Hubs in a plethora of networks. *Trends in Plant Science*, 19(3), 146–157. <https://doi.org/10.1016/j.tplants.2013.10.010>
- Hunt, Taylor, Nigel Paul., Charles, Graham., International Cactaceae Systematics Group., D. R. (2006). *The new cactus lexicon*. DH.
- Irikova, T., Grozeva, S., & Denev, I. (2012). Identification of BABY BOOM and LEAFY COTYLEDON genes in sweet pepper (*Capsicum annuum* L.) genome by their partial gene sequences. *Plant Growth Regulation*, 67(2), 191–198. <https://doi.org/10.1007/s10725-012-9676-4>
- Ivanov, V. B., & Dubrovsky, J. G. (2013). Longitudinal zonation pattern in plant roots: conflicts and solutions. *Trends in Plant Science*, 18(5), 237–243.
- Jiang, K., & Feldman, L. J. (2005). REGULATION OF ROOT APICAL MERISTEM DEVELOPMENT. *Annual Review of Cell and Developmental Biology*, 21(1), 485–509. <https://doi.org/10.1146/annurev.cellbio.21.122303.114753>
- Jiang, K., Moe-Lange, J., Hennet, L., & Feldman, L. J. (2016). Salt Stress Affects the Redox Status of *Arabidopsis* Root Meristems. *Frontiers in Plant Science*, 7, 81. <https://doi.org/10.3389/fpls.2016.00081>
- Jiao, X., Zheng, X., Ma, L., Kutty, G., Gogineni, E., Sun, Q., Sherman, B. T., Hu, X., Jones, K., Raley, C., Tran, B., Munroe, D. J., Stephens, R., Liang, D., Imamichi, T., Kovacs, J. A., Lempicki, R. A., & Huang, D. W. (2013). A Benchmark Study on Error Assessment and Quality Control of CCS Reads Derived from the PacBio RS. *Journal of Data Mining in Genomics & Proteomics*, 4(3). <https://doi.org/10.4172/2153-0602.1000136>
- Kareem, A., Durgaprasad, K., Sugimoto, K., Du, Y., Pulianmackal, A. J., Trivedi, Z. B., Abhayadev, P. V., Pinon, V., Meyerowitz, E. M., Scheres, B., & Prasad, K. (2015). PLETHORA genes control regeneration by a two-step mechanism. *Current Biology*, 25(8), 1017–1030. <https://doi.org/10.1016/j.cub.2015.02.022>
- Kenrick, P., & Strullu-Derrien, C. (2014). The Origin and Early Evolution of Roots. *Plant Physiology*, 166(2), 570–580. <https://doi.org/10.1104/pp.114.244517>
- Kerstens, M. H. L., Schranz, M. E., & Bouwmeester, K. (2020). Phylogenomic analysis of the APETALA2 transcription factor subfamily across angiosperms reveals both deep conservation and lineage-specific patterns. *The Plant Journal*, 103(4), 1516–1524. <https://doi.org/10.1111/tpj.14843>
- Kim, S., Soltis, P. S., Wall, K., & Soltis, D. E. (2006). Phylogeny and Domain Evolution in the APETALA2-like Gene Family. *Molecular Biology and*

- Evolution*, 23(1), 107–120. <https://doi.org/10.1093/molbev/msj014>
- Krizek, B. A., & Eaddy, M. (2012). AINTEGUMENTA-LIKE6 regulates cellular differentiation in flowers. *Plant Molecular Biology*, 78(3), 199–209. <https://doi.org/10.1007/s11103-011-9844-3>
- Li, P., Quan, X., Jia, G., Xiao, J., Cloutier, S., & You, F. M. (2016). RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics*, 17(1), 852. <https://doi.org/10.1186/s12864-016-3197-x>
- Liang, C., Mao, L., Ware, D., & Stein, L. (2009). Evidence-based gene predictions in plant genomes. *Genome Research*, 19(10), 1912–1923. <https://doi.org/10.1101/gr.088997.108>
- Mähönen, A. P., Tusscher, K. Ten, Siligato, R., Smetana, O., Díaz-Triviño, S., Salojärvi, J., Wachsman, G., Prasad, K., Heidstra, R., & Scheres, B. (2014). PLETHORA gradient formation mechanism separates auxin responses. *Nature*, 515(7525), 125–129. <https://doi.org/10.1038/nature13663>
- Makkena, S., & Lamb, R. S. (2013). The bHLH transcription factor SPATULA regulates root growth by controlling the size of the root meristem. *BMC Plant Biology*, 13(1), 1. <https://doi.org/10.1186/1471-2229-13-1>
- Malamy, J. (2005). Intrinsic and environmental response pathways that regulate root system architecture. *Plant, Cell & Environment*, 28, 67–77. <https://doi.org/10.1111/j.1365-3040.2005.01306.x>
- Martínez-Vázquez, O., & Rubluo, A. (1989). In-vitro mass propagation of the near-extinct *Mammillaria san-angelensis* Sánchez-Mejorada. *Journal of Horticultural Science*, 64(1), 99–105. <https://doi.org/10.1080/14620316.1989.11515933>
- Mouchel, C. F., Briggs, G. C., & Hardtke, C. S. (2004). Natural genetic variation in *Arabidopsis* identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. *Genes & Development*, 18(6), 700–714. <https://doi.org/10.1101/gad.1187704>
- Mudunkothge, J. S., & Krizek, B. A. (2012). Three *Arabidopsis* AIL/PLT genes act in combination to regulate shoot apical meristem function. *The Plant Journal : For Cell and Molecular Biology*, 71(1), 108–121. <https://doi.org/10.1111/j.1365-313X.2012.04975.x>
- Nakajima, K., Sena, G., Nawy, T., & Benfey, P. N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. *Nature*, 413(6853), 307–311. <https://doi.org/10.1038/35095061>
- Nibau, C., Gibbs, D., & Coates, J. (2008). Branching out in new directions: The control of root architecture by lateral root formation. *The New Phytologist*, 179, 595–614. <https://doi.org/10.1111/j.1469-8137.2008.02472.x>
- Okamuro, J. K., Caster, B., Villarroel, R., Van Montagu, M., & Jofuku, K. D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proceedings of the National Academy of Sciences of*

- the United States of America*, 94(13), 7076–7081.
<https://doi.org/10.1073/pnas.94.13.7076>
- Palomino, G., Dolezel, J., Cid, R., Brunner, I., Méndez, I., & Rubluo, A. (1999). Nuclear genome stability of *Mammillaria san-angelensis* (Cactaceae) regenerants induced by auxins in long-term in vitro culture. *Plant Science*, 141(2), 191–200. [https://doi.org/https://doi.org/10.1016/S0168-9452\(98\)00216-7](https://doi.org/https://doi.org/10.1016/S0168-9452(98)00216-7)
- Pasha, A., Subramaniam, S., Cleary, A., Chen, X., Berardini, T., Farmer, A., Town, C., & Provart, N. (2020). Araport Lives: An Updated Framework for Arabidopsis Bioinformatics. *The Plant Cell*, 32(9), 2683–2686.
<https://doi.org/10.1105/tpc.20.00358>
- Péret, B., Desnos, T., Jost, R., Kanno, S., Berkowitz, O., & Nussaume, L. (2014). Root Architecture Responses: In Search of Phosphate. *Plant Physiology*, 166(4), 1713–1723. <https://doi.org/10.1104/pp.114.244541>
- Posada, D. (2008). jModelTest: phylogenetic model averaging. In *Molecular biology and evolution* (Vol. 25, Issue 7, pp. 1253–1256).
<https://doi.org/10.1093/molbev/msn083>
- Prasad, K., Grigg, S. P., Barkoulas, M., Yadav, R. K., Sanchez-Perez, G. F., Pinon, V., Blilou, I., Hofhuis, H., Dhonukshe, P., Galinha, C., Mähönen, A. P., Muller, W. H., Raman, S., Verkleij, A. J., Snel, B., Reddy, G. V., Tsiantis, M., & Scheres, B. (2011). Arabidopsis PLETHORA transcription factors control phyllotaxis. *Current Biology*, 21(13), 1123–1128.
<https://doi.org/10.1016/j.cub.2011.05.009>
- Reinhardt, D., & Rost, T. (1995). Developmental Changes of Cotton Root Primary Tissues Induced by Salinity. *International Journal of Plant Sciences - INT J PLANT SCI*, 156. <https://doi.org/10.1086/297273>
- Rodriguez-Alonso, G., Matvienko, M., López-Valle, M. L., Lázaro-Mixteco, P. E., Napsucialy-Mendivil, S., Dubrovsky, J. G., & Shishkova, S. (2018). Transcriptomics insights into the genetic regulation of root apical meristem exhaustion and determinate primary root growth in *Pachycereus pringlei* (Cactaceae). *Scientific Reports*, 8(1), 1–11. <https://doi.org/10.1038/s41598-018-26897-1>
- Sabatini, S., Heidstra, R., Wildwater, M., & Scheres, B. (2003). SCARECROW is involved in positioning the stem cell niche in the Arabidopsis root meristem. *Genes & Development*, 17(3), 354–358. <https://doi.org/10.1101/gad.252503>
- Sablowski, R. (2007). Flowering and determinacy in Arabidopsis. *Journal of Experimental Botany*, 58(5), 899–907. <https://doi.org/10.1093/jxb/erm002>
- Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochemical and Biophysical Research*

- Communications*, 290(3), 998–1009. <https://doi.org/10.1006/bbrc.2001.6299>
- Sánchez-Calderón, L. (2013). *Root Development and Abiotic Stress Adaptation* (M. E. Ibarra-Cortés (ed.); p. Ch. 5). IntechOpen. <https://doi.org/10.5772/55043>
- Santuari, L., Sanchez-Perez, G. F., Luijten, M., Rutjens, B., Terpstra, I., Berke, L., Gorte, M., Prasad, K., Bao, D., Timmermans-Hereijgers, J. L. P. M., Maeo, K., Nakamura, K., Shimotohno, A., Pencik, A., Novak, O., Ljung, K., van Heesch, S., de Bruijn, E., Cuppen, E., ... Heidstra, R. (2016). The PLETHORA gene regulatory network guides growth and cell differentiation in Arabidopsis roots. *Plant Cell*, 28(12), 2937–2951. <https://doi.org/10.1105/tpc.16.00656>
- Sarkar, A. K., Luijten, M., Miyashima, S., Lenhard, M., Hashimoto, T., Nakajima, K., Scheres, B., Heidstra, R., & Laux, T. (2007). Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. *Nature*, 446(7137), 811–814. <https://doi.org/10.1038/nature05703>
- Scheres, B., Benfey, P., & Dolan, L. (2002). Root Development. *The Arabidopsis Book*, 2002(1). <https://doi.org/10.1199/tab.0101>
- Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E. H., Schmid, M., Jürgens, G., & Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature*, 464(7290), 913–916. <https://doi.org/10.1038/nature08836>
- Sena, G., Wang, X., Liu, H.-Y., Hofhuis, H., & Birnbaum, K. (2009). Organ regeneration does not require a functional stem cell niche in plants. *Nature*, 457, 1150–1153. <https://doi.org/10.1038/nature07597>
- Shigyo, M., Hasebe, M., & Ito, M. (2006). Molecular evolution of the AP2 subfamily. *Gene*, 366(2), 256–265. <https://doi.org/10.1016/j.gene.2005.08.009>
- Shishkova, S., Rost, T. L., & Dubrovsky, J. G. (2008). Determinate root growth and meristem maintenance in angiosperms. *Annals of Botany*, 101(3), 319–340. <https://doi.org/10.1093/aob/mcm251>
- Shishkova, Svetlana, Las Peñas, M. L., Napsucialy-Mendivil, S., Matvienko, M., Kozik, A., Montiel, J., Patiño, A., & Dubrovsky, J. G. (2013). Determinate primary root growth as an adaptation to aridity in Cactaceae: Towards an understanding of the evolution and genetic control of the trait. *Annals of Botany*, 112(2), 239–252. <https://doi.org/10.1093/aob/mct100>
- Simpson, M. G. (2010). *Plant Systematics*. Elsevier Science. <https://books.google.com.mx/books?id=dj8KRlmgYf4C>
- Sipos, R., Székely, A. J., Palatinszky, M., Révész, S., Márialigeti, K., & Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiology Ecology*, 60(2), 341–350. <https://doi.org/10.1111/j.1574-6941.2007.00283.x>
- Srinivasan, C., Liu, Z., Heidmann, I., Supena, E. D. J., Fukuoka, H., Joosen, R.,

- Lambalk, J., Angenent, G., Scorza, R., Custers, J. B. M., & Boutilier, K. (2007). Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (*Nicotiana tabacum* L.). *Planta*, *225*(2), 341–351. <https://doi.org/10.1007/s00425-006-0358-1>
- Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644. <https://doi.org/10.1093/bioinformatics/btn013>
- Steffens, B., & Rasmussen, A. (2016). The Physiology of Adventitious Roots. *Plant Physiology*, *170*(2), 603–617. <https://doi.org/10.1104/pp.15.01360>
- Taiz, L., & Zeiger, E. (2003). *Plant physiology*. 3rd edn. Oxford University Press.
- Trueba, S., Isnard, S., Barthélémy, D., & Olson, M. E. (2016). Trait coordination, mechanical behaviour and growth form plasticity of *Amborella trichopoda* under variation in canopy openness. *AoB PLANTS*, *8*, plw068. <https://doi.org/10.1093/aobpla/plw068>
- Tsukagoshi, H., Busch, W., & Benfey, P. N. (2010). Transcriptional Regulation of ROS Controls Transition from Proliferation to Differentiation in the Root. *Cell*, *143*(4), 606–616. <https://doi.org/10.1016/j.cell.2010.10.020>
- Udvardi, M. K., Czechowski, T., & Scheible, W.-R. (2008). Eleven golden rules of quantitative RT-PCR. In *The Plant cell* (Vol. 20, Issue 7, pp. 1736–1737). <https://doi.org/10.1105/tpc.108.061143>
- Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., & Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, *46*(D1), D1190–D1196. <https://doi.org/10.1093/nar/gkx1002>
- van den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P., & Scheres, B. (1997). Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature*, *390*(6657), 287–289. <https://doi.org/10.1038/36856>
- Verbelen, J.-P., De Cnodder, T., Le, J., Vissenberg, K., & Baluska, F. (2006). The Root Apex of *Arabidopsis thaliana* Consists of Four Distinct Zones of Growth Activities: Meristematic Zone, Transition Zone, Fast Elongation Zone and Growth Terminating Zone. *Plant Signaling & Behavior*, *1*(6), 296–304. <https://doi.org/10.4161/psb.1.6.3511>
- Wang, Z., Mao, J.-L., Zhao, Y.-J., Li, C.-Y., & Xiang, C.-B. (2015). L-Cysteine inhibits root elongation through auxin/PLETHORA and SCR/SHR pathway in *Arabidopsis thaliana*. *Journal of Integrative Plant Biology*, *57*(2), 186–197. <https://doi.org/10.1111/jipb.12213>
- Xiong, Y., McCormack, M., Li, L., Hall, Q., Xiang, C., & Sheen, J. (2013). Glucose–TOR signalling reprograms the transcriptome and activates meristems. *Nature*, *496*(7444), 181–186. <https://doi.org/10.1038/nature12030>
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L.

(2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134. <https://doi.org/10.1186/1471-2105-13-134>

Zhou, Y., Zhang, J., Xiong, X., Cheng, Z.-M., & Chen, F. (2022). De novo assembly of plant complete genomes. *Tropical Plants*, 1(1), 1–8. <https://doi.org/10.48130/TP-2022-0007>

Zluhan-Martínez, E., López-Ruíz, B. A., García-Gómez, M. L., García-Ponce, B., de la Paz Sánchez, M., Álvarez-Buylla, E. R., & Garay-Arroyo, A. (2021). Integrative Roles of Phytohormones on Cell Proliferation, Elongation and Differentiation in the *Arabidopsis thaliana* Primary Root. *Frontiers in Plant Science*, 12, 659155. <https://doi.org/10.3389/fpls.2021.659155>

Anexos

Gen	Primer	Secuencia	%GC	Tm (°C)
MhPLT1/2.1	LCMhPLT1/2.1DF	AACAACAACAGCGTCATGGC	50	56.7
	LCMhPLT1/2.1DR	TTAGCTGCCCCTGACCCTAT	55	57.9
	LCMhPLT1/2.1F	CCGCTTGGAATTGCAGGTTT	50	56.6
	LCMhPLT1/2.1R	ACCAATCGGTCAAAGCAAGC	50	56.3
MhPLT1/2.2	LCMhPLT1/2.2DF	TCTCTTCACCGACAACAGCC	55	57.2
	LCMhPLT1/2.2DR	GGCCTCTGAACTTGATGGCT	55	57.5
	LCMhPLT1/2.2F	AGAGTGTGTGGGTGAGAGGA	55	57.5
	LCMhPLT1/2.2R	CCACCATTATGGGGTCCCTG	60	57.7
MhPLT3/7.1	LCMhPLT3/7.1DF	CGGCCAGAGAACCTCCATTT	55	57.4
	LCMhPLT3/7.1DR	CCGAAGTGATGCAGGAGGTT	55	57.4
	LCMhPLT3/7.1F	CCCCACTACTTGACGAGTG	60	57.3
	LCMhPLT3/7.1R	TACGCCCTTCTTAAACCCGG	55	57.1
MhPLT3/7.2	LCMhPLT3/7.2DF	CGCAAATTACTTCGCCGACC	55	57
	LCMhPLT3/7.2DR	GGCTGCTATGTCGTATGCCT	55	57.2
	LCMhPLT3/7.2F	TGTCTCCACTCTTCGCAAC	55	57.2
	LCMhPLT3/7.2R	TCTTCGGCGGGTTAGTAGGA	55	57.5
MhPLT5	LCMhPLT5DF	GGTCGAAGGCTTTGGTCAGA	55	57.2
	LCMhPLT5DR	CCACCGTCACACCCATTGTA	55	57.3
	LCMhPLT5F	TCTTTGTTCTCCACCGCCTC	55	57.2
	LCMhPLT5R	CTCGCACTTTCTCCCTCCTC	60	57.3

Anexo 1. Especificaciones de los primers utilizados