



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS



## INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS

Control Escolar de Licenciatura



### VOTOS DE APROBATORIOS

Secretaria Ejecutiva del Instituto de Investigación en Ciencias Básicas Aplicadas de la Universidad Autónoma del Estado de Morelos.  
P r e s e n t e .

Por medio de la presente le informamos que después de revisar la versión escrita de la tesis que realizó el C. **VILLA NUÑEZ JUAN PABLO** con número de matrícula **20154009058** cuyo título es:

"Regulación postranscripcional mediada por microRNAs en el crecimiento determinado de la raíz primaria de *Pachycereus pringlei*"

Consideramos que **SI** reúne los méritos que son necesarios para continuar los trámites para obtener el título de **LICENCIADO EN CIENCIAS ÁREA TERMINAL DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR**

Cuernavaca, Mor a 26 de enero del 2022

Atentamente  
Por una universidad culta

Se adiciona página con la e-firma UAEM de los siguientes:

DRA. CARMEN NINA PASTOR COLÓN  
DRA. VERÓNICA LIRA RUAN  
DRA. SVETLANA SHISHKOVA  
DR. JORGE LUIS FOLCH MALLOL  
DR. GUSTAVO RODRÍGUEZ ALONSO

**PRESIDENTE**  
**SECRETARIO**  
**VOCAL**  
**PRIMER SUPLENTE**  
**SEGUNDO SUPLENTE**



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

### Sello electrónico

**JORGE LUIS FOLCH MALLOL** | Fecha:2022-01-26 08:56:01 | Firmante

Ot7xqTWNkXEXE/5alEvCpS2j4SLHvx8wvxq8vCdFTvSLn7GZsRJJLJcj7hG3ohePKYnVvBBRLw8TjkMaMkiASSJR8zg+MpyUOKeBgtiyM8O4J2ug5jyWDQRtqir8l5lqVlt9zgu+YyU0sgBnwrpgXET7O2p8Lq0vB4c9gYZTfO1aJ/G6LyW/Qju+MOnvQh7Pylp814LiSWFGaD1uech6UQR5cPIDU50G4qD17bBQjxLJXsd4CACkwoZM0LR5CnviEt9D682h4cPkVOvXhfuRo3QYSiQaH/GcLUOTtdS8qyK7bf/YOE8DVzITc+NT/YzLKuU39bVT/IMzx6PRFVBA==

**CARMEN NINA PASTOR COLON** | Fecha:2022-01-26 09:14:26 | Firmante

upsVrF4vOymVOr2+bgjapzdzuTXKy+PkQLan7oJ/UKxIG9H9q63hXfdiDnFfr+HJBhN1r17TCiNOvuoH59au0qg5BvICKtN45CT40k7FEFe19UadK8CmKhvLiC9LwbtJ+ZiyZSTsqo7Qf0byY+vRAHsiU6QGR2ThlxccaP57r7vNwRxBoPEuUOsMjBkbnNM1pvUheJ8lo3TC95iYw5CAqLUVGFu5H6lirgQu5PPsz1bzlnSE3Q1BA5QGcadCmiC/7pTsTUykPx9id7eTiC74ysZpqVWnMzd1Gko2lKvvh6HyFMTumZEZoEYhN7aUnq46dL3TtnbVvuU7RXf3OQ==

**SVETLANA SHISHKOVA .** | Fecha:2022-03-13 12:28:57 | Firmante

UvzLdufvMpwrgGxP7xNOqEMiNkaO2gEFqH1IP6h/b2/2GBKJzj3fbfYRQQWcftOPXevp3oji6DVMmQIIIRImfe9L0SLZ97GBYY5jxWPty5oUx8XrPKk4i4PI5WmEC3MCbTdRD4A5QQ//nCUfRnqYxoMv3hmlkK9K0vuVBlamfBStburLq4KYdAot+3BrAmqWkLNdOZksRM5Ardj5emf6l3XackoGUnlCV9XTc/dzyllR7CCoTOF0DQpVe0myW3naK7gCkw3CuUej+MupgFRy3euN+fSdxVaMiN3s0NzEWXYiT2kPOvnX9xSxN19sJCWfWFwN3uXChXecPJuZv1w==

**GUSTAVO RODRÍGUEZ ALONSO** | Fecha:2022-03-13 13:08:06 | Firmante

P5dNoRI/odVU9u8SLQRUjHZZwG5K8ZDJTjx5zlwDjW4wocB+tNZqET/BlfvaiJG1aNvIRtk1/56A9442d2vubi51ZYSwFqNwOX9AiL+vR3iyboKec4r3ZU2A5WwIDkdix7ZlrouoOp34WPY32PHop7QTNwAHfSO9KEFHZSjadXrvKDjMyalm6tzWg7fjLPhrjKMylDJeI+mApOwclVFMrR4dqWivR+NG9YclUJSBF+YL2wTGkn1Yibg5hOzkbY0Nuktghxi4vf1ypqqujOTllqzKaqDH/ot8z7+8NcLtkoeipEL2TxVevcoPWxyA3G1BpTuq5udy/gumeWJcQ==

**VERONICA LIRA RUAN** | Fecha:2022-03-30 10:13:55 | Firmante

Py2g8//ScxRFd9x9OTm0tff4xXx1hctFmxUxUlmjiEXypgj0zjW25096MS+XrxtuWr5WxjzeCtg0mXZZMYexqma/pwF6iukYD3h1RQ7zcqo5B3BmpczTmk4ck+HRMX9FPLDDzPRwqM6SIG8bRGNsFV37XJxLv21HPZNxa2ITpcXi30yByvcH3KJIO8iQoKIGwyzPjTOu7oXv2xbO2xhO/eumbI5sePqzCmTw0G5o8yjabeEZ+3JRRfsKAdrKjuUuBE2YliZ/FE52GUGh+YDx4ntWBC8GRryu3yXz+63u/J/stgsF/Hnys+XFxdyfyK6Z1Dsk5VZGITSuJsP0YgEQ==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



yIEp6Lca8

<https://efirma.uaem.mx/noRepudio/v9vcNudDzEwZyP5SnyZyqON8qaYE8po7>





UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS  
INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y  
APLICADAS  
CENTRO DE INVESTIGACIÓN EN DINÁMICA CELULAR

**Regulación postranscripcional mediada por microRNAs en el  
crecimiento determinado de la raíz primaria de *Pachycereus  
pringlei***

# TESIS

para obtener el título de:  
**LICENCIADO EN CIENCIAS**  
**ÁREA TERMINAL BIOQUÍMICA Y BIOLOGÍA MOLECULAR**

PRESENTA  
**JUAN PABLO VILLA NÚÑEZ**

DIRECTORA DE TESIS:  
Dra. Svetlana Shishkova

- **Presidente.** Dra. Carmen Nina Pastor Colón
  - **Secretario.** Dra. Verónica Lira Ruan
  - **Suplente 1.** Dr. Jorge Luis Folch Mallol
  - **Suplente 2.** Dr. Gustavo Rodríguez Alonso

La presente tesis se realizó en el departamento de Biología Molecular de Plantas del Instituto de Biotecnología-UNAM, bajo la asesoría de la Dra. Svetlana Shishkova.

Este proyecto fue parcialmente financiado por los donativos UNAM-DGAPA-PAPIIT IN201318, IN200818 y CONACyT CF2019-304301.

Juan Pablo Villa Núñez recibió una beca del proyecto IN200818 por concepto de Titulación de Licenciatura.

Agradecemos al Dr. Francisco Molina-Freaner y el Q.B. José Fulgencio Martínez Rodríguez por la donación de semillas de *Pachycereus pringlei* y a la Dra. Selene Napsucialy Mendivil por el apoyo técnico.

## Agradecimientos

Primero quiero agradecer a las personas que fueron una parte importante de mi formación académica: i) a los miembros de mi comité tutorial, Dra. Verónica Lira, Dra. Nina Pastor y Dr. Jordi Folch, por sus comentarios y su comprensión; ii) al Dr. Joseph Dubrovsky y a mis compañeros de laboratorio, Héctor M., Selene N., Ramces G., Alejandra L., Sofía H., y Ramsés A., por sus cálidas palabras y apoyo técnico cuando estuve en el laboratorio; y iii) especialmente a la Dra. Svetlana Shishkova y al Dr. Gustavo Rodríguez, quienes fungieron como asesores principales de este trabajo y estuvieron conmigo en los momentos más complicados.

Agradezco particularmente al Dr. Gustavo Rodríguez, quien se ha convertido en mi gran maestro y amigo, por haberme enseñado el camino de la ciencia en aquellos días del año 2010. La ciencia es, como dijo Carl Sagan: “lo máspreciado que tenemos”, pero es aún más importante nuestra salud física y mental, como Gustavo dice.

Extiendo mi agradecimiento a las personas que me apoyaron emocionalmente y que fueron los responsables de que no me haya dado por vencido. Doy gracias a Catalina Bucarey, quien ocupa un lugar especial en mi corazón, porque fue la luz en mis días más oscuros. Y agradezco a Andrid Clavijo porque creyó en mí cuando yo era incapaz de hacerlo.

Finalmente, agradezco profundamente a mi familia. Sin ellos yo no sería nada de lo que soy ahora. A mi padre, Angel V. M., porque siempre me dijo: “estudia”. A mi madre, María N. L., por hacer por mí hasta prácticamente lo que no le corresponde. A mi hermano, Luis V., por todos sus consejos y palabras de aliento. Y a mis mascotas, las que sólo puedo acariciar en mis recuerdos y las que afortunadamente aún me esperan en casa, porque son la alegría de mi vida.

# Índice

Índice de figuras .....	IV
Índice de tablas .....	V
Abreviaturas .....	VI
Resumen .....	VII
Abstract .....	VIII
1. Introducción .....	1
1.1 Generalidades de la raíz .....	1
1.1.1 Zonas de desarrollo de la raíz.....	1
1.1.2 Meristemo apical de la raíz .....	2
1.2 Crecimiento indeterminado y determinado de la raíz.....	4
1.3 Cactáceas .....	4
1.3.1 <i>Pachycereus pringlei</i> .....	4
1.4 Regulación postranscripcional de la expresión génica mediada por microRNAs .....	7
2. Antecedentes.....	10
3. Justificación .....	11
4. Objetivos.....	12
4.1 Objetivo general .....	12
4.2 Objetivos particulares .....	12
5. Materiales y métodos.....	13
5.1 Identificación de los transcritos <i>SPL</i> de <i>Pachycereus pringlei</i> .....	13
5.2 Extensión del marco de lectura de los transcritos .....	13
5.3 Predicción de transcritos blancos de miR156 .....	15
6. Resultados y discusión .....	16
6.1 Identificación de los transcritos <i>SPL</i> de <i>Pachycereus pringlei</i> .....	16
6.2 Extensión del marco de lectura de los transcritos incompletos .....	19
6.3 Árbol filogenético de las proteínas <i>SPL</i> .....	26
6.4 Predicción de posibles transcritos blancos de ppr-miR156.....	28
7. Conclusiones .....	36
8. Perspectivas a corto plazo .....	37
9. Bibliografía.....	38

## Índice de figuras

**Figura 1.** Organización de la raíz primaria de *Arabidopsis thaliana*.

**Figura 2.** El crecimiento determinado de la raíz primaria de *Pachycereus pringlei* se puede dividir en tres etapas del desarrollo.

**Figura 3.** Biogénesis de microRNAs y silenciamiento génico postranscripcional.

**Figura 4.** Marco de lectura de los transcritos *PprSPLs*, *CgiSPL6b* y *PprMATE*.

**Figura 5.** Árbol filogenético de las proteínas SPL de *Arabidopsis* y *P. pringlei/C. gigantea*.

**Figura 6.** Posición del sitio de unión de ppr-miR156 en los seis posibles transcritos-blancos *PprSPLs* y el posible transcrito-blanco *CgiSPL6b*.

**Figura 7.** Conservación de las secuencias de los sitios de unión de los posibles transcritos-blancos *PprSPLs*, *CgiSPL6b* y el transcrito *PprMATE*.

**Figura 8.** Conservación de las secuencias de los sitios de unión de ppr-miR156 en los ortólogos *CgiSPLs*.

## Índice de tablas

**Tabla 1.** Contigs del transcriptoma de *P. pringlei* identificados por el tBLASTn con la sonda AthSPL3.

**Tabla 2.** Anotación de los transcritos *PprSPLs* basada en mejor hit unidireccional utilizando BLASTn con el genoma de *Arabidopsis*.

**Tabla 3.** Hits (scaffolds de *P. pringlei*) de la búsqueda por BLASTn utilizando a los seis transcritos incompletos *PprSPLs* como sonda.

**Tabla 4.** Hits (scaffolds de *C. gigantea*) de la búsqueda por BLASTn utilizando a los seis transcritos incompletos *PprSPLs* como sonda para interrogar al genoma de *C. gigantea*.

**Tabla 5.** Porcentajes de identidad y similitud (calculada con la matriz BLOSUM62) entre las proteínas SPL de *Pachycereus pringlei* y *Carnegiea gigantea*.

**Tabla 6.** Alineamientos del sitio de unión de ppr-miR156a en los transcritos *PprSPLs* y *CgjSPL6b*.

**Tabla 7.** Alineamientos del sitio de unión de ppr-miR156b en los transcritos *PprSPLs* y *CgjSPL6b*.

**Tabla 8.** Alineamientos del sitio de unión de ppr-miR156c en los transcritos *PprSPLs* y *CgjSPL6b*.

**Tabla 9.** Alineamientos del sitio de unión de ppr-miR156d en los transcritos *PprSPLs* y *CgjSPL6b*.

**Tabla 10.** Alineamientos del sitio de unión de ppr-miR156e en los transcritos *PprSPLs* y *CgjSPL6b*.

**Tabla 11.** Alineamientos del posible sitio de unión de las cinco isoformas de ppr-miR156 en *PprMATE*.



## Abreviaturas

RAM: root apical meristem / meristemo apical de la raíz

*Arabidopsis*: *Arabidopsis thaliana*

CQ: centro quiescente

d.d.g.: días después de la germinación

miRNAs: microRNAs

pri-miRNA: miRNA primario

DCL1: DICER-LIKE 1

pre-miRNA: precursor del miRNA

HEN1: HUA ENHANCER 1

AGO: ARGONAUTA

RISC: RNA-induced silencing complex / complejo de silenciamiento inducido por RNA

smallRNA-seq: secuenciación de RNAs pequeños

SPL: SQUAMOSA PROMOTER BINDING PROTEIN LIKE

MATE: MULTIDRUG AND TOXIC COMPOUND EXTRUSION-TYPE TRANSPORTER

UTR: Un-Translated Regions / regiones no traducidas

CDS: CoDing Sequence / región codificante

Clustal-O: Clustal Omega

MFE: minimum free energy value / valor mínimo de energía libre

RLM-RACE (RNA Ligase-Mediated - Rapid Amplification of cDNA Ends)

## Resumen

La raíz es un órgano importante para las plantas vasculares porque mantiene a la planta anclada a su sustrato y absorbe agua y nutrientes que son transportados al resto de la planta para su supervivencia. Con base en su origen, las raíces se pueden clasificar como: raíz primaria, cuando se origina a partir de la radícula; raíz lateral, cuando se origina a partir de otra raíz; y raíz adventicia, cuando se origina de un órgano distinto a la raíz. En la mayoría de las plantas angiospermas la raíz primaria puede crecer de manera indefinida debido a la presencia y mantenimiento del meristemo apical de la raíz (RAM, por sus siglas en inglés) a lo largo de la vida de este órgano, es decir, exhibe crecimiento indeterminado. Un caso especial ocurre en la mayoría de las especies de cactáceas, en las que la raíz primaria deja de crecer pocos días después de la germinación porque el RAM se agota y, como consecuencia, todas las células del ápice de la raíz se diferencian. Este tipo de crecimiento se llama crecimiento determinado, y se ha propuesto que les confiere una ventaja adaptativa a los cactus en ambientes áridos. El programa genético que determina el agotamiento del RAM no se conoce. En este trabajo se hizo un análisis bioinformático sobre la regulación postranscripcional mediada por microRNAs en el crecimiento determinado de la raíz primaria del cactus *Pachycereus pringlei*. Se analizó la posible existencia de dos módulos microRNA-blancos, ppr-miR156–*PprSPL* y ppr-miR156–*PprMATE*, utilizando datos del transcriptoma y microtranscriptoma del ápice de la raíz primaria de *P. pringlei* en las etapas de desarrollo cuando el RAM está presente y activo y cuando el RAM está completamente agotado. Se escogió al módulo miR156–*SPL* porque está conservado desde musgos hasta plantas angiospermas, y a un posible módulo nuevo no reportado en la literatura, miR156–*MATE*, porque fue sugerida su existencia en un trabajo previo del laboratorio. En el transcriptoma del ápice de la raíz primaria de *P. pringlei* se identificaron 11 contigs que se anotaron como *PprSPLs*; seis de estos contigs, o transcritos *PprSPLs*, más uno sugerido por el análisis con el genoma de otro cactus, *C. gigantea*, utilizando la secuencia de un transcrito incompleto *PprSPL*, fueron identificados como posibles transcritos-blancos de cinco isoformas de ppr-miR156 identificadas en el trabajo

previo en el laboratorio. Además, el análisis bioinformático realizado en este trabajo sugiere que el transcrito *PprMATE* no es blanco de ninguna de las cinco isoformas de ppr-miR156. Estos resultados sugieren que el módulo miR156–*SPL* está conservado y es funcional en la raíz de *P. pringlei*, y, por lo tanto, en especies de cactáceas que presentan crecimiento determinado de la raíz primaria. Para confirmar la existencia del módulo ppr-miR156–*PprSPL* y rectificar datos sobre la posible existencia del módulo ppr-miR156–*PprMATE* es necesario hacer un análisis experimental.

## **Abstract**

The root is an essential organ of vascular plants as it anchors the plant to its substrate and absorbs water and nutrients necessary for the life of the plant. Roots can be classified as: (a) primary root formed as a radicle during embryogenesis, in many plant species it is the first root which emerges from the seed upon germination; (b) lateral root formed within a parental root; and (c) adventitious root which originates from any organ other than the root. In most angiosperms the primary root can grow for long periods due to the presence and maintenance of the root apical meristem (RAM) throughout the life of this organ, that is, the root exhibits indeterminate growth. However, in many Cactaceae species the primary root stops growing a few days after germination due to RAM exhaustion and subsequent differentiation of all the cells in the root apex. This type of growth is called determinate root growth and it has been proposed to provide cacti an advantage for survival in arid environments. The genetic program underlying RAM exhaustion and determinate root growth is not known. In this work, a bioinformatic analysis was performed to examine post-transcriptional regulation mediated by microRNAs in the determinate growth of the primary root of *Pachycereus pringlei*, a cactus species. Transcriptome and microtranscriptome data from the primary root apex of *P. pringlei* at two developmental stages, when the RAM is present and when it is fully exhausted, were used to evaluate the existence of two microRNA–target modules, ppr-miR156–*PprSPL* and ppr-miR156–*PprMATE*. The miR156–*SPL* interaction is widely conserved, from mosses to angiosperm plants,

and the existence of the new miR156-*MATE* module not yet reported in other plant species was suggested in a previous work in our laboratory. During this project, 11 contigs were identified in the transcriptome of the primary root apex of *P. pringlei* that were annotated as *PprSPLs*. Six of these *PprSPLs* contigs, or transcripts, and one more transcript deduced from the genomic sequence of a closely related cactus, *C. gigantea*, using a *PprSPL* contig sequence with incomplete ORF, were recognized as putative targets of the five ppr-miR156 isoforms identified in previous work in our laboratory. Furthermore, analysis of the putative miRNA binding site of the *PprMATE* transcript performed in this work suggests that the probability of existence of miR156-*PprMATE* module involving any of the five ppr-miR156 isoforms is very low. These results suggest that the miR156-*SPL* module is conserved in *P. pringlei*, and, therefore, in a species exhibiting determinate growth of the primary root. To confirm the existence of the ppr-miR156-*PprSPL* module, and to rectify data on the possible existence of the ppr-miR156-*PprMATE* module, experimental analysis should be performed in the future.

# **1. Introducción**

## **1.1 Generalidades de la raíz**

La raíz es un órgano indispensable para la supervivencia y el desarrollo de las traqueofitas, o plantas vasculares. Además de mantener a la planta anclada al suelo, este órgano absorbe agua y nutrientes que se transportan a través del xilema hacia los demás órganos de la planta.

Se conoce como sistema radical al conjunto de raíces de una planta, y éste puede estar conformado por una raíz primaria, raíces laterales y raíces adventicias. La raíz primaria es aquella que se forma a partir de la raíz embrionaria o radícula; generalmente esta raíz es la primera que emerge de la semilla. Las raíces laterales son las que se originan a partir de una raíz parental. Las raíces adventicias, a diferencia de las raíces laterales, se desarrollan a partir de cualquier órgano distinto de la raíz, tales como los tallos o las hojas (Raven *et al.*, 2013).

El crecimiento y desarrollo de la raíz depende de la división, elongación y diferenciación celular. En el caso del crecimiento longitudinal de la raíz, estos procesos ocurren en regiones o zonas específicas del ápice, tal y como se describen a continuación.

### **1.1.1 Zonas de desarrollo de la raíz**

La raíz primaria, al igual que otras raíces de una planta, puede dividirse longitudinalmente en tres zonas: la zona meristemática, o el meristemo apical de la raíz (RAM, por sus siglas en inglés), en donde las células son pequeñas y mitóticamente activas; la zona de elongación, en donde las células ya no se dividen, pero se alargan longitudinalmente; y la zona de diferenciación, en donde las células completamente elongadas adquieren características de tipos celulares especializados. La división de las células en la zona meristemática y el alargamiento de las células en la zona de elongación permiten el crecimiento longitudinal de la raíz, por lo que la zona meristemática y la zona de elongación conforman a la región de crecimiento. Durante el proceso de diferenciación, las

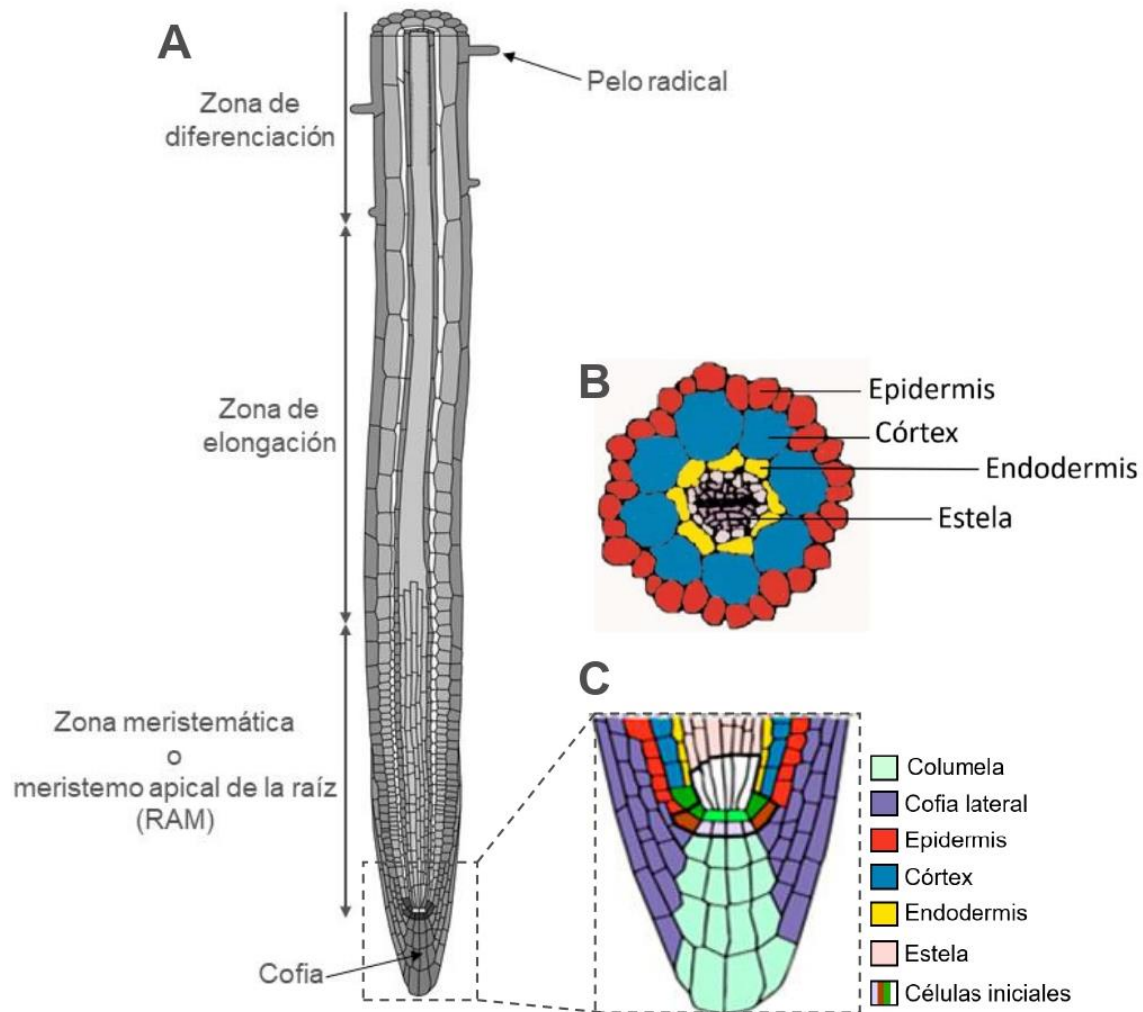
células epidérmicas forman extensiones laterales, llamados pelos radicales, los cuales aumentan la superficie de absorción de agua y nutrientes (Figura 1A), lo que permite fácilmente visualizar esta zona. La parte más apical de la zona meristemática está rodeada por una estructura llamada cofia. Las células de la cofia protegen al RAM contra el daño físico mediante la secreción de mucopolisacáridos cuando la raíz crece a través del suelo. Asimismo, la cofia juega un papel importante en las respuestas de la raíz a la gravedad, gradientes de humedad y obstáculos físicos, entre otras, llamadas tropismos (Raven *et al.*, 2013).

En un corte transversal de la parte distal de una raíz se pueden apreciar, de afuera hacia adentro, las capas concéntricas de células de epidermis, córtex y endodermis. En una raíz joven de la especie modelo *Arabidopsis thaliana* (*Arabidopsis*) cada uno de estos tipos celulares está representado por una sola capa (Figura 1B). En otras especies, como las cactáceas, generalmente hay más de una capa de córtex; la endodermis representa la capa más interna del córtex. En la región central de la raíz se encuentra la estela, o estele, también llamada cilindro central, o cilindro (pro)vascular, que contiene al xilema y al floema rodeados por la capa del periciclo (Figura 1B).

### **1.1.2 Meristemo apical de la raíz**

En la parte distal del RAM se encuentra el centro quiescente (CQ). El CQ está compuesto por un grupo de células con actividad mitótica casi nula. Las células adyacentes al CQ se llaman células iniciales, o células troncales, y se dividen con poca frecuencia. Una célula hija producto de esta división se quedará como célula inicial, y la otra, después de pasar por ciclos de división y luego por elongación celular, dará lugar a las células de los diferentes tipos celulares. En la raíz de *Arabidopsis* hay cuatro tipos de células iniciales (Figura 1C): la célula inicial del córtex y la endodermis dará origen a las capas del córtex y la endodermis, es decir, el tejido fundamental. El tejido vascular, que incluye el periciclo, floema y xilema, se formará a partir de células iniciales de la vasculatura. La célula inicial de la epidermis (tejido dermal) y la cofia lateral dará lugar a estos dos tipos celulares.

La parte central de la cofia, la columela, se originará de su propio tipo de células iniciales. Se ha demostrado que el contacto entre las células de CQ y las células iniciales mantiene a estas últimas en un estado indiferenciado (Scheres *et al.*, 2002). En algunos trabajos contemporáneos se llama nicho de células troncales al conjunto del CQ y células iniciales; en otros textos se considera solamente al CQ.



**Figura 4.** Organización de la raíz primaria de *Arabidopsis thaliana*. (A) Corte longitudinal de la raíz. Se señalan las tres zonas de desarrollo, así como la cofia y los pelos radicales. (B) Corte transversal de la raíz. Se observan las diferentes capas celulares. Imagen tomada de Scheres *et al.*, 2002. (C) El meristemo apical de la raíz (RAM, por sus siglas en inglés) contiene al centro quiescente y a las células iniciales que formarán cada tipo celular de la raíz. (A) y (C) son imágenes tomadas y modificadas de Peret, 2017.

## 1.2 Crecimiento indeterminado y determinado de la raíz

Generalmente, la raíz puede crecer de manera indefinida, o indeterminada, debido a la existencia y actividad del RAM por periodos largos de tiempo durante la vida de este órgano. Sin embargo, en algunos casos, el RAM se agota y todas las células del ápice se diferencian. Como resultado de este proceso, la raíz deja de crecer. Este tipo de crecimiento de la raíz se llama crecimiento determinado. Se conocen varios ejemplos de crecimiento determinado de las raíces laterales y adventicias (revisado en Shishkova *et al.*, 2008). Sin embargo, hasta el año 2022, las únicas especies reportadas que presentan crecimiento determinado de la raíz primaria de manera natural, y no en plantas mutantes, pertenecen a la familia Cactaceae (Dubrovsky, 1997; Shishkova *et al.*, 2013). Recientemente en el repositorio bioRxiv se depositó un preprint que reporta crecimiento determinado de la raíz primaria en plántulas de *Striga hermonthica* obtenidas de las semillas colectadas en la naturaleza (Xiao *et al.*, 2022).

## 1.3 Cactáceas

Las plantas de la familia Cactaceae, o cactáceas, son nativas de América. Se distribuyen principalmente en zonas áridas o semiáridas desde el sur de Canadá hasta el sur de Argentina (Anderson, 2001). Existe una prevalencia del crecimiento determinado de la raíz primaria en especies de cactáceas de las subfamilias Cactoideae y Opuntioideae, dos de las cuatro subfamilias de cactáceas (Shishkova *et al.*, 2013). Se ha propuesto que este tipo de crecimiento les confiere una ventaja adaptativa para su supervivencia debido a que permite un rápido establecimiento de la plántula en ambientes áridos y semiáridos (Dubrovsky, 1997). Sin embargo, la regulación genética del agotamiento del RAM como causa del crecimiento determinado no se conoce.

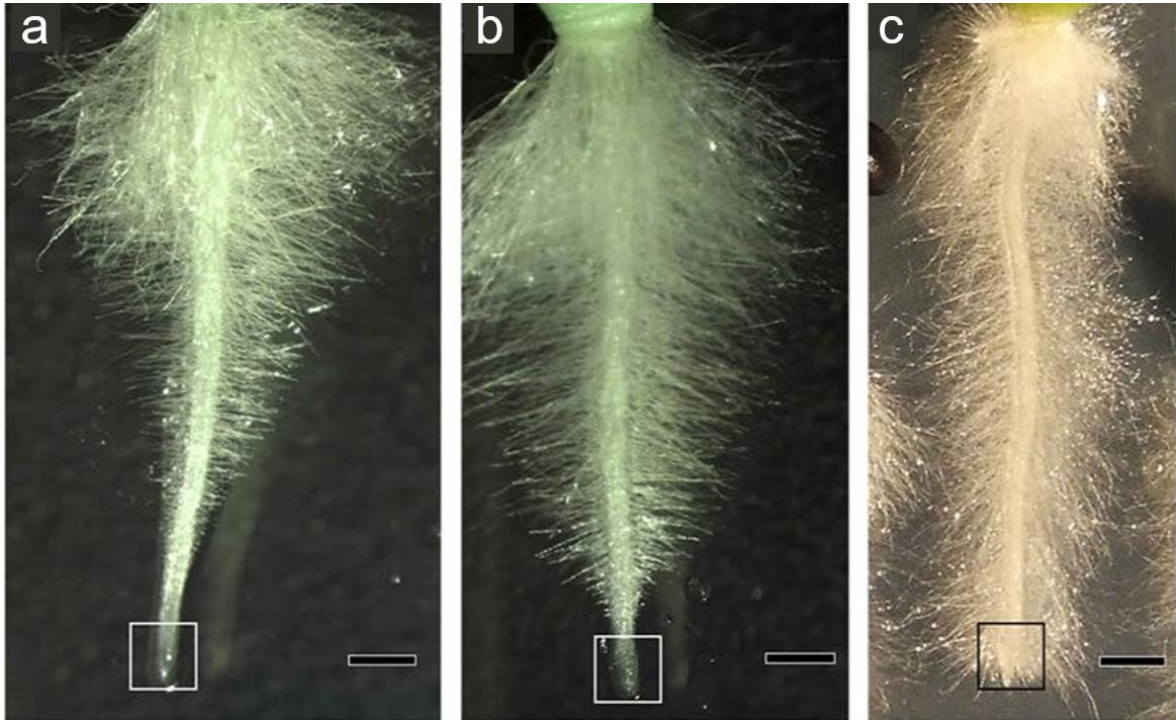
### 1.3.1 *Pachycereus pringlei*

Para estudiar los mecanismos moleculares del agotamiento del RAM, en el laboratorio se ha utilizado a la cactácea *Pachycereus pringlei*, perteneciente a la subfamilia Cactoideae. Esta especie es endémica del Desierto de Sonora; es una



planta arborescente columnar que puede llegar a medir hasta 20 metros de altura (Turner *et al.*, 1995). La raíz primaria de esta cactácea deja de crecer a los 7-10 días después de la germinación (d.d.g.) como resultado del agotamiento del RAM (Dubrovsky y Gómez-Lomelí, 2003). Para el estudio del crecimiento determinado, el crecimiento de la raíz primaria de *P. pringlei* puede dividirse en tres etapas: la etapa inicial, de 1-4 d.d.g. aproximadamente, en la que el RAM está presente y activo; la etapa intermedia, de 5-6 d.d.g., en la que el RAM está presente, pero es más pequeño; y la etapa terminal, de 9-13 d.d.g., en la que el RAM se ha agotado y, en consecuencia, todas las células del ápice de la raíz están diferenciadas (Figura 2). La distancia entre la zona de diferenciación y el ápice de la raíz primaria nos ayuda a diferenciar entre estas tres etapas; a su vez, podemos identificar a la zona de diferenciación por la presencia de pelos radicales en esta zona. Cabe mencionar que en plántulas diferentes el tiempo entre la germinación y cese del crecimiento de la raíz puede variar considerablemente, sin embargo, todas las plántulas exhiben el crecimiento determinado de la raíz primaria. En otras especies de cactáceas con este tipo de crecimiento de la raíz, como en *Stenocereus gummosus*, la raíz primaria puede crecer durante 2 a 3 días (Dubrovsky, 1997) o incluso por un período más corto (Shishkova *et al.*, 2013).

Debido a que el crecimiento determinado de la raíz primaria de *P. pringlei* es un programa genético altamente conservado (Dubrovsky y Gómez-Lomelí, 2003; Shishkova *et al.*, 2013) y a que el tiempo que tarda la raíz en dejar de crecer nos permite distinguir con facilidad cada una de las etapas del desarrollo para la colecta de muestras, esta especie es un buen modelo para el estudio del agotamiento del RAM. En el trabajo previo del laboratorio se caracterizó el transcriptoma del ápice de la raíz primaria de *P. pringlei* (Rodríguez-Alonso *et al.*, 2018) y posteriormente se empezó a estudiar la regulación de la expresión génica mediada por microRNAs (López-Valle, Tesis de doctorado en preparación).



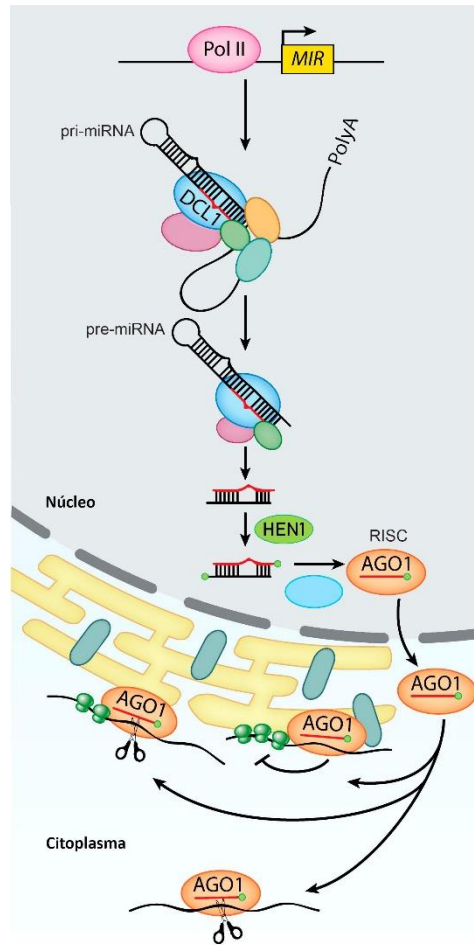
**Figura 5.** El crecimiento determinado de la raíz primaria de *Pachycereus pringlei* se puede dividir en tres etapas del desarrollo. (a) Etapa inicial, cuando el RAM está presente y activo. (b) Etapa intermedia, cuando el RAM está presente, pero es más pequeño. (c) Etapa terminal, cuando el RAM se ha agotado. La posición de los pelos radicales respecto al ápice nos ayuda a diferenciar entre las tres etapas del desarrollo. Imagen tomada de Rodríguez-Alonso *et al.*, 2018. Barra = 1 mm.

## 1.4 Regulación postranscripcional de la expresión génica mediada por microRNAs

Los microRNAs (miRNAs) son moléculas de RNA generalmente de 21-22 nucleótidos (nts) de longitud (Yu *et al.*, 2019), las cuales se originan de los genes *MIR*. En las plantas estos genes pueden estar localizados entre otros genes, es decir, en regiones intergénicas, o dentro de intrones e incluso en exones de otros genes (Wang *et al.*, 2019). Cuando la RNA polimerasa II transcribe a un gen *MIR* se produce un transcrito denominado miRNA primario (pri-miRNA). Una parte del pri-miRNA de plantas, modificado por adición del “cap” y la cola de poliadeninas, adquiere una estructura secundaria de tallo-asa; es decir, esta parte de la molécula de RNA de una sola cadena se pliega formando una doble cadena con una homología imperfecta (Figura 3). Posteriormente el pri-miRNA se procesa en dos pasos en los que en plantas participa principalmente una endonucleasa de tipo DICER, generalmente la enzima DICER-LIKE 1 (DCL1, de aquí en adelante se utiliza la nomenclatura para *Arabidopsis*). En plantas, a diferencia de los animales, ambos pasos se llevan a cabo en el núcleo y por la misma enzima (Figura 3). El resultado del primer paso es el precursor del miRNA (pre-miRNA), y el resultado del segundo paso es un dúplex miRNA-miRNA\*. Los extremos 3' de cada cadena del dúplex son metilados por una enzima metiltransferasa, generalmente HUA ENHANCER 1 (HEN1). La cadena miRNA del dúplex miRNA-miRNA\* se separa y se une a una proteína llamada ARGONAUTA (AGO). La unión del miRNA funcional con AGO forma al complejo de silenciamiento inducido por RNA (RISC por sus siglas en inglés). Durante los primeros años del estudio de miRNAs se pensaba que la cadena miRNA\* se degradaba; posteriormente los resultados de secuenciación de RNAs pequeños (smallRNA-seq) evidenciaron que en muchas ocasiones no es así. Además, se han acumulado más evidencias sobre las funciones de cada una de las cadenas del dúplex (revisado en Liu *et al.*, 2017), por lo que la base de datos de miRNAs miRbase en su versión 22 (<https://www.mirbase.org>) ya no utiliza la terminología “miRNA/miRNA\*”. En esta base de datos actualmente se le denomina miRNA-5p a la cadena del dúplex que

se originó a partir del brazo 5' del pre-miRNA, y miRNA-3p a la cadena del dúplex que se originó a partir del brazo 3' del pre-miRNA.

El complejo RISC es transportado al citoplasma, donde puede unirse a un mRNA blanco específico por complementariedad de bases entre dicho transcrito y el miRNA que forma parte del complejo RISC. Esta complementariedad requerida es más alta en plantas que en animales bilaterales, por lo que los miRNAs de plantas generalmente tienen pocos blancos, y los de animales pueden tener muchos. RISC puede interferir en la expresión de los genes-blancos por dos mecanismos: (i) afectando la estabilidad del transcrito-blanco mediante el corte del transcrito debido a la actividad de endonucleasa de una proteína AGO, generalmente AGO1, lo cual conduce a la degradación del transcrito por exonucleasas; o (ii) inhibiendo su traducción, actuando en varias etapas de la interacción entre el transcrito y el ribosoma (Gautam *et al.*, 2017; Yu *et al.*, 2017; Yu *et al.*, 2019). En las plantas, el primer mecanismo es el más común. Este tipo de regulación de la expresión génica es uno de los tipos de silenciamiento génico postranscripcional y juega un papel importante en la regulación de la expresión génica (Figura 3). En plantas existen familias de miRNAs conservados en muchos o varios linajes, así como miRNAs linaje-específicos (Wang *et al.*, 2019).



**Figura 6.** Biogénesis de microRNAs y silenciamiento génico postranscripcional. La RNA polimerasa II transcribe al gen *MIR*. El transcrito miRNA primario (pri-miRNA) se procesa en dos pasos por una endonucleasa DICER-LIKE (DCL1 en la imagen) a pre-miRNA y después a dúplex miRNA-miRNA\*. El dúplex se metila por HEN1 (círculos verdes). Posteriormente las cadenas del dúplex se separan y una de ellas se une a la proteína ARGONAUTA (AGO1 en la imagen). La unión del miRNA con la proteína AGO forma el complejo de silenciamiento inducido por RNA (RISC, por sus siglas en inglés), RISC se transporta del núcleo al citoplasma, se une a un mRNA blanco por complementariedad de bases entre el miRNA y dicho transcrito, e interfiere con su traducción, ya sea por el mecanismo de corte del transcrito o por el mecanismo de inhibición de la traducción. Imagen tomada y modificada de Yu *et al.*, 2019.

## **2. Antecedentes**

El crecimiento determinado de la raíz primaria de las cactáceas es un programa genético robusto (Dubrovsky *et al.*, 2007; Dubrovsky y Gómez-Lomelí, 2003; Shishkova *et al.*, 2013). Sin embargo, se desconocen los mecanismos moleculares que conducen al agotamiento del RAM. Con el objetivo de dilucidar la regulación genética del crecimiento determinado, en el laboratorio se secuenció, se ensambló *de novo* y se caracterizó el transcriptoma del ápice de la raíz primaria de *P. pringlei* de las tres etapas del desarrollo mencionadas arriba (Rodríguez-Alonso *et al.*, 2018). También, mediante la secuenciación masiva “smallRNA-seq” y “degradome-seq” y su análisis se ha obtenido información sobre el microtranscriptoma y el degradoma de la raíz de *P. pringlei*. Entre otros resultados, se encontró que algunos miRNAs de *P. pringlei* y sus transcritos blancos tienen correlación inversa de abundancia en el ápice de la raíz primaria de este cactus en las etapas del desarrollo, cuando tiene el RAM o cuando el RAM está agotado (López-Valle, Tesis de doctorado en preparación). Hasta el momento no existen datos del análisis funcional de los miRNAs y sus blancos en cactáceas; para empezar este tipo de análisis se recomienda establecer las condiciones experimentales, empleando un módulo miRNA-blanco conservado.

Algunos de los miRNAs altamente conservados en plantas son los miRNAs que pertenecen a la familia miR156 (Gautam *et al.*, 2017). Estos miRNAs regulan a los genes de la familia *SQUAMOSA PROMOTER BINDING PROTEIN LIKE (SPL)*, los cuales codifican para factores de transcripción. El módulo miR156–*SPL* es un módulo muy conservado en plantas que participa en procesos de transición del desarrollo de las hojas, transición a floración, en el desarrollo de raíces laterales y en el desarrollo de la raíz primaria de *Arabidopsis* (Poethig, 2009; Yu *et al.*, 2015; Barrera-Rojas *et al.*, 2020). En nuestro laboratorio, y como resultado de un análisis del degradoma del ápice de la raíz primaria de *P. pringlei*, además de los blancos conservados en otras especies vegetales de la familia *SPL*, se propuso un blanco nuevo de miR156: un transcrito de la familia *MULTIDRUG AND TOXIC COMPOUND EXTRUSION-TYPE TRANSPORTER (MATE)* (López-Valle, Tesis de

doctorado en preparación). La familia *MATE* es un grupo importante de transportadores de compuestos orgánicos en plantas. Además, se reportó que la sobreexpresión de dos proteínas *MATE* de arroz acelera la transición del desarrollo de las hojas en plantas transgénicas de *Arabidopsis* (Tiwari *et al.*, 2014).

En este trabajo se identificaron los posibles ortólogos de los genes *SPL* de *Arabidopsis* que se expresan en el ápice de la raíz primaria de *P. pringlei* y se analizó si dichos transcritos *SPL* de *P. pringlei* (*PprSPLs*), así como un transcrito *MATE* de *P. pringlei* (*PprMATE*), son posibles blancos de diferentes isoformas de ppr-miR156 (miR156 de *P. pringlei*) identificadas en el trabajo previo del laboratorio.

### **3. Justificación**

El módulo miR156—*SPL* es idóneo para iniciar con el análisis de la regulación de la expresión génica mediada por miRNAs en el crecimiento determinado de la raíz primaria de *P. pringlei* porque es un módulo que está conservado desde briofitas hasta plantas angiospermas (Cho *et al.*, 2012; Wang *et al.*, 2019). Además, existe un trabajo cuyos resultados sugieren que este módulo participa en el desarrollo de la raíz primaria de *Arabidopsis* (Barrera-Rojas *et al.*, 2020).

En el proyecto de doctorado de Mayra L. López Valle se sugirió un nuevo blanco de ppr-miR156, *PprMATE*. Si bien los transcritos *SPL* de *Arabidopsis* son blancos de miR156, los transcritos *MATE* de *Arabidopsis* no son blancos de estos miRNA. Por lo mencionado, fue importante examinar la posibilidad del reclutamiento de un nuevo blanco que, al parecer, participa en el desarrollo de plantas, en adición a los blancos conservados que participan en estos procesos.

En este proyecto de tesis, con el uso de los datos del transcriptoma y microtranscriptoma del ápice de la raíz primaria de *P. pringlei* se realizó un análisis bioinformático de los posibles módulos de regulación postranscripcional ppr-miR156—*PprSPL* y ppr-miR156—*PprMATE*. Los resultados de este trabajo permiten que en el futuro se empiece con el análisis experimental para confirmar

la funcionalidad de estos módulos en el ápice de la raíz primaria de *P. pringlei* durante el agotamiento del RAM.

## **4. Objetivos**

### **4.1 Objetivo general**

- Realizar el análisis bioinformático de los posibles módulos de regulación postranscripcional ppr-miR156–*PprSPL* y ppr-miR156–*PprMATE* en la raíz primaria de *Pachycereus pringlei*.

### **4.2 Objetivos particulares**

- Identificar y anotar a los transcritos *SPL* de *Pachycereus pringlei* que se expresan en el ápice de la raíz primaria.
- Extender los transcritos incompletos *PprSPLs*, utilizando secuencias genómicas de *Pachycereus pringlei* y *Carnegieia gigantea*.
- Caracterizar los posibles sitios de unión de los ppr-miR156 en las secuencias de los transcritos *SPL* y del transcrito *MATE* de *Pachycereus pringlei*.



## **5. Materiales y métodos**

### **5.1 Identificación de los transcritos *SPL* de *Pachycereus pringlei***

Las secuencias de las proteínas *SPL* de *Arabidopsis* se obtuvieron de la base de datos de TAIR (<https://www.arabidopsis.org/>). Se utilizó la herramienta tBLASTn del programa CLC Genomics Workbench (v9.5.4) utilizando como sonda a la proteína *SPL3* de *Arabidopsis* para explorar las secuencias de la base de datos del transcriptoma del ápice de la raíz primaria de *P. pringlei* ensamblado *de novo* en el trabajo previo del laboratorio (Rodríguez-Alonso *et al.*, 2018). Los contigs se consideraron transcritos *SPL* cuando el e-value fue menor o igual a  $1 \times 10^{-5}$ . También se empleó el programa BLASTx de NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) para explorar las diferentes secuencias de las proteínas *SPL* de la base de datos 'Non-redundant protein sequences (nr)'. Se escogieron los cinco mejores hits de este BLAST para analizar el sentido del marco de lectura de los contigs y verificar si su marco de lectura estaba completo o incompleto. Para obtener la secuencia reversa complementaria de nts se utilizó el programa Reverse Complement ([https://www.bioinformatics.org/sms/rev\\_comp.html](https://www.bioinformatics.org/sms/rev_comp.html)) y para traducir las secuencias de nts se usó la herramienta Translate de ExPASy (<https://web.expasy.org/translate/>). Las secuencias transcriptómicas de los ortólogos probables *SPL* de *P. pringlei* se utilizaron como sonda para asignar sus ortólogos en *Arabidopsis* por mejor hit unidireccional al utilizar el algoritmo BLASTn de la base de datos Phytozome.

### **5.2 Extensión del marco de lectura de los transcritos incompletos**

Con base en los alineamientos de los cinco mejores hits del BLASTx realizado en NCBI se identificó si el marco de lectura de los transcritos *SPL* de *P. pringlei* estaba completo o incompleto. Las secuencias de nts de los ortólogos probables *SPL* de *P. pringlei* con un marco de lectura incompleto se mapearon sobre el genoma preliminar de baja cobertura de *P. pringlei* (Copetti *et al.*, 2017) para intentar extender el marco de lectura. Los transcritos cuyo marco de lectura no se

pudo extender sobre el genoma de *P. pringlei*, se mapearon sobre el genoma del cactus *Carnegiea gigantea* (Copetti *et al.*, 2017). Primero, algunas de las secuencias genómicas identificadas se sometieron al programa Reverse Complement para tener el marco de lectura de los genes SPL codificados en la cadena '+'. Después, las secuencias genómicas se alinearon con las secuencias de sus transcritos correspondientes. Estos alineamientos de pares de secuencias se llevaron a cabo con la herramienta EMBOSS Needle disponible en [https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/). Con base en estos alineamientos, se seleccionó una parte de la secuencia genómica (para cada caso) que correspondía a la región donde sí se alineó con la secuencia del transcrito incompleto, más una región de aproximadamente 10,000 nts más allá de la parte faltante de los transcritos incompletos, cuando el tamaño de la secuencia genómica fue lo suficientemente grande, para utilizarla como sonda con el algoritmo BLASTx de NCBI utilizando la base de datos 'Non-redundant protein sequences (nr)'. Con base en los alineamientos de los cinco mejores hits de esta búsqueda se identificó a la región codificante de cada una de las secuencias genómicas.

Las regiones 5' y 3' UTR (regiones no traducidas, o Un-Translated Regions, por sus siglas en inglés), así como la CDS (región codificante, o CoDing Sequence, por sus siglas en inglés), se anotaron con el programa CLC Genomics Workbench (v9.5.4).

También se realizó un alineamiento múltiple de secuencias de las proteínas SPL de *Arabidopsis*, *P. pringlei* y *C. gigantea* con el programa Clustal Omega (Clustal-O; <https://www.ebi.ac.uk/Tools/msa/clustalo/>). Este alineamiento se utilizó para construir un árbol filogenético no enraizado con el software MEGA X (Kumar *et al.*, 2018) utilizando el algoritmo Maximum Likelihood basado en el modelo de sustitución Jones-Taylor-Thornton (JTT). Se probó la robustez estadística del árbol mediante el método de Bootstrap con 1,000 réplicas, así como un modelo de distribución gamma y sitios invariantes (G+I). Se usó delección parcial de gaps en las columnas con <65% de cobertura. Este árbol fue exportado como un archivo

NWK y fue visualizado y editado con el programa FigTree (v1.4.4) que puede descargarse en <https://github.com/rambaut/figtree/releases>.

### **5.3 Predicción de transcritos blancos de ppr-miR156**

Para el análisis de los posibles sitios de unión de las diferentes isoformas de ppr-miR156 en las secuencias de los transcritos *SPL* de *P. pringlei* y *C. gigantea*, y el transcrito *PprMATE*, se utilizó el programa RNAhybrid (v2.2) (Rehmsmeier *et al.*, 2004) (<https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid/submission.html/>).

Además, se realizó un alineamiento múltiple de los posibles sitios de unión de miRNAs en los probables transcritos-blancos de *P. pringlei* identificados y sus ortólogos de *C. gigantea*. Este alineamiento se leyó como matriz, asignando valores numéricos a cada nucleótido, y a partir de esta matriz se hizo un heat-map utilizando el paquete pheatmap y la librería ggplot2 de R (<https://cran.r-project.org>).

## **6. Resultados y discusión**

### **6.1 Identificación de los transcritos *SPL* de *Pachycereus pringlei***

Para identificar a los transcritos *SPL* que se expresan en el ápice de la raíz primaria de *P. pringlei* durante las tres etapas del desarrollo del crecimiento determinado, se empleó el programa tBLASTn del software QIAGEN CLC Genomics Workbench (v9.5.4) y el transcriptoma ensamblado *de novo* (Rodríguez-Alonso *et al.*, 2018). Como sonda utilizamos a la secuencia de aminoácidos de la proteína SPL3 de *Arabidopsis* (AthSPL3). Nosotros escogimos como sonda a AthSPL3 porque en un trabajo previo del laboratorio un contig de *P. pringlei* se anotó como posible ortólogo de SPL3 de *Arabidopsis*. Esta búsqueda arrojó como resultado a 15 contigs, en adición al contig 0027386 anotado previamente. Tres de estos hits se descartaron debido a que tuvieron un e-value mayor a 1, lo cual indica que se pudieron obtener por azar. Los contigs seleccionados como ortólogos probables se muestran en la tabla 1.

En total identificamos 13 contigs, o transcritos, *SPL* probables de *P. pringlei* (*PprSPLs*). Este resultado es congruente con la variación reportada para diferentes especies del número de genes de la familia *SPL*. Por ejemplo, en *Arabidopsis* se han identificado 15 genes *SPL* (Cardon *et al.*, 1999; Xing *et al.*, 2010), en *Hordeum vulgare* L. 17 (Tripathi *et al.*, 2018), en *Oryza sativa* 19. Un caso particular es *Malus domestica*, cuya familia génica *SPL* que incluye 33 genes es la más grande descrita hasta el momento (Wang *et al.*, 2019).

**Tabla 1.** Contigs del transcriptoma de *P. pringlei* identificados por el tBLASTn con la sonda AthSPL3.

Hit	Tamaño (nts)	E-value	Bit-score
contig 0001120	3,569	1.62 x10 <sup>-32</sup>	288.00
contig 0000626	4,104	4.29 x10 <sup>-32</sup>	286.00
contig 0012348	1,326	1.66 x10 <sup>-30</sup>	272.00
contig 0007929	1,757	3.83 x10 <sup>-30</sup>	271.00
contig 0030438	604	5.09 x10 <sup>-30</sup>	262.00
contig 0000505	4,344	5.18 x10 <sup>-30</sup>	271.00
contig 0007106	1,860	6.39 x10 <sup>-30</sup>	269.00
contig 0024104	740	2.43 x10 <sup>-29</sup>	259.00
contig 0014860	1,146	3.15 x10 <sup>-27</sup>	249.00
contig 0004230	2,304	1.50 x10 <sup>-26</sup>	246.00
contig 0001824	3,094	1.99 x10 <sup>-21</sup>	211.00
contig 0016662	1,034	9.32 x10 <sup>-6</sup>	98.00

Las secuencias transcriptómicas *PprSPLs* se utilizaron como sonda con el algoritmo BLASTn en Phytozome para encontrar sus posibles ortólogos en Arabidopsis. Con base en el mejor hit de cada sonda se sugirió la anotación para cada uno de los transcritos *PprSPLs*, con el marco de lectura tanto completo, como incompleto. Cuando para dos o más transcritos *PprSPLs* se obtuvo como el mejor hit el mismo *SPL* de Arabidopsis, se añadió una letra en orden alfabético, de acuerdo con su nivel de abundancia en el transcriptoma del ápice de la raíz primaria (Rodríguez-Alonso *et al.*, 2018), siendo el más abundante el que lleva la letra a, por ejemplo: *PprSPL2a* (Tabla 2).

**Tabla 2.** Anotación de los transcritos *PprSPLs* basada en mejor hit unidireccional utilizando BLASTn con el genoma de Arabidopsis.

Contig	Nombre sugerido
contig 0000626	<i>PprSPL1a</i>
contig 0001120	<i>PprSPL1b</i>
contig 0007929	<i>PprSPL2a</i>
contig 00272386	<i>PprSPL2b</i>
contig 0012348	<i>PprSPL2c</i>
contig 0004230	<i>PprSPL6a</i>
contig 0014860	<i>PprSPL6b</i>
contig 0001824	<i>PprSPL7</i>
contig 0030438	<i>PprSPL13a</i>
contig 0007106	<i>PprSPL13b</i>
contig 0000505	<i>PprSPL14</i>
contig 0016662	<i>PprSPL15a</i>
contig 0024104	<i>PprSPL15b</i>

Los 13 transcritos *PprSPLs* se utilizaron como sonda para encontrar sus secuencias ortólogas en la base de datos 'Non-redundant protein sequences (nr)' con el programa BLASTx de NCBI. Estos resultados permitieron identificar si el marco de lectura de los transcritos estaba completo o incompleto, y también identificamos en qué dirección estaban ensamblados los contigs, es decir, si las proteínas SPL estaban codificadas en la cadena '+' o en la cadena '-'. Siete de los 13 transcritos (*PprSPL1a*, *PprSPL1b*, *PprSPL2a*, *PprSPL6a*, *PprSPL7*, *PprSPL13b* y *PprSPL14*) tuvieron un marco de lectura completo, mientras que los seis restantes (*PprSPL2b*, *PprSPL2c*, *PprSPL6b*, *PprSPL13a*, *PprSPL15a* y *PprSPL15b*) lo tuvieron incompleto. Estos seis transcritos incompletos se extendieron para poder identificar los posibles sitios de unión a miRNAs como se explica en el apartado 6.2. Sólo los transcritos *PprSPL2b* y *PprSPL15b* tuvieron su marco abierto de lectura en la cadena '+', mientras que el marco abierto de lectura de los otros 11 *PprSPLs* se halló en la cadena complementaria. Por esta razón, estos 11 transcritos se sometieron al programa Reverse Complement para

reportarlos con el marco de lectura en dirección 5' a 3' que es como se reporta normalmente a los transcritos en las bases de datos.

Por otra parte, también nos interesaba analizar a un transcrito de la familia *MATE* (*MULTIDRUG AND TOXIC COMPOUND EXTRUSION-TYPE TRANSPORTER*), *PprMATE*, el cual se propuso como un blanco nuevo de ppr-miR156 (López-Valle, Tesis de doctorado en preparación). Por esta razón analizamos si el marco de lectura del transcrito *PprMATE* que se propuso como blanco de miR156 estaba completo o incompleto. Para esto, *PprMATE* se sometió al programa BLASTx de NCBI utilizando la base de datos 'Non-redundant protein sequences (nr)'. Con base en los resultados de esta búsqueda se identificó que el marco de lectura estaba completo y que estaba en dirección 5' a 3', por lo que no fue necesario obtener su secuencia reversa complementaria.

## **6.2 Extensión del marco de lectura de los transcritos incompletos**

Para extender el marco de lectura de los seis transcritos *PprSPLs* incompletos identificamos las secuencias genómicas de las cuales se originan estos transcritos para reconocer las partes faltantes en cada uno de ellos. Para esto tomamos a los seis transcritos incompletos como sonda con el programa BLASTn del software CLC Genomics Workbench (v9.5.4) utilizando el genoma preliminar de baja cobertura de *P. pringlei* ensamblado en el trabajo de Copetti *et al.*, 2017. Como resultado de esta búsqueda obtuvimos seis scaffolds, o secuencias genómicas, correspondientes a cada uno de los transcritos incompletos (Tabla 3). Para cada una de estas secuencias genómicas se obtuvo su secuencia reversa complementaria. Después alineamos por pares a cada una de las secuencias genómicas (las seis secuencias originales y las seis secuencias reversas complementarias) con sus respectivos transcritos. El mejor alineamiento resultante en cada caso correspondía a la secuencia genómica con un marco de lectura en dirección 5' a 3'. Con base en estos alineamientos identificamos la región de las secuencias genómicas que se alineó con los transcritos (dicha región corresponde a una parte de la región codificante, o CDS, CoDing Sequence, por sus siglas en inglés), planeando recortar la secuencia de los scaffolds dejando

aproximadamente 10,000 nts río arriba o río abajo, si el transcrito estaba incompleto en las regiones 5' o 3', respectivamente, y del otro lado recortar hasta donde se terminaba la secuencia del UTR presente en el transcrito. Sin embargo, solo uno de los seis scaffolds identificados tuvo el tamaño mayor de 10,000 nts, y otros cinco fueron más cortos, entre 1,776 y 5,204 nts (Tabla 3). Por lo anterior, estos cinco transcritos no se recortaron.

**Tabla 3.** Hits (scaffolds de *P. pringlei*) de la búsqueda por BLASTn utilizando a los seis transcritos incompletos *PprSPLs* como sonda.

Query (contig)	Hit (scaffold)	Tamaño (nts)	E-value
<b><i>PprSPL2b</i></b>	NCQS01028287.1	1,776	0.0
<b><i>PprSPL2c</i></b>	NCQS01025644.1	2,917	0.0
<b><i>PprSPL6b</i></b>	NCQS01119680.1	4,987	0.0
<b><i>PprSPL13a</i></b>	NCQS01078155.1	5,204	0.0
<b><i>PprSPL15a</i></b>	NCQS01156571.1	11,307	0.0
<b><i>PprSPL15b</i></b>	NCQS01010913.1	1,928	0.0

Posteriormente, las seis secuencias genómicas de la tabla 3 se utilizaron como sonda con el programa BLASTx de NCBI usando la base de datos 'Non-redundant protein sequences (nr)' para identificar su marco abierto de lectura. Muchos genes de organismos eucarióticos, entre ellos las plantas, contienen intrones. Por esta razón los transcritos maduros son producto del procesamiento de pre-mRNAs que consiste en la eliminación de los intrones, llamado empalme, o splicing, entre otras etapas de procesamiento. Entonces, consideramos que las regiones de las sondas que sí se alinearon con los hits anotados como *SPL* correspondían a exones en cada secuencia genómica, y las regiones que no se alinearon con los hits correspondían a intrones (aunque también cabía la posibilidad de que las regiones donde no hubo alineamiento correspondieran a artefactos por errores del ensamblado del genoma, puesto que es un genoma de baja cobertura). Como se pronosticó por el tamaño de los scaffolds (Tabla 3), utilizando el genoma de *P. pringlei* sólo pudimos extender el marco de lectura de uno de los seis transcritos incompletos, *PprSPL13a* (Figura 4). Esto se debió a que el genoma de *P. pringlei*



se encuentra muy fragmentado: está ensamblado en 171,584 scaffolds con una media de longitud de solo 2,347pb.

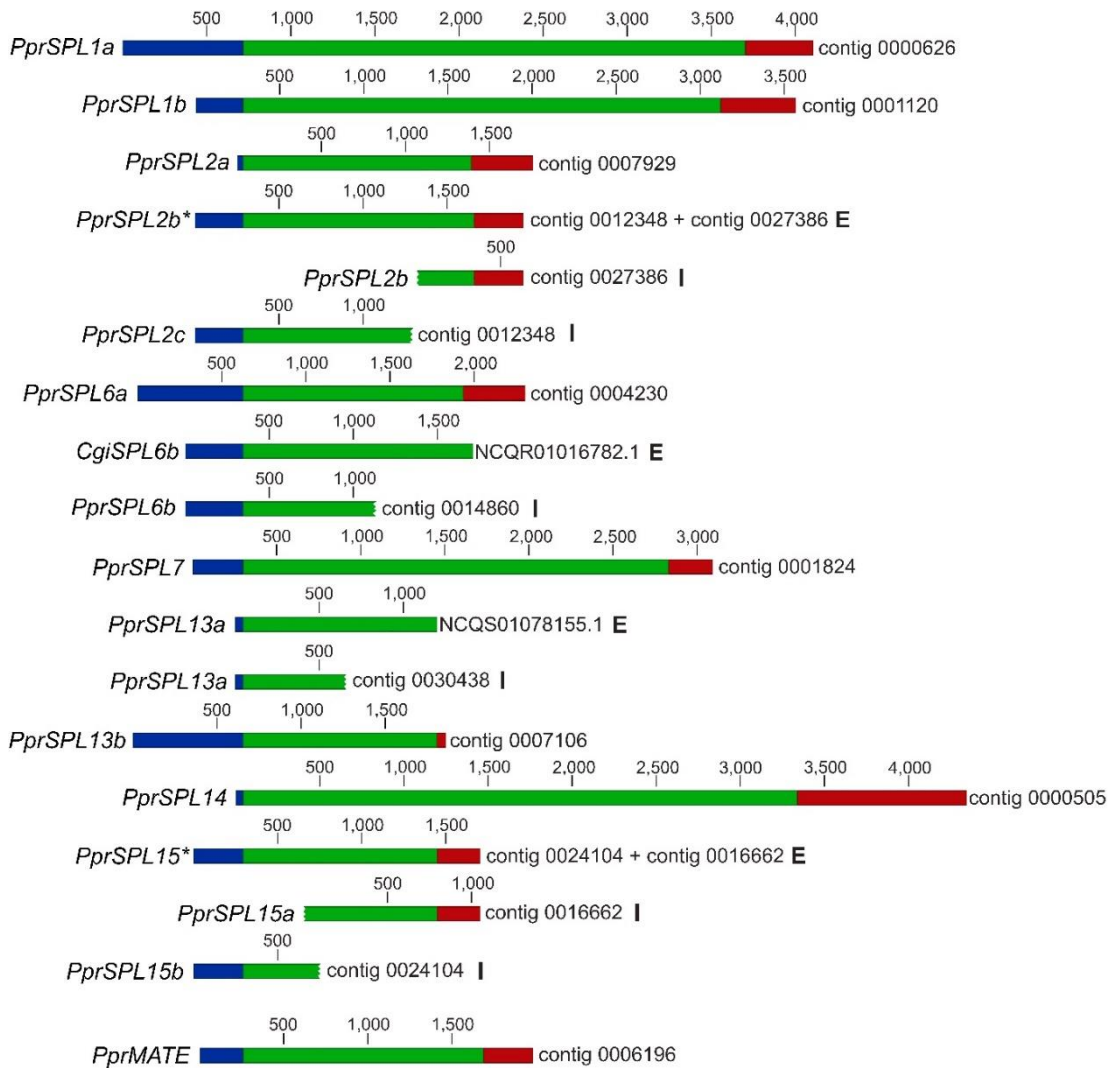
Como no pudimos extender cinco de los seis transcritos incompletos con el genoma de *P. pringlei*, decidimos repetir el mismo análisis utilizando un genoma de otro cactus columnar, *Carnegiea gigantea* (Copetti *et al.*, 2017). Entre los genomas de cactáceas reportados, escogimos a este cactus porque su genoma en aquel momento fue el mejor ensamblado de los cinco genomas disponibles de cactáceas. Además, es la especie de cactus más cercana filogenéticamente a *P. pringlei*; ambos pertenecen a la tribu Pachycereeae de la subfamilia Cactoideae. Como se esperaba, y como se aprecia en la tabla 4, el tamaño de los scaffolds del genoma de *C. gigantea*, es considerablemente mayor que el de los scaffolds del genoma de *P. pringlei*. Por esta razón, fue posible identificar a la secuencia completa del ortólogo *SPL6b* de *C. gigantea* (*CgiSPL6b*) al utilizar a la secuencia incompleta del transcrito *PprSPL6b*. El transcrito *PprSPL6b* no se extendió porque su ortólogo *CgiSPL6b* contiene únicamente nucleótidos de la secuencia genómica de *C. gigantea*, pero a partir de aquí nos referiremos a *CgiSPL6b* como un transcrito extendido debido a que el e-value entre las secuencias de *P. pringlei* y *C. gigantea* es de 0.0 (Tabla 4), por lo cual suponemos que el transcrito *PprSPL6b* tiene la misma secuencia, o una con muy pocas sustituciones, que *CgiSPL6b*.

Para los transcritos incompletos *PprSPL2b* y *PprSPL2c* se obtuvo un mismo scaffold como hit (Tabla 4). Estos dos contigs mapearon sobre las partes 3' y 5', respectivamente, de la CDS del gen ortólogo, el cual se identificó como *CgiSPL2b*. Esto indica que ambos contigs (*PprSPL2b* y *PprSPL2c*) conforman un único gen. Utilizando la secuencia deducida a partir del transcrito *CgiSPL2b* y uniendo las secuencias de *PprSPL2b* y *PprSPL2c*, se infirió la CDS completa del transcrito que fue nombrado como *PprSPL2b\** (Figura 4). Lo mismo ocurrió para los contigs *PprSPL15a* y *PprSPL15b*: ambos mapearon sobre un mismo scaffold de *C. gigantea* en las partes 3' y 5' de la CDS, respectivamente. Al fusionar a estos dos contigs en un solo transcrito se obtuvo la CDS completa; a este transcrito fusionado lo nombramos *PprSPL15\** (Figura 4).

**Tabla 4.** Resultados significativos de la búsqueda por BLASTn al utilizar a los seis transcritos incompletos *PprSPLs* como sonda para interrogar al genoma de *C. gigantea*.

Query (contig)	Hit (Scaffold)	Tamaño (nts)	E-value
<b><i>PprSPL2b</i></b>	NCQR01000926.1	101,689	0.0
<b><i>PprSPL2c</i></b>	NCQR01000926.1	101,689	0.0
<b><i>PprSPL6b</i></b>	NCQR01016782.1	104,471	0.0
<b><i>PprSPL13a</i></b>	NCQR01053264.1	30,535	0.0
<b><i>PprSPL15a</i></b>	NCQR01009966.1	242,072	0.0
<b><i>PprSPL15b</i></b>	NCQR01009966.1	242,072	0.0

En la figura 4 se muestran los datos obtenidos en los análisis descritos arriba: (1) las regiones 5' UTR (región que no se traduce, o UnTranslated Region, por sus siglas en inglés), CDS y 3' UTR de los siete transcritos *PprSPLs* completos: *PprSPL1a*, *PprSPL1b*, *PprSPL2a*, *PprSPL6a*, *PprSPL7*, *PprSPL13b*, *PprSPL14* y el transcrito *PprMATE*; (2) las regiones CDS incompletas y las regiones 5' o 3' UTR (según corresponda) de los seis transcritos incompletos *PprSPLs* (marcados con una **I**) acompañados por (3) las CDS completas inferidas a partir de las secuencias genómicas para los genes correspondientes a los transcritos incompletos (*PprSPL2b\**, *CgiSPL6b*, *PprSPL13a* y *PprSPL15\**) (marcados con una **E**) que también incluyen la región 5' o 3' UTR presente en los transcritos incompletos. Debido a que la región UTR faltante no se puede identificar por métodos bioinformáticos, esta no se muestra en la figura. Las regiones 5' UTR, CDS y 3' UTR fueron delimitadas con el programa CLC Genomics Workbench (v9.5.4). En la figura 4, por la longitud de las CDSs se puede apreciar que las diferentes proteínas de cactáceas de la familia SPL tienen tamaños bastante diferentes, similar a lo reportado para proteínas SPL de otras especies vegetales (Zhang *et al.*, 2016; Zhong *et al.*, 2019; Zhu *et al.*, 2020).



**Figura 4.** Marco de lectura de los transcritos *PprSPLs*, *CgiSPL6b* y *PprMATE*. Del lado izquierdo se muestran los nombres y del lado derecho el ID para cada transcritor. Todos los transcritos incompletos (señalados con una I) lograron extenderse y están señalados con una E. El transcritor extendido *PprSPL2b\** es la unión de los transcritos incompletos *PprSPL2b* y *PprSPL2c*, mientras que el transcritor extendido *PprSPL15\** es la unión de los transcritos incompletos *PprSPL15a* y *PprSPL15b*, ambos transcritos extendidos se infirieron a partir de secuencias genómicas de *C. gigantea*. La secuencia codificante, o CDS (CoDing Sequence, por sus siglas en inglés) del transcritor extendido *PprSPL13a* se recuperó completa con el genoma de *P. pringlei* y la CDS del transcritor extendido *CgiSPL6b* sólo se pudo recuperar con el genoma de *C. gigantea*, utilizando como referencia la secuencia del transcritor incompleto *PprSPL6b* y la base de datos de 'Non-redundant protein sequences (nr)' de NCBI. Los números arriba de las secuencias representan el número de nts. El color azul representa a la región 5' UTR (Un-Translated Region, por sus siglas en inglés), el color verde representa la CDS, y el color rojo representa a la región 3' UTR.

También se identificaron los ortólogos *SPL* de *C. gigantea* correspondientes a los nueve transcritos *PprSPLs* con CDS completo, incluidos *PprSPL2b\** y *PprSPL15\**, realizando el mismo análisis explicado en los párrafos anteriores. Todos los *SPL* de *C. gigantea* se nombraron de acuerdo con la anotación de sus ortólogos en *P. pringlei*. En la tabla 5 se muestran los porcentajes de identidad y similitud entre las proteínas *PprSPL* y *CgiSPL*. Como se puede observar, seis pares de proteínas ortólogas *SPL* tienen el mismo tamaño, la longitud de tres pares de ortólogos se difiere en un residuo (aa, aminoácido) y de otro par en 2 aa. Los porcentajes de identidad de los pares de proteínas ortólogas varían entre 100% y 97.1%. Esta observación indica que las secuencias de los ortólogos *SPL* de *P. pringlei* y *C. gigantea* están altamente conservadas y que, por lo tanto, podemos suponer que la secuencia de la proteína *CgiSPL6b*, identificada a partir del transcrito incompleto *PprSPL6b*, será casi idéntica a las de su ortólogo *PprSPL6b*.

**Tabla 5.** Porcentajes de identidad y similitud (calculada con la matriz BLOSUM62) entre las proteínas SPL de *Pachycereus pringlei* y *Carnegieia gigantea*.

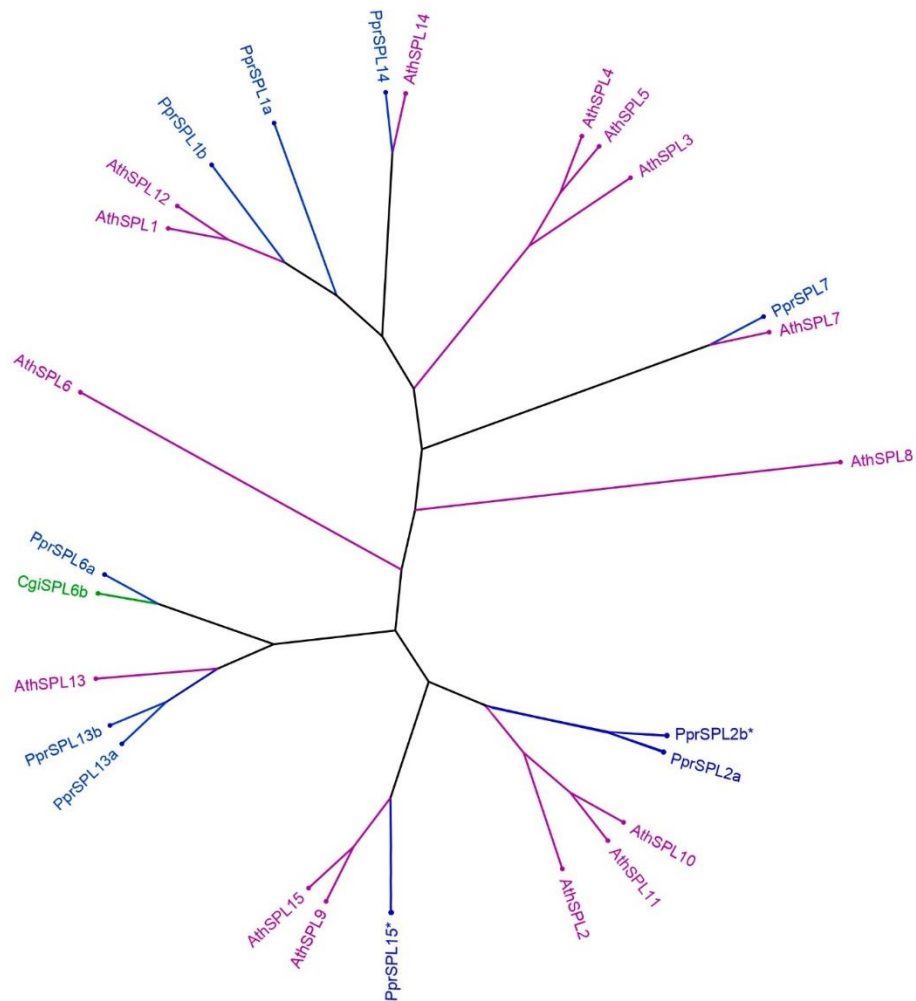
Par de ortólogos	Tamaño (aa)		Identidad (%)	Similitud (%)
	PprSPL	CgiSPL		
PprSPL1a vs CgiSPL1a	PprSPL1a	CgiSPL1a	99.1	99.5
	995	995		
PprSPL1b vs CgiSPL1b	PprSPL1b	CgiSPL1b	98.6	99.3
	945	945		
PprSPL2a vs CgiSPL2a	PprSPL2a	CgiSPL2a	98.2	98.7
	451	453		
PprSPL2b* vs CgiSPL2b	PprSPL2b*	CgiSPL2b	98.7	98.9
	458	457		
PprSPL6a vs CgiSPL6a	PprSPL6a	CgiSPL6a	97.7	97.9
	435	435		
PprSPL7 vs CgiSPL7	PprSPL7	CgiSPL7	100	100
	769	769		
PprSPL13a vs CgiSPL13a	PprSPL13a	CgiSPL13a	98.4	99.2
	385	384		
PprSPL13b vs CgiSPL13b	PprSPL13b	CgiSPL13b	97.1	98.2
	383	383		
PprSPL14 vs CgiSPL14	PprSPL14	CgiSPL14	97.3	98.3
	1,097	1,097		
PprSPL15* vs CgiSPL15	PprSPL15*	CgiSPL15	98.2	98.4
	383	384		

### 6.3 Árbol filogenético de las proteínas SPL

Para confirmar la anotación de los transcritos *SPL* de *P. pringlei*, basada en el mejor hit unidireccional obtenido con el algoritmo BLASTn utilizando el genoma de *Arabidopsis*, se realizó un alineamiento múltiple con las secuencias de las proteínas PprSPL1a, PprSPL1b, PprSPL2a, PprSPL2b\*, PprSPL6a, CgiSPL6b, PprSPL7, PprSPL13a, PprSPL13b, PprSPL14 y PprSPL15\*, y las 15 proteínas SPL de *Arabidopsis*. Las proteínas ortólogas SPL de *P. pringlei* y *C. gigantea* son idénticas o tienen un porcentaje de similitud >98%, por lo que decidimos utilizar al ortólogo CgiSPL6b en lugar de PprSPL6b, porque no contamos con la secuencia completa de esta proteína. Posteriormente se construyó un árbol filogenético no enraizado con este alineamiento, al utilizar el algoritmo Maximum-Likelihood, con 1,000 réplicas de Bootstrap bajo el modelo de sustitución Jones-Taylor-Thornton (JTT). El árbol fue exportado como un archivo NWK y fue editado con el programa FigTree v1.4.4 (Figura 5).

El resultado de este análisis muestra que las proteínas PprSPL7 y PprSPL14 son ortólogos probables de las proteínas AthSPL7 y AthSPL14, respectivamente, puesto que PprSPL7 se agrupó en un clado con AthSPL7 y PprSPL14 se agrupó en un clado con AthSPL14. Las proteínas PprSPL13a y PprSPL13b se agruparon en un clado con AthSPL13, lo cual indica que dos parálogos PprSPL son ortólogos de una sola proteína AthSPL. Las proteínas provisionalmente anotadas como PprSPL1a y PprSPL1b se agruparon en un clado con las proteínas AthSPL1 y AthSPL12; las proteínas PprSPL2a y PprSPL2b\* se agruparon con las proteínas AthSPL2, AthSPL10 y AthSPL1; y la proteína PprSPL15\* se agrupó con las proteínas AthSPL9 y AthSPL15, por lo que la relación de paralogía/ortología para estas proteínas es más compleja. Las proteínas PprSPL6a y CgiSPL6b no se agruparon con AthSPL6, sino que se agruparon como un clado hermano del clado que contiene a PprSPL13a, PprSPL13b y AthSPL13. Cabe señalar que hubo clados de proteínas SPL de *Arabidopsis* en los que no se agruparon las proteínas de *P. pringlei* o *C. gigantea* e.g., AthSPL3-AthSPL4-AthSPL5. Un análisis filogenético más detallado que incluya proteínas SPL de más especies de distintas

familias vegetales podrá rectificar y precisar la anotación preliminar de las PprSPLs/CgiSPLs realizada en este trabajo. Además, se debe tener en cuenta que los *PprSPLs* se identificaron en el transcriptoma del ápice de la raíz primaria ensamblado *de novo* y que existen más genes *PprSPLs* que no se expresan en el ápice de la raíz.



**Figura 5.** Árbol filogenético de las proteínas SPL de *Arabidopsis*, *P. pringlei* y *C. gigantea*. En lugar de la proteína PprSPL6b, para la cual no se logró recuperar su secuencia completa, se utilizó la proteína CgiSPL6b, aprovechando que las proteínas SPL ortólogas de estas dos especies de cactáceas son idénticas o tienen un alto porcentaje de identidad (Tabla 4). Este árbol filogenético se obtuvo con el método de Maximum-Likelihood con 1,000 réplicas de Bootstrap bajo el modelo JTT y fue editado con el programa FigTree (v1.4.4). En color morado se muestran las proteínas SPL de *Arabidopsis* (AthSPLs), en color azul las proteínas SPL de *P. pringlei* (PprSPLs) y en color verde la proteína SPL6b de *C. gigantea* (CgiSPL6b).

#### 6.4 Predicción de posibles transcritos blancos de ppr-miR156

Se utilizaron cinco isoformas de ppr-miR156 identificadas en un trabajo previo del laboratorio (López-Valle, Tesis de doctorado en preparación) para predecir sus posibles sitios de unión en los transcritos que aparecen en la figura 4. Estas cinco isoformas son ppr-miR156a, ppr-miR156b, ppr-miR156c, ppr-miR156d y ppr-miR156e; sus nombres fueron asignados con base en su abundancia durante la etapa inicial y la etapa terminal del desarrollo de la raíz primaria de *P. pringlei*, siendo ppr-miR156a la isoforma más abundante y ppr-miR156e la isoforma menos abundante.

La presencia de posibles sitios de unión de las cinco isoformas ppr-miR156 en los transcritos *PprSPL*, *CgiSPL6b* y *PprMATE*, se analizó con el programa RNAhybrid (<https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid/submission.html/>). Con base en los alineamientos miRNA–transcrito resultantes, se consideró como posibles transcritos-blancos a : (i) los transcritos cuyo sitio de unión no tuviera más de tres nts desapareados en la región semilla del miRNA (nts que ocupan las posiciones 2 al 13 en dirección 5' a 3' del miRNA) pues la complementariedad de bases en esta región juega un papel importante para llevar a cabo el corte del transcrito por el complejo RISC (Yu *et al.*, 2019); (ii) los transcritos cuyo valor de energía libre mínima (MFE: minimum free energy, por sus siglas en inglés) del alineamiento fuera menor a -30.0 kcal/mol y (iii) los transcritos cuyos alineamientos del sitio de unión y el miRNA no tuvieran indeles, que son inserciones o deleciones de nts en una secuencia. Se identificaron seis transcritos *PprSPLs* (*PprSPL2a*, *PprSPL2b\**, *PprSPL6a*, *PprSPL13a*, *PprSPL13b* y *PprSPL15\**) y al transcrito *CgiSPL6b*, con la suposición de que el gen *CgiSPL6b* se transcribe, como posibles blancos de las cinco isoformas ppr-miR156 (Tablas 6-10).



**Tabla 6.** Alineamientos del sitio de unión de ppr-miR156a en los transcritos *PprSPLs* y *CgiSPL6b*.

<b>ppr-miR156a—SPL</b>		
<b>Alineamiento</b>	<b>Transcrito</b>	<b>MFE (kcal/mol)</b>
5' GUGCUCUCUCUCUCUCUGUCA 3'      —        3' CACGAGUGAGAGAAGACAGU 5'	<i>PprSPL2a</i>	-38.2
	<i>PprSPL2b*</i>	
	<i>PprSPL6a</i>	
	<i>CgiSPL6b</i>	-37.6
	<i>PprSPL13a</i>	-38.0
	<i>PprSPL13b</i>	-37.9
	<i>PprSPL15*</i>	

La secuencia anotada en dirección 5'→3' corresponde al sitio de unión en el transcrito (20 nts), y la secuencia anotada en dirección 3'→5' corresponde a ppr-miR156a (20 nts). El símbolo “|” indica apareamiento de bases y el desapareamiento está indicado por “-”. MFE: valor mínimo de energía libre. Las diferencias del valor MFE se deben a las secuencias nucleotídicas en la vecindad del probable sitio de unión de miRNA.

**Tabla 7.** Alineamientos del sitio de unión de ppr-miR156b en los transcritos *PprSPLs* y *CgiSPL6b*.

<b>ppr-miR156b—SPL</b>		
<b>Alineamiento</b>	<b>Transcrito</b>	<b>MFE (kcal/mol)</b>
5' GUGCUCUCUCUCUCUCUGUCA 3'      —       — 3' CACGAGAGAUAGAAGACAGUC 5'	<i>PprSPL2a</i>	-37.0
	<i>PprSPL2b*</i>	
	<i>PprSPL6a</i>	-36.7
	<i>CgiSPL6b</i>	
5' GUGCUCUCUCUCUCUCUGUCAG 3'      —        3' CACGAGAGAUAGAAGACAGUC 5'	<i>PprSPL13a</i>	-38.6
	<i>PprSPL13b</i>	
5' GUGCUCUCUCUCUCUCUGUCAC 3'      —       — 3' CACGAGAGAUAGAAGACAGUC 5'	<i>PprSPL15*</i>	-36.4

La secuencia anotada en dirección 5'→3' corresponde al sitio de unión en el transcrito (21 nts), y la secuencia anotada en dirección 3'→5' corresponde a ppr-miR156b (21 nts). El símbolo “|” indica apareamiento de bases y el desapareamiento está indicado por “-”. MFE: valor mínimo de energía libre. Las diferencias del valor MFE se deben a las secuencias nucleotídicas en la vecindad del probable sitio de unión de miRNA.

**Tabla 8.** Alineamientos del sitio de unión de ppr-miR156c en los transcritos *PprSPLs* y *CgiSPL6b*.

<b>ppr-miR156c–SPL</b>		
<b>Alineamiento</b>	<b>Transcrito</b>	<b>MFE (kcal/mol)</b>
5' GUGCUCUCUCUCUCUUCUGUCA 3'      -----      3' CACGAGAGAUAGAAGACAGUU 5'	<i>PprSPL2a</i>	-38.8
	<i>PprSPL2b*</i>	
	<i>PprSPL6a</i>	
	<i>CgiSPL6b</i>	
5' GUGCUCUCUCUCUCUUCUGUCAG 3'      -----     : 3' CACGAGAGAUAGAAGACAGUU 5'	<i>PprSPL13a</i>	-38.5
	<i>PprSPL13b</i>	
5' GUGCUCUCUCUCUCUUCUGUCAC 3'      -----     ----- 3' CACGAGAGAUAGAAGACAGUU 5'	<i>PprSPL15*</i>	-37.8

La secuencia anotada en dirección 5'→3' corresponde al sitio de unión en el transcrito (21 nts), y la secuencia anotada en dirección 3'→5' corresponde a ppr-miR156c (21 nts). El símbolo “|” indica apareamiento de bases, el desapareamiento está indicado por “-” y el símbolo “:” indica un apareamiento guanina-uracilo. MFE: valor mínimo de energía libre. Las diferencias del valor MFE se deben a las secuencias nucleotídicas en la vecindad del probable sitio de unión de miRNA.

**Tabla 9.** Alineamientos del sitio de unión de ppr-miR156d en los transcritos *PprSPLs* y *CgiSPL6b*.

<b>ppr-miR156d–SPL</b>		
<b>Alineamiento</b>	<b>Transcrito</b>	<b>MFE (kcal/mol)</b>
5' UCCGUGCUCUCUCUCUUCUGU 3'   -     -----      3' AGUCACGAGAGAUAGAAGACA 5'	<i>PprSPL2a</i>	-33.4
	<i>PprSPL2b*</i>	
5' GUCGUGCUCUCUCUCUUCUGU 3' -:-     -----      3' AGUCACGAGAGAUAGAAGACA 5'	<i>PprSPL6a</i>	-32.9
	<i>CgiSPL6b</i>	
5' AUUGUGCUCUCUCUCUUCUGU 3' -:-     -----      3' AGUCACGAGAGAUAGAAGACA 5'	<i>PprSPL13a</i>	-32.6
5' AUCGUGCUCUCUCUCUUCUGU 3' -:-     -----      3' AGUCACGAGAGAUAGAAGACA 5'	<i>PprSPL13b</i>	-32.9
5' GCUGUGCUCUCUCUCUUCUGU 3' - -----     ----- 3' AGUCACGAGAGAUAGAAGACA 5'	<i>PprSPL15*</i>	-33.0

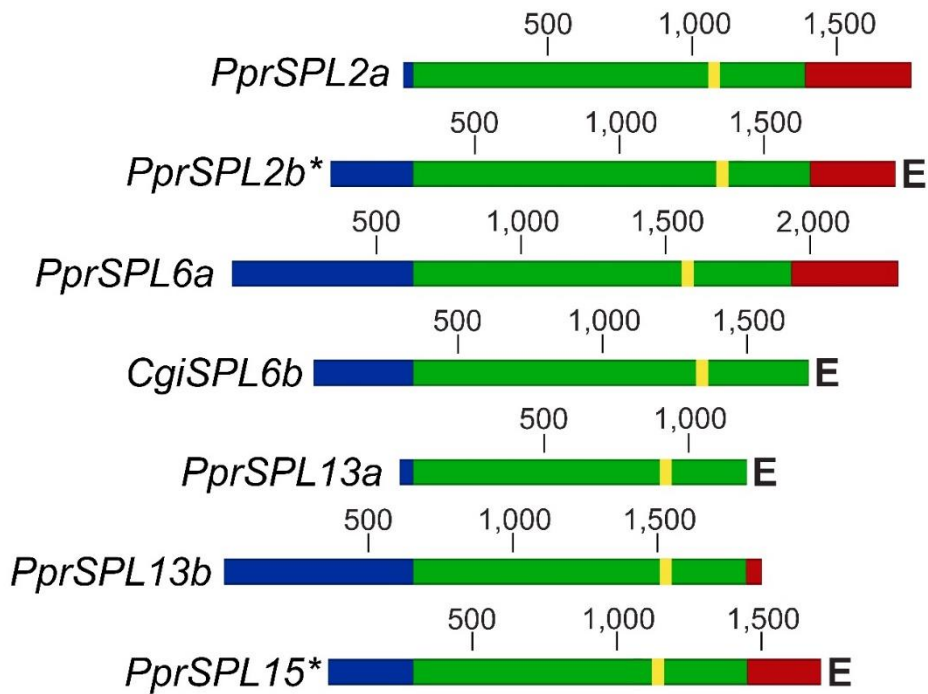
La secuencia anotada en dirección 5'→3' corresponde al sitio de unión en el transcrito (21 nts), y la secuencia anotada en dirección 3'→5' corresponde a ppr-miR156d (21 nts). El símbolo “|” indica apareamiento de bases, el desapareamiento está indicado por “-” y el símbolo “:” indica un apareamiento guanina-uracilo. MFE: valor mínimo de energía libre. Las diferencias del valor MFE se deben a las secuencias nucleotídicas en la vecindad del probable sitio de unión de miRNA.

**Tabla 10.** Alineamientos del sitio de unión de ppr-miR156e en los transcritos *PprSPLs* y *CgiSPL6b*.

<b>ppr-miR156e—SPL</b>		
<b>Alineamiento</b>	<b>Transcrito</b>	<b>MFE (kcal/mol)</b>
5' GUGCUCUCUCUCUUCUGUCAACA 3'      -----   ----- 3' CACGAGUGAGAGAAGACAGUAAA 5'	<i>PprSPL2a</i>	-39.1
	<i>PprSPL2b*</i>	
5' GUGCUCUCUCUCUUCUGUCAGCU 3'      -----   -----: 3' CACGAGUGAGAGAAGACAGUAAA 5'	<i>PprSPL6a</i>	-38.8
	<i>CgiSPL6b</i>	
5' GUGCUCUCUCUCUUCUGUCACCA 3'      -----   ----- 3' CACGAGUGAGAGAAGACAGUAAA 5'	<i>PprSPL13a</i>	-37.8
5' GUGCUCUCUCUCUUCUGUCAUCA 3'      -----   ----- 3' CACGAGUGAGAGAAGACAGUAAA 5'	<i>PprSPL13b</i>	-38.2
5' GUGCUCUCUCUCUUCUGUCAAU 3'      -----   ----- 3' CACGAGUGAGAGAAGACAGUAAA 5'	<i>PprSPL15*</i>	-39.1

La secuencia anotada en dirección 5'→3' corresponde al sitio de unión en el transcrito (23 nts), y la secuencia anotada en dirección 3'→5' corresponde a ppr-miR156e (23 nts). El símbolo "|" indica apareamiento de bases, el desapareamiento está indicado por "-" y el símbolo ":" indica un apareamiento guanina-uracilo. MFE: valor mínimo de energía libre. Las diferencias del valor MFE se deben a las secuencias nucleotídicas en la vecindad del probable sitio de unión de miRNA.

Asimismo, se identificó la posición de los posibles sitios de unión de ppr-miR156 en los siete posibles transcritos-blancos identificados. Como se puede apreciar en la figura 6, en todos los transcritos *PprSPLs* y *CgiSPL6b* estos posibles sitios de unión de ppr-miR156 se sitúan dentro de la CDS. Además, son sitios únicos, lo cual se muestra en las tablas 5 a 9. Estos datos confirman las diferencias descritas previamente sobre la regulación postranscripcional mediada por miRNAs en plantas y animales: los transcritos-blancos de animales bilaterales pueden tener varios sitios de unión a miRNAs con menor complementariedad de bases entre el miRNA y el sitio de unión, y se encuentran generalmente en el 3' UTR del transcrito-blanco; mientras que los transcritos-blancos de plantas generalmente tienen un único sitio de unión a miRNAs con complementariedad de bases casi perfecta entre el miRNA y el sitio de unión, y este sitio puede estar tanto en la CDS como en los UTRs, dependiendo de la familia génica (Yu *et al.*, 2019).



**Figura 6.** Posición del sitio de unión de ppr-miR156 en los seis posibles transcritos-blancos *PprSPLs* y el posible transcrito-blanco *CgiSPL6b*. Los transcritos extendidos están marcados con una **E**. El color azul representa a la región 5' UTR, el color verde representa la secuencia codificante, el color rojo representa a la región 3' UTR y el color amarillo representa el sitio de unión de ppr-miR156.

De acuerdo con los criterios usados para determinar posibles transcritos-blancos (Yu *et al.*, 2019), tomando en cuenta la complementariedad de bases entre el posible sitio de unión en el transcrito *PprMATE* y el umbral de la energía libre mínima, este transcrito no se identificó como probable blanco de ninguna de las cinco isoformas de ppr-miR156. En la tabla 11 se muestran los alineamientos ppr-miR156–*PprMATE* que arrojó el programa.

**Tabla 11.** Alineamientos del posible sitio de unión de las cinco isoformas de ppr-miR156 en *PprMATE*.

<b>ppr-miR156–<i>PprMATE</i></b>		
<b>Alineamiento</b>	<b>miRNA</b>	<b>MFE (kcal/mol)</b>
5' AUGCUCAUUCUCGUGCUUGUAA 3' -     :    -      -  3' CACGAGUGAGAGAA-G-ACAGU 5'	ppr-miR156a	-25.6
5' GUGCUCUCUCUGCC-UCAA 3'      :    -      -  3' CACGAGAGAUAGAAGACAGUC 5'	ppr-miR156b	-28.3
5' AUGCUCAUUCUCGUGCUUGUAAG 3' -     :    -      - : 3' CACGAGUGAGAGAA-G-ACAGUU 5'	ppr-miR156c	-25.6
5' UCGGUGCUCUCUGCCUCAAU 3'   :     :        --  3' AGUCACGAGAGAUAGA---AGACA 5'	ppr-miR156d	-33.0
5' AUGCUCAUUCUCGUGCUUGUAAGCA 3' -     :    -      - :-- 3' CACGAGUGAGAGAA-G-ACAGUUA 5'	ppr-miR156e	-25.6

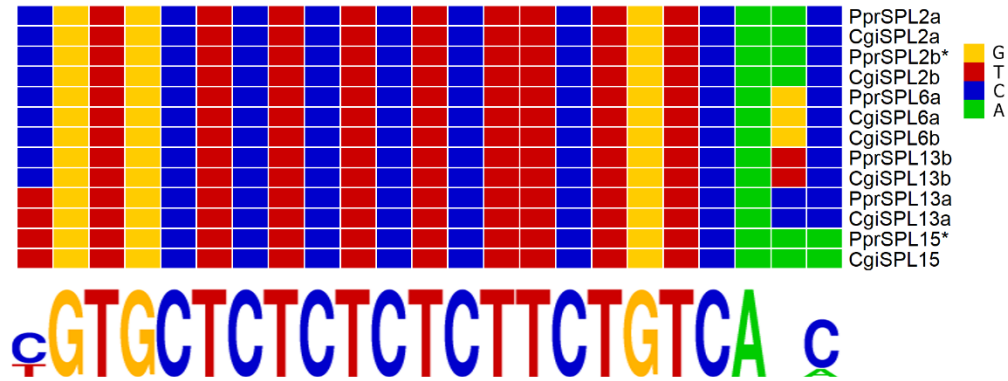
La secuencia anotada en dirección 5'→3' corresponde al sitio de unión en el transcrito, y la secuencia anotada en dirección 3'→5' corresponde al miRNA. El símbolo “|” indica apareamiento de bases, el despareamiento está indicado por “-”, el símbolo “:” indica un apareamiento guanina-uracilo y los indeles están representados por “-”. MFE: valor mínimo de energía libre. Las diferencias del valor MFE se deben a las secuencias nucleotídicas en la vecindad del probable sitio de unión de miRNA.

En la figura 7 se muestra un alineamiento entre los posibles sitios de unión de ppr-miR156 en los seis posibles transcritos-blancos *PprSPLs*, *CgiSPL6b* y *PprMATE*. Como se puede observar, los sitios de unión de todos los probables *SPLs*-blancos de ppr-miR156 tienen un alto porcentaje de identidad, lo que explica por qué se conserva su complementariedad con las cinco isoformas de ppr-miR156. Además, confirma las observaciones reportadas, las cuales sugieren que los polimorfismos de nts en los extremos 5' o 3' de los sitios de unión de miRNAs pueden contribuir a la regulación diferencial de distintos transcritos-blancos de una familia génica por la regulación postranscripcional mediada por diferentes isoformas de miRNAs (Pradhan *et al.*, 2020). Sin embargo, el probable sitio de unión de *PprMATE* tiene menor identidad con los sitios de unión de los *SPLs*, por lo que la complementariedad de bases entre *PprMATE* y las isoformas ppr-miR156 es mucho menor.

<i>PprSPL2a</i>	CGTGCTCTCTCTTTCTGTCAAC
<i>PprSPL2b*</i>	CGTGCTCTCTCTTTCTGTCAAC
<i>PprSPL6a</i>	CGTGCTCTCTCTTTCTGTCAAG
<i>CgiSPL6b</i>	CGTGCTCTCTCTTTCTGTCAAG
<i>PprSPL13a</i>	TGTGCTCTCTCTTTCTGTCAAC
<i>PprSPL13b</i>	CGTGCTCTCTCTTTCTGTCAAT
<i>PprSPL15*</i>	TGTGCTCTCTCTTTCTGTCAAA
<i>PprMATE</i>	GGTGCTCTCTGTCTGCCTCAATA

**Figura 7.** Conservación de las secuencias de los sitios de unión de los posibles transcritos-blancos *PprSPLs*, *CgiSPL6b* y el transcrito *PprMATE*. En el alineamiento se muestra la identidad con resaltado de color verde y la diferencia de las bases con resaltado de color rojo.

En total identificamos seis transcritos *SPL* de *P. pringlei* (*PprSPL2a*, *PprSPL2b\**, *PprSPL6a*, *PprSPL13a*, *PprSPL13b* y *PprSPL15\**), y un hipotético transcrito *SPL* de *C. gigantea* (*CgiSPL6b*), como posibles transcritos-blancos de ppr-miR156. Además, al analizar la presencia de los posibles sitios de unión de ppr-miR156 en los ortólogos identificados *CgiSPLs* (*CgiSPL2a*, *CgiSPL2b*, *CgiSPL6a*, *CgiSPL13a*, *CgiSPL13b* y *CgiSPL15*), se observó que sí están presentes e incluso son idénticos a los posibles sitios de unión en sus ortólogos *PprSPLs* (Figura 8). Este resultado sugiere que el posible sitio de unión en *CgiSPL6b* está conservado en su ortólogo *PprSPL6b* y que, por lo tanto, siete de los 11 transcritos *PprSPLs* identificados en este trabajo son posibles transcritos-blancos de ppr-miR156. Esto es congruente con lo observado en otras especies, en las cuales algunos genes *SPLs* son blancos de los miR156 (tienen sitios de unión a diferentes isoformas de la familia miR156 y están confirmadas en experimentos la actividad de RISC de corte de transcrito-blanco, y/o inhibición de la traducción); mientras que otros genes *SPLs* no son blancos de los miRNAs de esta familia. Por ejemplo, en *Arabidopsis* 10 de 15 *SPLs* son blancos de 8 isoformas de miR156, en *O. sativa* 10 de 19 *SPLs* son blancos de 11 isoformas de miR156 y en *M. domestica* 20 de 33 *SPLs* pueden ser blancos de 32 isoformas de miR156 (Wang *et al.*, 2019).



**Figura 8.** Conservación de la secuencia del sitio de unión de ppr-miR156 en los ortólogos *CgiSPLs*. Este *heatmap* fue hecho con el paquete *pheatmap* y la librería *ggplot2* de R. Abajo del *heatmap* se muestra la matriz de peso de los sitios de unión; como puede observarse, en la penúltima posición no aparece ningún nucleótido porque los cuatro nucleótidos (G, T, C y A) aparecen en esta posición en el alineamiento. El marco de lectura está en dirección 5'→3' y las secuencias están alineadas en pares de ortólogos *PprSPLs/CgiSPLs*, excepto *CgiSPL6b* puesto que no se cuenta con la secuencia completa de su ortólogo *PprSPL6b*.

Para confirmar que los transcritos *PprSPLs* mencionados arriba en efecto son blancos de las cinco isoformas de ppr-miR156, más la existencia o no del módulo ppr-miR156–*PprMATE*, es necesario hacer un análisis experimental. Para esto, por ejemplo, se puede utilizar el método de RLM-RACE (RNA Ligase-Mediated - Rapid Amplification of cDNA Ends), con el cual se puede comprobar la actividad de corte del transcrito después de la unión del complejo RISC por complementariedad de bases entre el sitio de unión en el transcrito y el miRNA. Por lo tanto, esta es una de las perspectivas.

## **7. Conclusiones**

- En el ápice de la raíz primaria de *P. pringlei* se expresan 11 transcritos *PprSPLs*, lo cual sugiere la importancia de esta familia génica en el desarrollo de la raíz de esta especie. El tamaño de la familia génica *PprSPL* concuerda con lo reportado para otras especies de angiospermas.
- En seis de los once transcritos *PprSPLs*, más uno sugerido por el análisis con el genoma de *C. gigantea*, se encontró la secuencia de unión de las cinco isoformas ppr-miR156, indicando que la regulación de la expresión de genes *SPL* por miR156 también existe en las especies de cactáceas con crecimiento determinado de la raíz primaria. Tanto *PprSPLs*, como isoformas de ppr-miR156, se expresan en el ápice de la raíz primaria de *P. pringlei*, lo cual apunta a la participación de la regulación de los *PprSPLs* por ppr-miR156 en el desarrollo de la raíz de esta especie.
- El probable sitio de unión del transcrito *PprMATE* tiene menor porcentaje de identidad con ppr-miR156 comparado con el sitio de unión de los transcritos *PprSPLs*, por lo que se necesita más evidencia experimental para confirmar o rechazar la existencia del módulo miR156–*MATE* sugerido en el análisis anterior en el grupo de investigación.



## **8. Perspectivas**

Los resultados del presente trabajo aportan información que permitirá la realización de experimentos que puedan confirmar la existencia, funcionalidad y papel de los módulos de regulación de expresión génica ppr-miR156–*PprSPL* y ppr-miR156–*PprMATE* en el ápice de la raíz primaria de *P. pringlei*. Entre los posibles experimentos a realizar destacan:

- Comprobar la actividad de ppr-miR156 sobre los transcritos seleccionados *PprSPLs*.
- Analizar la posible actividad de ppr-miR156 sobre el transcrito *PprMATE*.
- Analizar la abundancia de las isoformas de ppr-miR156 en el ápice de la raíz primaria de *P. pringlei* en dos etapas del desarrollo, con el meristemo apical de la raíz activo o agotado.
- Analizar la abundancia de los transcritos seleccionados *PprSPLs*, posibles blancos de ppr-miR156, en el ápice de la raíz primaria de *P. pringlei* en dos etapas del desarrollo.
- Analizar la abundancia de *PprMATE*, posible transcrito blanco de ppr-miR156, en el ápice de la raíz primaria de *P. pringlei* en dos etapas del desarrollo.

## **9. Bibliografía**

- Anderson, E. F. (2001). The cactus family. Timber Press, Portland, OR.
- Barrera-Rojas, C. H., Rocha, G. H. B., Polverari, L., Pinheiro Brito, D. A., Batista, D. S., Notini, M. M., ... & Nogueira, F. T. S. (2020). miR156-targeted SPL10 controls Arabidopsis root meristem activity and root-derived de novo shoot regeneration via cytokinin responses. *Journal of Experimental Botany*, **71**, 934-950.
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281-297.
- Cho, S. H., Coruh, C., & Axtell, M. J. (2012). miR156 and miR390 regulate tasiRNA accumulation and developmental timing in *Physcomitrella patens*. *The Plant Cell*, **24**, 4837-4849.
- Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J. L., Childs, K. L., Eguiarte, L. E., ... & Sanderson, M. J. (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences*, **114**, 12003-12008.
- Dubrovsky, J. G. (1997). Determinate primary-root growth in seedlings of Sonoran Desert Cactaceae; its organization, cellular basis, and ecological significance. *Planta*, **203**, 85-92.
- Dubrovsky, J. G., & Gómez-Lomelí, L. F. (2003). Water deficit accelerates determinate developmental program of the primary root and does not affect lateral root initiation in a Sonoran Desert cactus (*Pachycereus pringlei*, Cactaceae). *American Journal of Botany*, **90**, 823-831.
- Gautam, V., Singh, A., Verma, S., Kumar, A., Kumar, P., Singh, S., ... & Sarkar, A. K. (2017). Role of miRNAs in root development of model plant *Arabidopsis thaliana*. *Indian Journal of Plant Physiology*, **22**, 382-392.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, **35**, 1547.

- Pradhan, M., Pandey, P., Baldwin, I. T., & Pandey, S. P. (2020). Argonaute4 modulates resistance to *Fusarium brachygibbosum* infection by regulating jasmonic acid signaling. *Plant Physiology*, **184**, 1128-1152.
- Poethig, R. S. (2009). Small RNAs and developmental timing in plants. *Current Opinion in Genetics & Development*, **19**, 374-378.
- Raven, P. H., Evert, R. F. y Eichhorn, S. E. (2013). *Biology of plants*. New York: W. H. Freeman and Co.
- Rehmsmeier, M., Steffen, P., Höchsmann, M., & Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507-1517.
- Rodríguez-Alonso, G. (2018). Exploración mediante RNAseq de la regulación genética del crecimiento determinado de la raíz primaria de la cactácea *Pachycereus pringlei*. Tesis de doctorado. Universidad Nacional Autónoma de México.
- Rodríguez-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**, 1-11.
- Scheres, B., Benfey, P., & Dolan, L. (2002). Root development. *The Arabidopsis book/American Society of Plant Biologists*, **1**.
- Shishkova, S., Rost, T. L., & Dubrovsky, J. G. (2008). Determinate root growth and meristem maintenance in angiosperms. *Annals of Botany*, **101**, 319-340.
- Shishkova, S., Las Peñas, M. L., Napsucialy-Mendivil, S., Matvienko, M., Kozik, A., Montiel, J., ... & 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**, 239-252.
- Tiwari, M., Sharma, D., Singh, M., Tripathi, R. D., & Trivedi, P. K. (2014). Expression of OsMATE1 and OsMATE2 alters development, stress

responses and pathogen susceptibility in Arabidopsis. *Scientific Reports*, **4**, 1-12.

- Turner, R. M., Bowers, J. E., & Burgess, T. L. (2005). *Sonoran Desert Plants: An Ecological Atlas*. University of Arizona Press.
- Wang, C., Wang, Q., Zhu, X., Cui, M., Jia, H., Zhang, W., ... & Shen, W. (2019). Characterization on the conservation and diversification of miRNA156 gene family from lower to higher plant species based on phylogenetic analysis at the whole genomic level. *Functional & Integrative Genomics*, **19**, 933-952.
- Xiao, T. T., Kirschner, G. K., Kountche, B. A., Jamil, M., Maria, S., Lube, V., ... & Blilou, I. (2021). The PLETHORA/PIN-FORMED/AUXIN network mediates terminal prehaustorium formation in the parasitic plant *Striga hermonthica*. *bioRxiv*.
- Yu, N., Niu, Q. W., Ng, K. H., & Chua, N. H. (2015). The role of miR156/SPL modules in Arabidopsis lateral root development. *The Plant Journal*, **83**, 673-685.
- Yu, Y., Jia, T., & Chen, X. (2017). The 'how' and 'where' of plant microRNAs. *New Phytologist*, **216**, 1002-1017.
- Yu, Y., Zhang, Y., Chen, X., & Chen, Y. (2019). Plant noncoding RNAs: hidden players in development and stress responses. *Annual Review of Cell and Developmental Biology*, **35**, 407-431.
- Zhang, W., Bei, L. I., & Bin, Y. U. (2016). Genome-wide identification, phylogeny and expression analysis of the SBP-box gene family in maize (*Zea mays*). *Journal of Integrative Agriculture*, **15**, 29-41.
- Zhong, H., Kong, W., Gong, Z., Fang, X., Deng, X., Liu, C., & Li, Y. (2019). Evolutionary analyses reveal diverged patterns of SQUAMOSA promoter binding protein-like (SPL) gene family in *Oryza* genus. *Frontiers in Plant Science*, **10**, 565.
- Zhu, T., Liu, Y., Ma, L., Wang, X., Zhang, D., Han, Y., ... & Ma, L. (2020). Genome-wide identification, phylogeny and expression analysis of the SPL gene family in wheat. *BMC Plant Biology*, **20**, 1-14.