



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**Maestría en Ciencias Bioquímicas**

Análisis molecular de los genes *PpPLT* en el desarrollo  
de la raíz primaria del cardón *Pachycereus pringlei*

TESIS

QUE PARA OPTAR POR EL GRADO DE:  
Maestra en Ciencias

PRESENTA:  
MAYRA LILIANA LÓPEZ VALLE

TUTOR PRINCIPAL  
DRA. SVETLANA SHISHKOVA  
INSTITUTO DE BIOTECNOLOGÍA

MIEMBROS DEL COMITÉ TUTOR  
DRA. ELIZABETH CORDOBA MARTINEZ  
DR. LUIS CÁRDENAS TORRES

INSTITUTO DE BIOTECNOLOGÍA

CUERNAVACA, MOR. MAYO, 2015



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.



Esta tesis fue realizada en el Instituto de Biotecnología de la UNAM en el departamento de biología molecular de plantas en el laboratorio a cargo del Dr. Joseph Dubrovsky, bajo la asesoría de la Dra. Svetlana Shishkova. Este proyecto fue financiado por la DGAPA-PAPIIT UNAM a través de los proyectos IN204912 y IN207115 y por CONACYT a través del proyecto 240055. Mayra Liliana López Valle recibió una beca de maestría por parte del CONACYT (No. Registro Becario 275133) y por concepto de obtención de grado de maestría por parte de la DGAPA-PAPIIT UNAM a través de los proyectos IN204912 y IN207115.

## **Agradecimientos**

En primer lugar quiero agradecer a mi madre Alicia Valle, a mi padre Guillermo López y mi hermano León Yakov, ya que sin el apoyo incondicional que siempre me han dado no me hubiera sido posible realizar este proyecto. Dedico también esta tesis a mi abuela Micaela De Jesús Domínguez de quien aprendí que las metas se alcanzan trabajando.

Agradezco a la Dra. Svetlana Shishkova por brindarme la oportunidad y la confianza para retomar este proyecto y al Dr. Joseph Dubrovsky, que dirige nuestro grupo de trabajo y quien inició este proyecto.

Debo agradecer al M. en C. Gustavo Rodríguez quien además de ser amigo y compañero de estudios colaboró en el desarrollo de los análisis bioinformáticos plasmados en este trabajo. Agradezco al Dr. Jesús Montiel González de quien aprendí la técnica de RT-qPCR. Debo agradecer también a la I.B.I. Marcela Ramírez Yarza, Técnico Académico, quien junto con Sinhue Albarrán y Javier Yael Solís colaboraron con el cultivo del material biológico. Sinhue Albarrán colaboró además en parte bioinformática de este proyecto.

Por sus aportaciones durante la revisión de este escrito agradezco a la Dra. Elizabeth Córdoba Martínez, Dra. Rosana Sánchez López, Dra. Claudia Díaz Camino, Dra. Denhi Schnabel Peraza, Dra. Esperanza Martínez y al Dr. Enrique Merino Pérez.

Agradecemos a Selene Napsupcialy Mendivil, Técnico Académico de este laboratorio; al Dr. Francisco Molina Freaner quien donó las semillas utilizadas en este proyecto y al Q.B. José F. Martínez Rodríguez, técnico académico del Instituto de Ecología - Unidad Hermosillo, UNAM, quien recolectó dichas semillas.

A todos mis compañeros de laboratorio Jazmín, Héctor, Ramcés, Delia gracias por todas las risas que compartimos. Agradezco además a mis amigos y a todos los que me impulsaron durante estos años para continuar con mis estudios.

# Tabla de contenido

<b>Tabla de contenido</b> .....	v
<b>Lista de abreviaturas empleadas</b> .....	viii
<b>Lista de símbolos</b> .....	ix
<b>Resumen</b> .....	1
<b>Summary</b> .....	2
<b>1 Introducción</b> .....	3
1.1 Características del crecimiento post-embionario en plantas .....	3
1.2 Importancia y desarrollo del sistema radical en plantas .....	6
1.3 El crecimiento de la raíz en plantas .....	7
1.4 La función del meristemo apical de la raíz en plantas .....	8
1.5 Características del crecimiento determinado de la raíz en plantas .....	9
1.6 Importancia ecológica del crecimiento determinado de la raíz primaria de las cactáceas desérticas .....	11
1.7 Distribución del crecimiento determinado de la raíz primaria en la familia Cactaceae .....	15
Mutantes con crecimiento determinado de la raíz identificadas en <i>A. thaliana</i> ...	17
1.8 Identificación de los factores de transcripción <i>SHORT ROOT</i> y <i>SCARECROW</i> .....	17
1.9 El papel de los factores de transcripción <i>PLETHORA</i> en <i>A. thaliana</i> .....	19
1. 9.1 La función de los genes <i>PLT</i> en el RAM en <i>A. thaliana</i> .....	19
1.9.2 Gradientes de expresión de los genes <i>PLT</i> en el RAM de <i>A. thaliana</i> .....	21
1.9.3 Control de la expresión de <i>PLT</i> .....	23
1.9.4 Los genes <i>PLT</i> en el SAM .....	23
1.9.5 La participación de genes <i>PLT</i> en filotaxis y rizotaxis .....	24
1.9.6 Los genes blanco de <i>PLT</i> .....	25
<b>2. Antecedentes</b> .....	26
2.1 Generación de bibliotecas sustractivas de expresión génica .....	26
2.2 mRNA-seq: transcriptomas de puntas de raíces de <i>P. pringlei</i> .....	26
<b>3. Justificación</b> .....	29
<b>4. Hipótesis</b> .....	30
<b>5. Objetivo general</b> .....	30
5.1 Objetivos particulares .....	30

<b>6. Material y Métodos</b> .....	31
Glosario de definiciones bioinformáticas .....	31
6.1 Análisis filogenéticos preliminares de genes <i>PLT</i> .....	32
6.1.1 Identificación de <i>contigs</i> con similitud a <i>PLT</i> .....	32
6.1.2 Identificación de posibles proteínas ortólogas a los <i>contigs PpPLT</i> .....	32
6.2 Análisis filogenético de genes <i>AIL</i> .....	33
6.2.1 Búsqueda en bases de datos de proteínas AIL .....	33
6.2.2 Alineamiento de secuencias y reconstrucción filogenética de proteínas AIL .....	33
6.2.3 Alineamiento de secuencias y reconstrucción filogenética de proteínas PLT .....	33
6.3 Germinación de semillas de <i>Pachycereus pringlei</i> y su cultivo <i>in vitro</i> .....	34
6.4 Colecta de muestras biológicas para caracterizar la expresión de <i>PpPLT</i> .....	34
6.5 Extracción de RNA total.....	35
6.6 Preparación de geles de agarosa y electroforesis de RNA y DNA .....	35
6.7 Tratamiento de RNA con DNasa .....	35
6.8 Síntesis de cDNA a partir de RNA (Retro-transcripción) .....	35
6.9 RT-qPCR (PCR cuantitativa con transcriptasa reversa).....	36
6.9.1 Diseño de oligonucleótidos para RT-qPCR.....	36
6.9.2 Prueba de oligonucleótidos en PCR (Reacción en cadena de la polimerasa) de punto final.....	36
6.9.3 Ensayo de RT-qPCR).....	36
6.9.4 Normalización de la expresión de genes blanco .....	37
6.10 Amplificación de los extremos de cDNA por RACE-PCR.....	37
6.10.1 Diseño de oligonucleótidos para RACE-PCR .....	38
6.10.2 Síntesis de poblaciones de cDNAs con extremo 5' modificado.....	40
6.10.3 Síntesis de población de cDNAs con extremo 3' modificado.....	40
6.10.4 Síntesis de oligonucleótidos .....	40
6.10.5 Síntesis de población de cDNAs con extremo 5' modificado.....	40
6.10.6 Ensayo de amplificación de extremos de cDNA por PCR (RACE PCR) .....	41
Reacción primaria de RACE PCR 5'.....	41
Reacción secundaria de RACE PCR (anidada) 5' .....	41
Reacción primaria de RACE PCR 3'.....	42
Reacción secundaria de RACE PCR (anidada) 3' .....	42

<b>7. Resultados y discusión</b> .....	43
7.1 Análisis de secuencias obtenidas a partir del ensamblaje realizado <i>in silico</i> .....	43
7.2 Análisis filogenético.....	50
7.2.1 Clados A y B de <i>PLT</i> .....	56
7.2.2 Proteínas PLT en plantas monocotiledóneas y dicotiledóneas.....	58
Caracterización de la expresión de los genes <i>PpPLT</i> .....	59
7.3 Expresión Virtual de los <i>contigs</i> .....	59
7.4 Caracterización de la expresión de los genes <i>PpPLT</i> por RT-qPCR.....	60
7.4.1 Diseño de oligonucleótidos para RT-qPCR.....	60
7.4.2 Prueba de oligonucleótidos para RT-qPCR.....	61
7.4.3 Extracción de RNA total de muestras biológicas de puntas de la raíz primaria de <i>P. pringlei</i> en cuatro etapas del desarrollo para RT-qPCR.....	63
7.4.5 PCR en Tiempo Real (RT-qPCR).....	64
7.5 RACE-PCR.....	68
7.5.4 Síntesis de población de cDNAs con extremo modificado.....	68
7.5.5 RACE-PCR 5'.....	69
7.5.6 RACE-PCR 3'.....	70
<b>8. Nuevo mRNA-seq</b> .....	72
8.1 <i>PLT</i> en el nuevo ensamblaje del transcriptoma.....	73
8.2 Expresión virtual de los <i>contigs</i> de <i>PpPLT</i> del último mRNA-seq.....	76
8.3 Genes de referencia para RT-qPCR.....	79
8.4 Comparación entre la cuantificación por RT-qPCR y el análisis de expresión virtual del nuevo mRNA-seq.....	82
<b>9. Conclusiones</b> .....	85
<b>10. Perspectivas</b> .....	86
<b>11. Bibliografía</b> .....	87



## Lista de abreviaturas empleadas

*A. thaliana: Arabidopsis thaliana*

*AIL: AINTEGUMENTA LIKE*

*ANT: AINTEGUMENTA*

*AP2: APETALA2*

*ARF: Factor de respuesta a auxina (Auxin Response Factor)*

*At: Prefijo referente a Arabidopsis thaliana*

*BBM: BABY BOOM*

*cDNA: Ácido desoxirribonucleico complementario*

*CLV3: CLAVATA3*

*CQ: Centro quiescente*

*Cq: Ciclo en el cual la cantidad de producto de RT-qPCR alcanza un umbral preestablecido*

*DNA: Ácido desoxirribonucleico*

*d.d.g.: Días después de la germinación*

*ERF: Factor de unión a un elemento de respuesta a etileno (Ethylene-Responsive Element Binding FACTOR)*

*et al.: Y otros*

*nt: nucleótidos*

*ORF: Marco abierto de lectura (Open Reading Frame)*

*pb: pares de bases*

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

*PLT: PLETHORA*

*Pp: Prefijo referente a Pachycereus pringlei*

*RACE PCR: PCR de amplificación rápida de extremos de cDNA (Rapid Amplification cDNA Ends PCR)*

*RAM: Meristemo apical de la raíz (Shoot Apical Meristem)*

*RFG: Factor de crecimiento de la raíz (Root Growth Factor)*

*RPKM: Lecturas por kilobase por millón de lecturas mapeadas (Reads per Kilobase per Million Mapped Reads)*

*qPCR: PCR cuantitativo con transcripción inversa (quantitative reverse transcriptase PCR)*

*SAM: Meristemo apical del brote (Shoot Apical Meristem)*

*SCR: SCARECROW*

*SHR: SHORT ROOT*

*STM: SHOOT MERISTEMLESS*

*Tm: Temperatura de fusión (Melting Temperature)*

*UTR: Región no traducida (UnTranslated Region)*

*WT: Tipo silvestre (Wild Type)*

*WUS: WUSCHEL*

## Lista de símbolos

gr	gramos
hr	horas
l	Litro
μ	micro
m	mili ( $10^{-3}$ )
min	minutos
M	Molar
n	nano ( $10^{-9}$ )
pH	menos el logaritmo del potencial de hidrógeno
s	segundos
TM	Marca comercial
X	por
μ	Micro ( $10^{-6}$ )
-	Menos
®	Marca comercial registrada
%	Porcentaje
°C	Grados Celsius
~	aproximadamente

## Resumen

El meristemo apical de la raíz (RAM) que se localiza en la punta de la raíz contiene a las células pluripotenciales que usualmente mantienen el crecimiento de la raíz al generar nuevas células a lo largo de la vida de este órgano. Sin embargo, la raíz primaria de muchas especies de cactáceas presenta crecimiento determinado como consecuencia del agotamiento del RAM y diferenciación de las células de la punta de la raíz. Esta característica única entre angiospermas es considerada una ventaja adaptativa en hábitats de clima árido. En *Arabidopsis thaliana* los factores de transcripción PLETHORA (PLT), que son parcialmente redundantes, regulan el establecimiento y mantenimiento del RAM. Nuestra investigación se enfoca en elucidar el papel de los genes *PLT* en el crecimiento determinado de la raíz primaria de la cactácea *Pachycereus pringlei*.

Usando el ensamblaje *de novo* de lecturas de mRNA-seq (lecturas de 85 nt no pareadas) identificamos seis posibles homólogos a *AtPLT* en la raíz primaria de *P. pringlei*. Se generó un amplio árbol filogenético para analizar la diversificación de PpPLTs en el contexto de las especies de angiospermas. Todas las secuencias de PpPLT se agruparon con secuencias homólogas de *A. thaliana*. El alto grado de conservación de las proteínas PLT sugiere que las funciones de estas proteínas puedan estar conservadas.

La división en clados de las proteínas PLT de angiospermas que encontramos concuerda con la especificación funcional de los genes previamente documentada. Uno de los clados contiene a posibles homólogos de *PLT1*, *PLT2* y *PLT4/Baby Boom (BBM)* que son necesarios para establecer y mantener el RAM en la raíz primaria de *Arabidopsis*. Mientras que, los de *PLT3*, *PLT5* y *PLT7*, involucrados en el posicionamiento y establecimiento de órganos laterales de raíz y brote de *Arabidopsis*, se agrupan en un segundo clado.

Para el mRNA-seq se utilizaron muestras de la punta de la raíz primaria de *P. pringlei* de dos etapas del desarrollo: la etapa inicial (cuando el RAM está presente) y la etapa terminal (cuando las células en al punta de la raíz ya se han diferenciado). El análisis de expresión virtual para los transcritos de *PpPLTs* indicaba una disminución en su abundancia en la etapa terminal del desarrollo. Para corroborar estos datos se realizaron ensayos de RT-qPCR para tres de estos genes. En este análisis se incluyó una sub-división de la etapa inicial y una etapa intermedia del desarrollo de la raíz primaria de *P. pringlei*. Las secuencias de los transcritos de *PpPLTs* resultado del ensamblaje *de novo* se encontraban incompletas, por lo cual en el presente trabajo se buscó obtener las secuencias completas de los transcritos de *PpPLT* mediante RACE-PCR.

A la par que se realizaba este trabajo se resecuenció el transcriptoma de la punta de la raíz primaria de *P. pringlei* con dos réplicas biológicas, incluyendo, además de las etapas inicial y terminal, una etapa intermedia. Se generaron 280 millones de lecturas pareadas de 100 nt.

Como resultado del reciente ensamblaje *de novo* se cuenta ahora con las secuencias casi completas de los transcritos *PpPLT*. Dichos transcritos presentan un patrón de expresión virtual en el que los valores de RPKM disminuyen en la etapa terminal. Este es el comportamiento que esperábamos encontrar para factores de transcripción *PpPLT*, es decir una menor expresión en la etapa del desarrollo en la que ya no hay células meristemáticas. Para los genes *PpPLT2a* y *PpAIL* la menor abundancia de transcrito en la etapa terminal pudo corroborarse mediante RT-qPCR.

Los datos de expresión virtual y RT-qPCR indican que los *PpPLT* posibles ortólogos de *PLT1*, *PLT2* y *BBM* tienen una muy baja expresión en la punta de la raíz. Los resultados obtenidos sugieren que la expresión de *PpPLT2a*, *PpPLT2b* y *PpBBM* puede no ser suficiente para alcanzar los umbrales necesarios en el mantenimiento del RAM.

## Summary

The root apical meristem (RAM) is localized at the root tip and usually generates new cells over the root lifespan thus maintaining its indeterminate growth. Nevertheless, the primary root of many Cactaceae species exhibits determinate growth as a result of RAM exhaustion and differentiation of the root tip cells. This unique feature among angiosperms is considered an adaptive advantage to arid habitats. In the *Arabidopsis thaliana* root, the partially redundant PLETHORA (PLT) transcription factors regulate establishment and maintenance of the RAM. Our research is focused on elucidating the role of *PLT* genes in the determinate growth of the primary root of the *Pachycereus pringlei* cactus.

Using the *de novo* assembly (mRNA-seq, single-end reads, 85 cycles) we identified six *AtPLT* homologs in *P. pringlei*. An exhaustive PLT phylogenetic tree was generated to analyze the diversification of PpPLTs in the context of angiosperm species. Each PpPLT sequences clustered an *A. thaliana* counterpart. The high degree of conservation in the PLT protein sequences suggest that the functions of these proteins might also be conserved.

We found that the resulting clade division of the angiosperm PLT proteins is congruent with previously reported gene specification. The angiosperm PLT proteins are divided in two major clades. One of the clades contains the PLT1, PLT2 and PLT4/Baby Boom (BBM) putative orthologs which are needed for the RAM establishment and maintenance in the *Arabidopsis* primary root. Meanwhile, PLT3, PLT5 and PLT7 putative orthologs, involved in the correct positioning and establishment of lateral organs in *Arabidopsis* shoot and root, were grouped in a second clade.

The *P. pringlei* primary root tip samples used for mRNA-seq were generated from two developmental stages: initial stage (when RAM is present) and terminal stage (when cells are differentiated). The virtual expression analysis of the assembled *PpPLTs* transcripts indicated a decrease in abundance at the terminal stage of root development. This observation was further confirmed by the RT-qPCR assays, in which we also included a subdivision of the initial growth stage and an intermediate stage. The assembled *PpPLT* sequences were incomplete, lacking the UTR regions and in some cases a part of the ORF. Therefore, we used RACE-PCR with the aim of obtaining the complete transcript sequence. By the time this work was completed, a new mRNA-seq transcriptome of *P. pringlei* primary root tips was sequenced and *de novo* assembled. In addition to the initial and terminal stages of root development the new transcriptome sequenced included an intermediate stage, and two biological replicates per sample. Almost 280 millions of 100 nt paired-end reads were generated.

The larger number of reads of better quality allowed assembling of the nearly complete sequences of the *PpPLT* transcripts. A virtual expression pattern in which the RPKM values decrease in the terminal stage was confirmed for these *PpPLT* transcripts. This behavior, that is, the lower transcript abundance at the terminal stage when the RAM is consumed, could be expected for PLT transcription factors. In the case of *PpPLT2a* and *PpAIL* this was corroborated by RT-qPCR.

The virtual expression data and the RT-qPCR results indicated that in general the *PLT1*, *PLT2* y *BBM* putative orthologs of *P. pringlei* have a low expression in the root tip.

Our results suggest that the expression level of *PpPLT2a*, *PpPLT2b* y *PpBBM* might be insufficient to reach the threshold needed for the maintenance of the RAM.

# 1 Introducción

Uno de los grandes retos de la biología es entender los procesos del desarrollo de organismos multicelulares, es decir los procesos que dan origen a un individuo a partir del huevo o cigoto. Al igual que en otros organismos multicelulares, el estudio del proceso del desarrollo de las plantas incluye la investigación de la diferenciación celular y de la organogénesis. Parte de estos estudios han contribuido de manera importante al conocimiento de diversos procesos moleculares y celulares durante el desarrollo.

Al combinarse los procesos del desarrollo con las variaciones genéticas presentes en la naturaleza, obtenemos como resultado la gran diversidad de formas que existe entre las especies vegetales y la plasticidad innata que poseen para responder ante condiciones variables en el ambiente. Las plantas han logrado adaptarse a la mayoría de los ecosistemas terrestres, y son constituyentes estructurales y funcionales claves del planeta, al desempeñar un importante papel en el ciclo del carbono, en la producción de oxígeno, además de proveer de materias primas y alimentos a la humanidad. A pesar de la gran biodiversidad que existe, se observan características anatómicas similares entre diferentes especies de angiospermas. Esta similitud sugiere que existen mecanismos similares en los procesos del desarrollo vegetal, por lo que es posible que dichos mecanismos se encuentren ampliamente conservados entre distintas plantas (Gilbert, 2005).

Las plantas son organismos sésiles, es por ello que han tenido que desarrollar la capacidad de obtener energía y nutrientes orgánicos por medio de la fotosíntesis. Además, la falta de motilidad de las plantas impide que estas puedan desplazarse hacia el ambiente más propicio para su crecimiento, es por esto que las plantas deben adaptarse al ambiente en el que se desarrollan (Taiz & Zaiger, 2006).

La adaptación de las plantas ante las condiciones adversas del ambiente puede ocurrir a nivel fisiológico, en gran parte debido a la flexibilidad de sus patrones de desarrollo. Dichos procesos de desarrollo se retroalimentan continuamente de las condiciones ambientales locales, es por eso que el desarrollo vegetal es mucho más plástico que el animal. A diferencia de los animales, cuyo plano corporal adulto se establece durante la embriogénesis, las plantas moldean su forma a lo largo de su vida a través de la regulación de la proliferación, expansión y diferenciación celular, lo cual les permite producir formas variables y complejas que se adaptan mejor al ambiente (Taiz & Zaiger, 2006).

## 1.1 Características del crecimiento post-embrionario en plantas

En las plantas vasculares después de la germinación de la semilla se desarrolla la plántula, en la cual quedan establecidos los ejes corporales básicos. En los ápices (o extremos) de la plántula se encuentran grupos de células pluripotenciales, la mayoría

de estas con una alta tasa de división celular. A estos grupos de células se les denomina meristemos apicales (Fig. 1). Las células que conforman a los meristemos apicales se encuentran poco diferenciadas, conservan la capacidad de proliferación y pueden auto-renovarse.

Es hasta el crecimiento post-embriionario que se forman la mayoría de las células de los tejidos que conformarán los órganos del cuerpo de la planta, estas células se originan en los meristemos apicales por divisiones mitóticas. Es en el meristemo apical del brote (SAM por sus siglas en inglés) en donde se generan las nuevas células que, tras diferenciarse, formarán la parte aérea de la planta; mientras que en el meristemo apical de la raíz (RAM por sus siglas en inglés) se generan las células que formarán el sistema radical (Weigel & Jürgens, 2002).

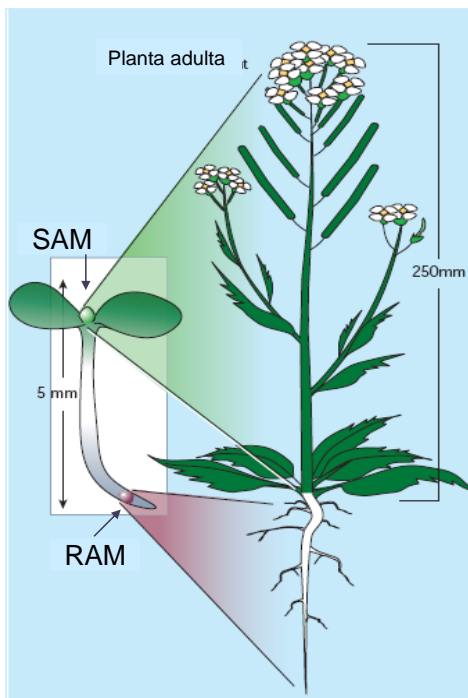


Fig. 1. Los meristemos apicales de la raíz y del brote dan origen a los órganos de la planta adulta. Del lado izquierdo en la plántula se señalan los meristemos apicales. Del lado derecho en la planta adulta se señalan los órganos a los que darán origen el SAM y RAM respectivamente (Modificado de Weigel & Jürgens, 2002).

Las divisiones en las células meristemáticas están reguladas mediante programas de desarrollo que deben especificar y sincronizar el tiempo de división celular, la orientación del plano de división celular y la expansión celular para dar origen a los órganos y tejidos (Steeves & Sussex, 1989). La actividad meristemática, es decir, la formación de nuevas células en los meristemos, varía de una especie a otra, y de un órgano a otro. Para mantener la actividad de los meristemos apicales se necesita de un balance entre la generación de nuevas células y la diferenciación de las células que se generaron en el meristemo, pues de esta manera el tamaño del meristemo puede conservarse relativamente constante (Sablowski, 2007).

En las raíces de algunas gimnospermas se ha observado que los meristemos pueden pausar su actividad y, como consecuencia, el crecimiento de la raíz, para reanudarlo tiempo después (Wilcox, 1962). Los meristemos que permanecen durante toda la vida de los órganos se conocen como meristemos indeterminados. Sin embargo, existen casos en los que el meristemo parece estar programado para detener irreversiblemente la proliferación celular en una etapa específica del desarrollo; a este tipo de meristemo se le llama meristemo determinado. En estos casos las células que dejaron de dividirse, se diferencian y el meristemo desaparece, o se agota. Los meristemos indeterminados producen partes de la planta cuyo tamaño y forma se ven afectados por las condiciones ambientales, los brotes por ejemplo, cuyo número de hojas es variable. Por otra parte los meristemos determinados producen órganos o tejidos cuya forma, tamaño y número son predecibles, las flores por ejemplo (Sablowski, 2007).

La estructura y organización celular del SAM es diferente a la organización celular del RAM. Las células del SAM dan lugar a varios órganos como las hojas, el tallo, y los órganos florales; mientras que el meristemo primario de la raíz da lugar a un solo órgano: la raíz primaria (Scheres *et al.*, 1995). Los primordios de las raíces laterales de varias especies vegetales se originan de las células del periciclo con o sin la participación de células de córtex en especies diferentes a cierta distancia del RAM (Dolan *et al.*, 1993). Las raíces laterales de las plantas se originan por restablecimiento de divisiones celulares en la zona proximal al RAM y por el establecimiento *de novo* de nuevos RAM. La parte aérea de la planta sufre una transición de crecimiento vegetativo a crecimiento generativo, es decir, que se produce un cambio en el tipo de órganos que se generan (Esau, 1977). En contraste, el RAM no presenta este tipo de transiciones durante el desarrollo.

Una pregunta fundamental en el campo del desarrollo de plantas es ¿cómo está regulada la formación de órganos? y más específicamente ¿cuál es el mecanismo por el que se logra mantener el balance necesario para que el meristemo permanezca activo? Aun cuando la organización celular del SAM y del RAM es muy diferente, los componentes de la señalización requeridos para el mantenimiento de ambos meristemos parecen estar relativamente conservados. Ambos tipos de meristemos apicales utilizan la comunicación célula-célula para mantener la actividad y organización meristemática y para coordinar la formación de nuevos órganos (Miwa *et al.*, 2009). Un ejemplo de ello es el módulo genético que en la planta modelo *Arabidopsis thaliana* está formado por CLAVATA3 (CLV3) y WUSCHEL (WUS) y que ajusta la homeostasis de la población de células madre en el SAM. El gen CLV3 que codifica el péptido ligando del receptor tipo cinasa CLV1, se expresa en las células meristemáticas, reprimiendo al factor de transcripción WUSCHEL (WUS) en las células adyacentes y activando su propia expresión por retroalimentación negativa. En el RAM existe un módulo genético análogo en el que el péptido CLE40, por medio del receptor tipo cinasa *Arabidopsis homolog of Crinkly4* (ACR4), reprime al homólogo de WUS (WOX5) regulando así el tamaño del RAM (Stahl & Simon, 2009).

## 1.2 Importancia y desarrollo del sistema radical en plantas

El conjunto de raíces de una planta conforma el sistema radical, el cual realiza funciones esenciales para la supervivencia y adaptación de las plantas, como son la absorción de agua, nutrientes y solutos. El sistema radical sirve de anclaje al sustrato. Además, la porción de suelo en donde se localizan las raíces, la rizósfera, es un sitio en el que se establecen interacciones bióticas con diversos organismos como bacterias y hongos (Malamy & Benfey, 1997).

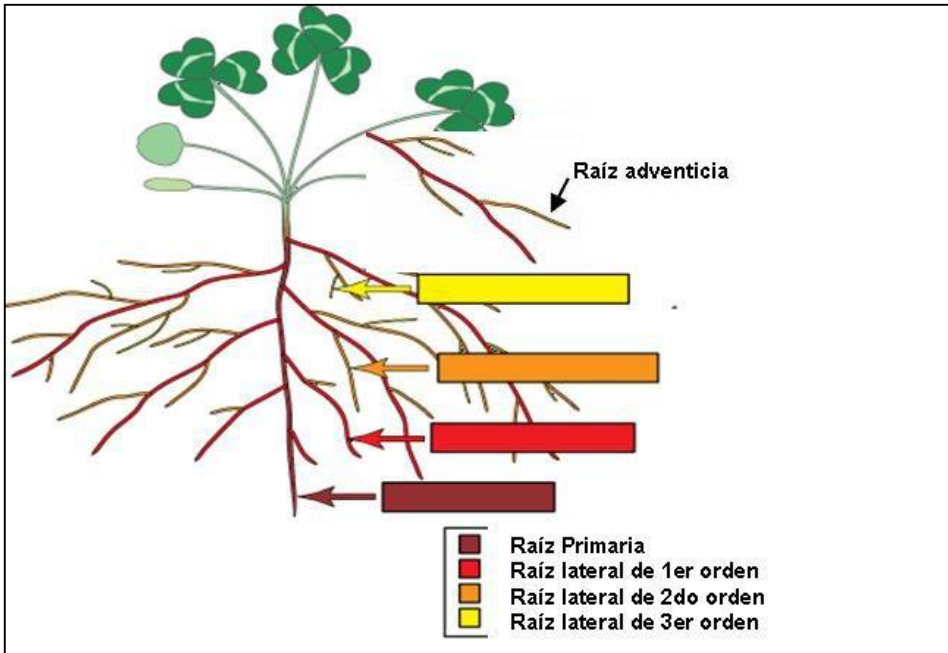


Fig. 2. Sistema radical. En el esquema se señalan el tipo de raíces que pueden integrar el sistema radical (Modificado de <http://mycorrhizas.info/root.html>).

La radícula o raíz embrionaria, emerge de la semilla durante la germinación y se elonga para originar la raíz primaria. A partir de la raíz primaria se formarán las raíces laterales de primer orden, y a partir de estas se formarán las raíces laterales de segundo orden, y así sucesivamente (Fig. 2). Si bien en las plantas dicotiledóneas la mayor parte del sistema radical se desarrolla a partir de las raíces, existen algunas raíces que se desarrollan a partir de órganos distintos, a estas raíces se les conoce como raíces adventicias.

El desarrollo de raíces es muy sensible a las condiciones ambientales y aun cuando el desarrollo del sistema radical de las plantas es de gran importancia en la agronomía, para la mayoría de las especies vegetales se conoce poco de los procesos y señales que regulan su desarrollo. En este trabajo nos centramos en el estudio de la raíz primaria, ya que a partir de ésta se desarrolla la mayor parte del sistema radical de plantas dicotiledóneas. *Pachycereus pringlei*, la planta modelo de estudio que elegimos en el presente trabajo, es una de las únicas especies entre las angiospermas con crecimiento determinado de la raíz primaria (Dubrovsky, 1997).



### 1.3 El crecimiento de la raíz en plantas

Cuando el crecimiento de la raíz es indeterminado, el proceso de desarrollo es continuo. En el proceso de desarrollo de una raíz los diferentes tipos celulares emergen en filas a partir del meristemo apical durante toda la vida del órgano. Es por ello que en todo momento las diferentes etapas del desarrollo están presentes en diferentes regiones distinguibles a lo largo de la raíz. Las zonas del desarrollo se dividen en zona meristemática, de elongación y de diferenciación (Fig. 3) (Scheres *et al.*, 2002).

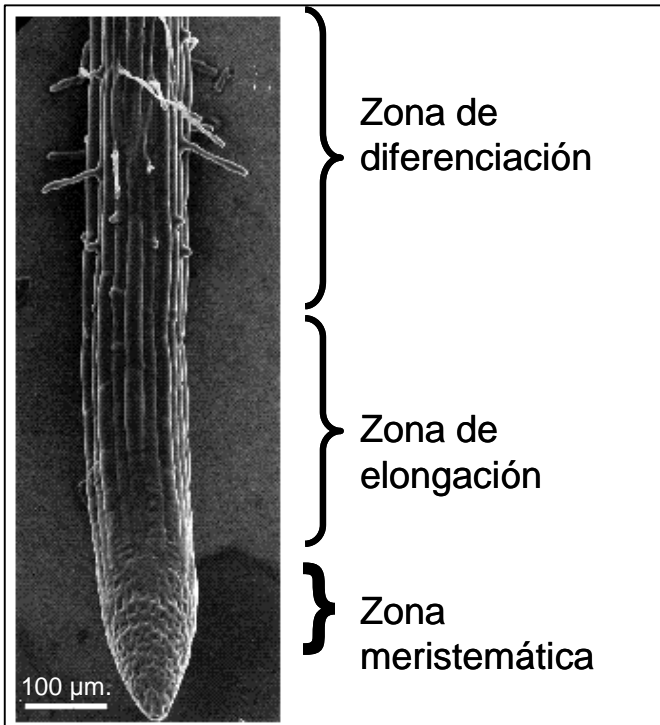


Fig. 3. Zonas de la raíz. Vista exterior de una raíz de *A. thaliana* en la que se señalan las diferentes zonas de la raíz (Dolan *et al.*, 1993).

La zona meristemática es la más cercana al ápice (Fig. 3), en la parte más distal está delimitada y protegida por la cofia. En dicha zona se encuentra el RAM, en donde se forman nuevas células, por lo que podemos encontrar células de menor tamaño que en el resto de la raíz.

La siguiente zona es la de elongación (Fig. 3), en donde se alargan las células que se formaron en la zona meristemática. Las células que se van elongando van empujando a las células de la zona meristemática; esto hace que el ápice se desplace dentro del sustrato y la raíz crezca.

Finalmente, en la zona de diferenciación celular (Fig. 3), las células adquieren su identidad final; es por ello que en esta zona podemos encontrar células totalmente diferenciadas, como lo son las células epidérmicas que al diferenciarse forman los pelos radicales (Fig. 3, Dolan *et al.*, 1993).

## 1.4 La función del meristemo apical de la raíz en plantas

A diferencia del SAM, se dice que el RAM es sub-terminal, ya que está cubierto por un grupo de células, denominado cofia, que brindan protección al meristemo cuando la raíz va abriéndose camino en el substrato (Dolan *et al.*, 1993), además de ejercer funciones indispensables para la planta como lo son el gravitropismo y el hidrotropismo (Sievers & Braun, 1996).

La raíz primaria de *A. thaliana* se ha convertido en uno de los modelos más estudiados de la biología del desarrollo de plantas debido a su semi-transparencia y elegante simplicidad. Al realizar un corte transversal a la raíz primaria de *A. thaliana* (Fig. 4) su organización celular puede ser vista como una serie de cilindros concéntricos, a esto se le conoce como patrón radial.

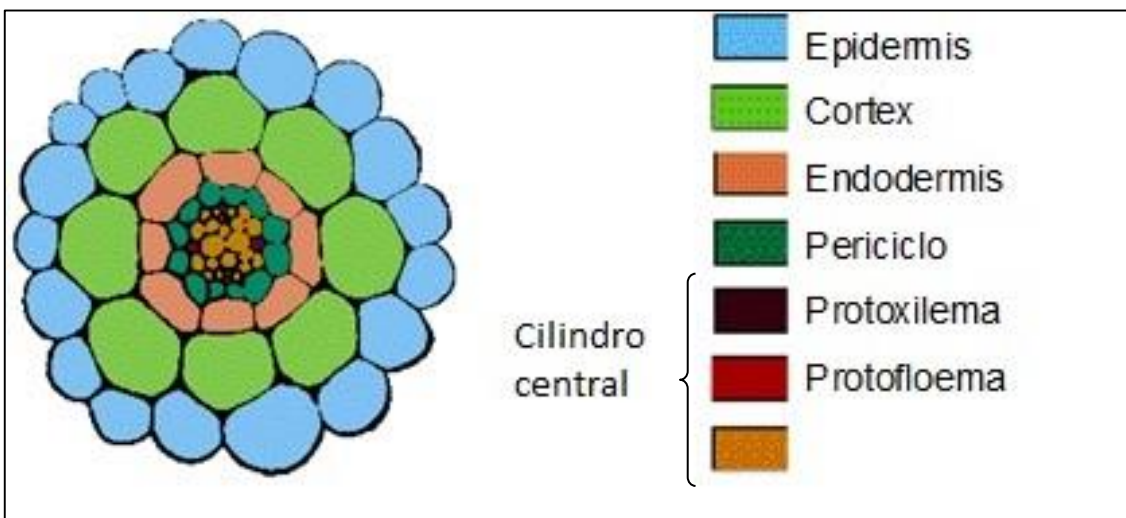


Fig. 4. Esquema del corte transversal de la raíz de *A. thaliana* a aproximadamente 1mm del ápice, en el que se hace distinción de los diferentes tipos celulares que conforman la raíz (Dolan *et al.*, 1993).

En la parte más apical del RAM se encuentra el centro quiescente (CQ), que en la raíz primaria joven de *A. thaliana* está conformado por cuatro a ocho células completamente indiferenciadas con una actividad mitótica casi nula. El centro quiescente se encuentra rodeado por una capa de células con mayor actividad mitótica llamadas células iniciales (también conocidas como células madre o células troncales) (Fig. 5). Las células iniciales de córtex/endodermis, epidermis/cofia lateral, columela y cilindro vascular dan origen a los tipos celulares correspondientes de la raíz, mediante divisiones mitóticas orientadas. Del lado derecho de la figura 5 se muestran las divisiones que ocurren en las células iniciales de córtex/endodermis para generar células de ambos tipos celulares (Scheres *et al.*, 2002).

Las células del RAM que se encuentran en la zona más proximal a las células iniciales presentan mayor actividad mitótica. El RAM se delimita en la zona en donde cesa la actividad mitótica (Scheres *et al.*, 2002).

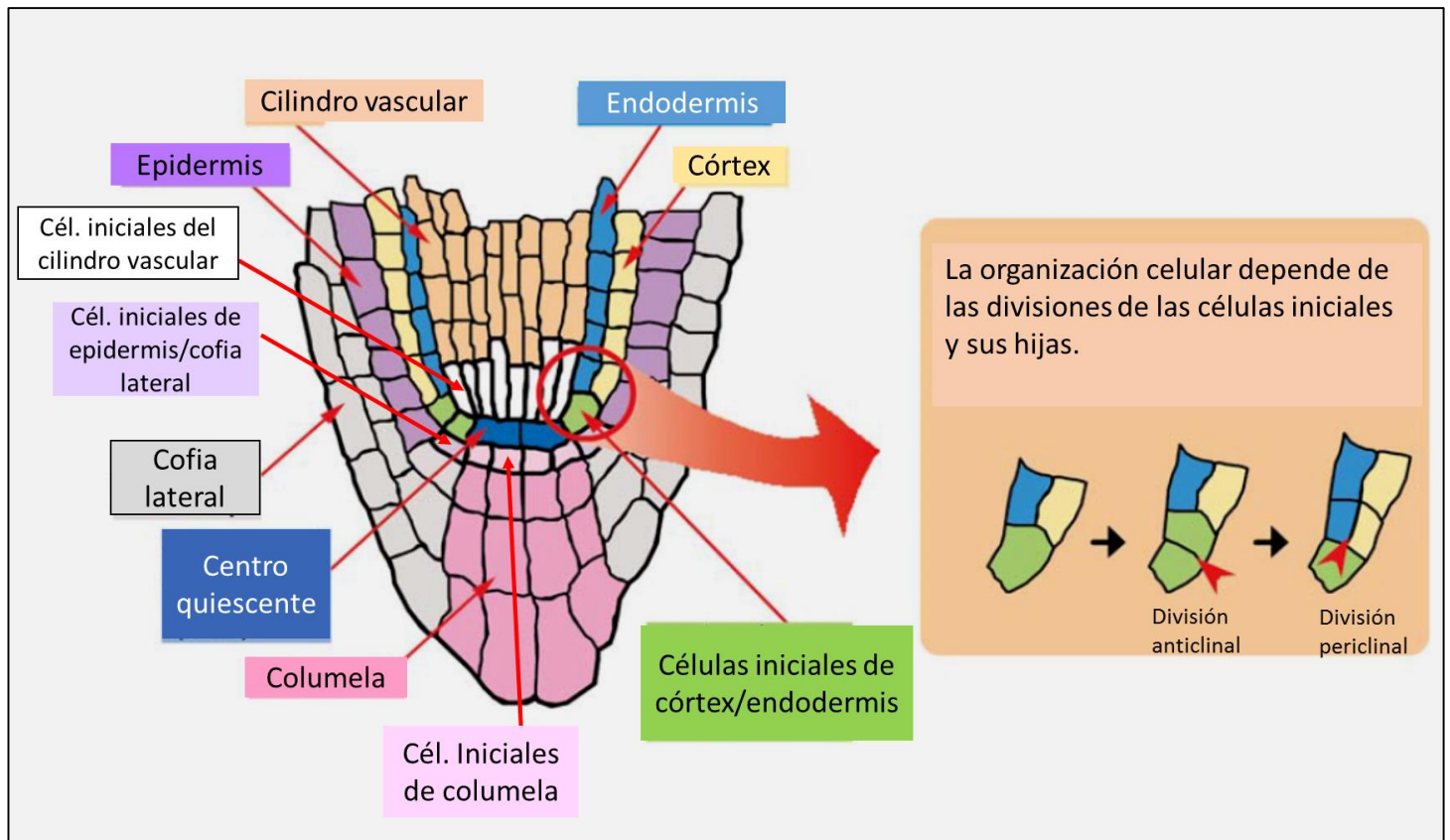


Fig. 5. Meristemo apical de la raíz primaria de *A. thaliana*. En la imagen se ilustra un corte longitudinal del ápice de la raíz, los colores denotan los diferentes tipos celulares presentes en el RAM. Del lado derecho se esquematizan las dos divisiones asimétricas que sufren las células iniciales de córtex/endodermis. División periclinal: paralela a la superficie de la raíz. División anticlinal: perpendicular a la superficie de la raíz (modificado de Scheres *et al.*, 2002).

En *A. thaliana* el linaje celular puede rastrearse hasta unas cuantas células iniciales (Fig. 5), por lo que se dice que su meristemo es de tipo cerrado (Dolan *et al.*, 1993).

### 1.5 Características del crecimiento determinado de la raíz en plantas

En las raíces la presencia del RAM, en el órgano maduro, se considera el principal criterio de indeterminación, sin importar si este se encuentra activo o inactivo (es decir sin actividad mitótica). El RAM se compone por células pequeñas, no diferenciadas, que presentan un citoplasma denso, con vacuolas pequeñas y paredes celulares delgadas. El crecimiento determinado se presenta cuando el RAM se agota y las células de la punta de la raíz se diferencian (Shishkova *et al.*, 2008).

Para ilustrar las diferencias entre el crecimiento determinado e indeterminado se muestra en la figura 6 una raíz primaria en la que la zona meristemática está presente (es decir que tiene crecimiento indeterminado) y otra raíz en la que las células del ápice están diferenciadas, por lo tanto no tiene una zona meristemática y no continuará su crecimiento (es decir que tiene crecimiento determinado).

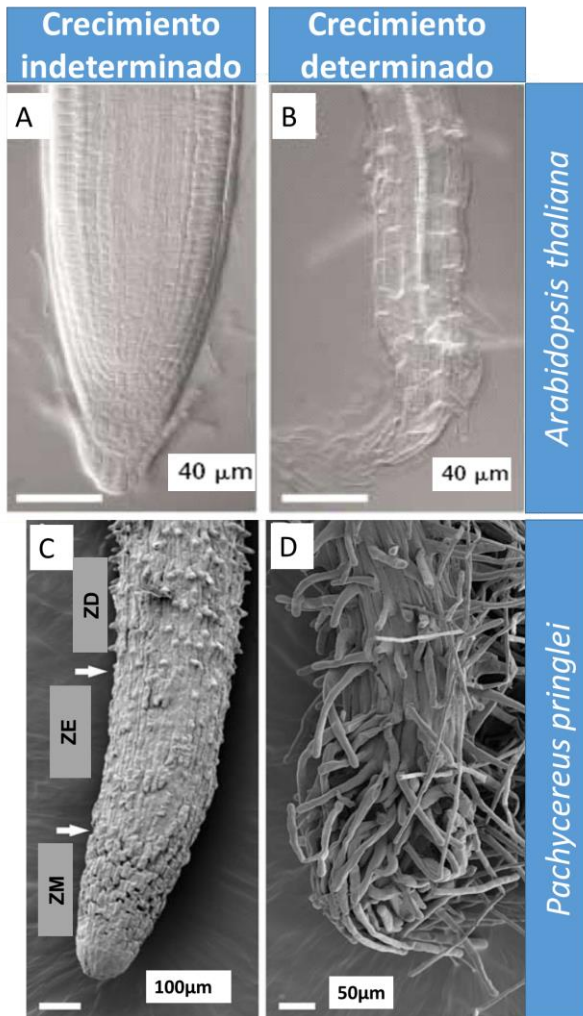


Fig. 6. Crecimiento indeterminado *Vs* crecimiento determinado de la raíz. A) y B) Raíces primarias de *Arabidopsis thaliana* a los 8 d.d.g. (días después de la germinación). A) planta de tipo silvestre; la raíz primaria mantiene el RAM y crece por largos periodos de tiempo B) planta mutante *plt1 plt2*; el RAM de la raíz primaria no se mantiene y la raíz deja de crecer, de lo que se hablará más tarde en el texto (Aida *et al.*, 2004).

C) y D) Raíces primarias de plantas de la cactácea *Stenocereus gummosus*, microscopia electrónica. En esta especie el RAM se diferencia a los pocos días de haber emergido la raíz. C) Raíz recientemente emergida contiene las tres zonas de desarrollo. Z.D (zona de diferenciación), Z.E. zona de elongación, Z.M. (zona meristemática). D) Raíz en la que se agotó el RAM y las células del ápice están totalmente diferenciadas (Shishkova & Dubrovsky, 2005).

El crecimiento determinado puede ser inducido bajo algunas condiciones. Por ejemplo, en el trigo, cuando el eje de la raíz se ve obstaculizado, el crecimiento de la raíz se detiene, el ápice de la raíz se cubre de pelos radicales, y empiezan a emerger raíces laterales cerca del ápice de esta raíz (Watt *et al.*, 2006).

En *A. thaliana*, el crecimiento determinado de la raíz primaria puede inducirse por algunos factores ambientales, en particular, cuando las plantas crecen en condiciones limitantes de fósforo. Se ha demostrado experimentalmente que en plántulas germinadas en medio nutritivo que contiene 1  $\mu\text{M}$  de  $\text{NaH}_2\text{PO}_4$ , el número de células en el meristemo apical de la raíz primaria gradualmente decrece hasta que no pueden encontrarse células meristemáticas. En estas raíces, todas las células de la punta de la raíz se diferencian completamente y el agotamiento del RAM se hace irreversible. Sin embargo, si estas plántulas se transfieren a un medio nutritivo con el nivel óptimo de fósforo (1 mM de  $\text{NaH}_2\text{PO}_4$ ) en la etapa en la que aún existen células meristemáticas, el proceso de agotamiento del RAM puede revertirse y las raíces restablecen su crecimiento (Sánchez-Calderón *et al.*, 2005). Debe considerarse además que la biodisponibilidad de fósforo se ve afectada por la presencia de hierro o por un pH ácido en el sustrato. Es decir que, cuando hay una baja cantidad de fósforo y además

el sustrato es ácido o contiene hierro, el fosfato se vuelve no disponible para la planta y se agota el RAM. En cambio si no hay hierro presente, aún a una baja concentración de fosfato, el RAM seguirá disponible y no se agotará (Svistoonoff *et al.*, 2007).

El crecimiento determinado en raíces laterales ha sido reportado en varias especies vegetales. Existen reportes de raíces adventicias con crecimiento determinado, como es el caso de algunas pteridofitas, entre ellas *Azolla* (Webster & MacLeod, 1996). Un ejemplo de raíces laterales con crecimiento determinado son las raíces proteoides, que son muchas raíces pequeñas que se desarrollan en fragmentos cortos del eje de una raíz, dándole a la raíz una apariencia de escobillón (Skene *et al.*, 1998). Las raíces proteoides se encontraron por primera vez en la familia Proteaceae. Posteriormente se encontraron raíces proteoides en miembros de otras siete familias: Casuarinaceae, Myricaceae, Fabaceae, Moraceae, Betulaceae, Cucurbitaceae y Elaeagnaceae (Shane & Lambers, 2005).

En general, el tiempo de vida de diferentes tipos de raíces es altamente variable, desde unas pocas semanas hasta varios años (Eissenstat *et al.*, 2000). En la mayoría de las angiospermas la raíz primaria crece por un periodo prolongado de tiempo. Sin embargo, en muchas especies pertenecientes a la familia Cactaceae se ha encontrado crecimiento determinado de la raíz primaria (Dubrovsky, 1997; Shishkova *et al.*, 2013).

### **1.6 Importancia ecológica del crecimiento determinado de la raíz primaria de las cactáceas desérticas**

El primer y hasta hoy en día el único grupo de angiospermas en las que se reportó crecimiento determinado de la raíz *primaria* en plantas de tipo silvestre y en todas las condiciones de crecimiento son las cactáceas del Desierto de Sonora (Dubrovsky, 1997). Este patrón de crecimiento de la raíz primaria y raíces laterales da como resultado un sistema radical compacto y cercano a la superficie del suelo, que facilita la adquisición de la escasa agua de lluvia. Esta es una de las adaptaciones evolutivas que han permitido a las cactáceas establecerse en condiciones de extrema sequía (Dubrovsky, 1997).

Las plantas suculentas del desierto usualmente presentan un sistema radical que no es muy profundo. La relación entre el peso seco de la raíz y el peso seco de la parte aérea es de 0.12 en promedio; esta proporción es 10 veces mayor en otros grupos de plantas desérticas (Nobel, 2002).

En el Desierto de Sonora además de que la disponibilidad de agua es muy limitada, la temperatura de la superficie del suelo puede exceder los 70°C (Nobel, 1988). Los periodos óptimos de germinación en el desierto dependen de las precipitaciones, por lo que son periodos muy cortos. Tras la lluvia el potencial hídrico del suelo decrece rápidamente, y la supervivencia de las plántulas depende de su biomasa (Dubrovsky, 1996). Bajo estas condiciones extremas el establecimiento de las plántulas que germinan tiene que ser un proceso sumamente eficiente.

Entre las primeras especies de cactáceas del Desierto de Sonora en las que se estudió el patrón de crecimiento de la raíz primaria se encuentran: el cardón *Pachycereus pringlei*, la pitaya agria *Stenocereus gummosus*, la pitaya dulce *Stenocereus thurberi* y la biznaga *Ferocactus peninsulae* (Dubrovsky, 1997, Dubrovsky y Gómez-Lomelí, 2003).

En estas especies una vez que ha emergido la raíz primaria, esta crece únicamente por unos días, para después detener su crecimiento. En nuestro laboratorio se demostró que durante el período del desarrollo post-embrionario en *P. pringlei* el CQ existe solamente por un tiempo corto para luego desaparecer, mientras que en el caso de *S. gummosus* el CQ no se establece (Rodríguez-Rodríguez *et al.*, 2003).

Dicho patrón se ha estudiado en nuestro laboratorio en plántulas crecidas *in vitro* en diferentes substratos: en papel filtro, en arena, tierra, y diferentes medios de cultivo. También se ha analizado el crecimiento de dichas plántulas bajo diversas condiciones variando el potencial hídrico, las concentraciones de sacarosa, las concentraciones de fosfatos, tras adicionar glutatión (que es importante en los cambios del estado RedOx), auxinas o inhibidores de transporte polar de auxinas, entre otras condiciones (Dubrovsky, 1997; Shishkova & Dubrovsky, 2008 y datos no publicados del laboratorio).

Cada una de las miles de plántulas analizadas presentó crecimiento determinado de la raíz primaria en todas las condiciones, como se ejemplifica en la gráfica de la cinética de crecimiento de la raíz primaria mostrada en la figura 7D (Dubrovsky, 1997).

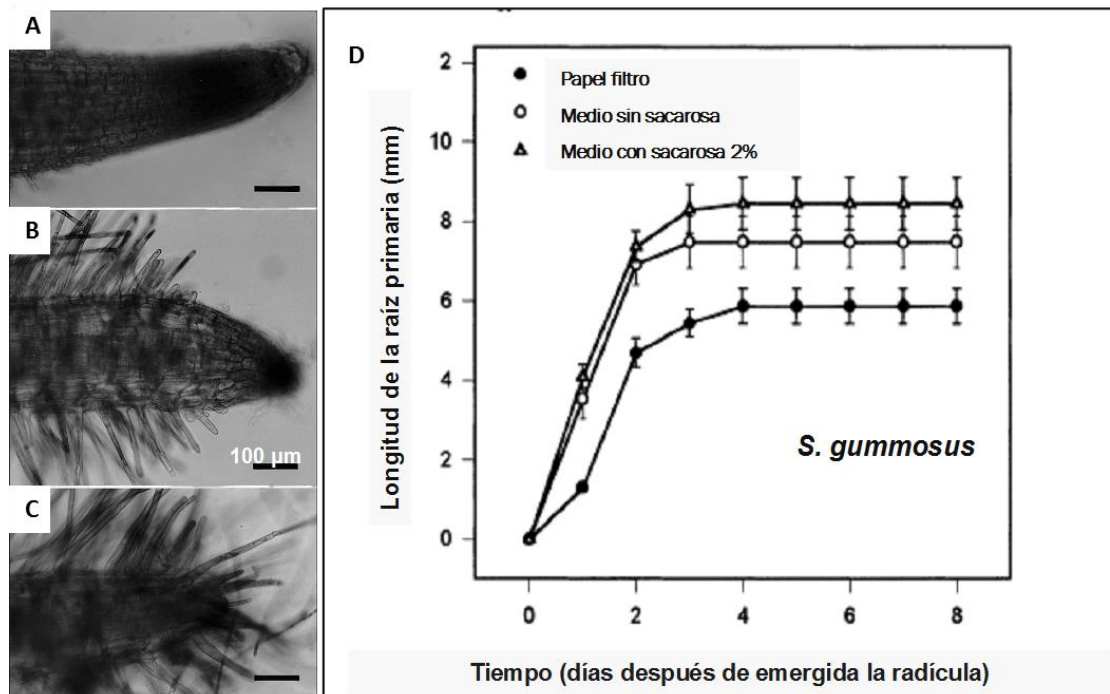


Fig. 7. Crecimiento determinado de la raíz primaria de *Stenocereus gummosus*. A-C) Ápices de la raíz primaria. A) Etapa inicial del desarrollo (el RAM aún está presente) 1 d.d.g. B) Etapa intermedia del desarrollo, 1 d.d.g. C) Etapa terminal del desarrollo (las células del ápice ya se han diferenciado), 2 d.d.g. La presencia de pelos radicales que cubren por completo la punta de la raíz indica que el RAM se ha agotado. Raíces teñidas con A)

tinción Feulgen, B-C) azul de toluidina. D) Cinética de crecimiento de la raíz primaria de *S. gummosus* (Dubrovsky, 1997).

Al diferenciarse, todas las células de la punta de la raíz, donde estuvo el RAM localizado, forman pelos radicales y esta característica morfológica es un signo de que el RAM se ha agotado (Fig. 7 A-C). Los numerosos pelos radicales que se forman por todas las células epidérmicas promueven la absorción de agua y nutrientes (Dubrovsky, 1998).

En cuanto a los mecanismos que conllevan al crecimiento determinado, se ha evidenciado que la muerte celular programada no está involucrada en el agotamiento del RAM (Shishkova *et al.*, 2005). Debido a esto, la pérdida de la actividad meristemática puede atribuirse por completo a un fenómeno de diferenciación celular.

El crecimiento se frena cuando el RAM se diferencia, lo que induce la formación de raíces laterales. En *S. gummosus*, *S. thurberi* y *F. peninsulæ*, se encontró una relación directa entre el agotamiento del RAM de la raíz primaria y la formación de raíces laterales (Fig. 8) (Dubrovsky, 1997). Las raíces laterales de las especies estudiadas también presentan crecimiento determinado. Por ello el sistema radical que se genera es compacto y ramificado (Dubrovsky, 1997).

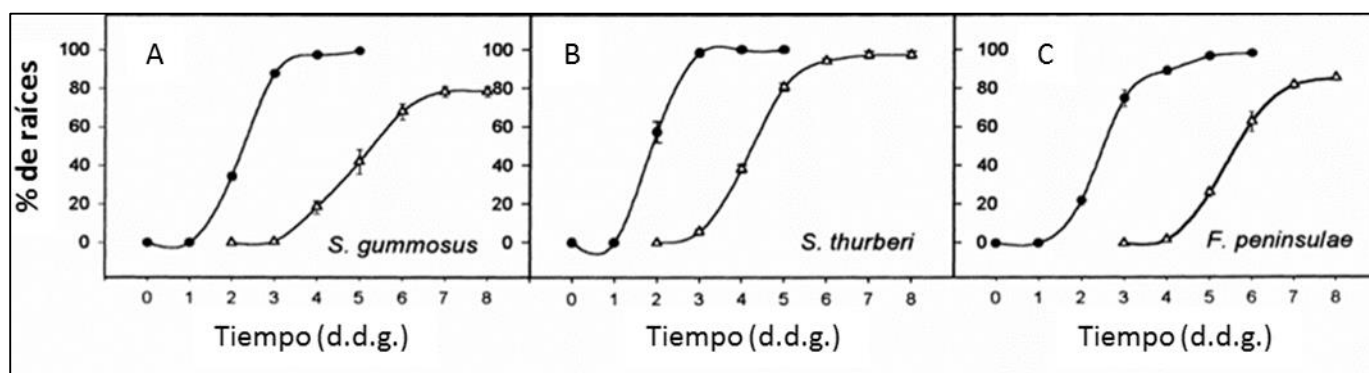


Fig. 8. Dinámicas de pérdida del meristemo (círculos) y formación de raíces laterales (triángulos) en tres especies de cactáceas del Desierto de Sonora (Dubrovsky, 1997).

Posteriormente se encontró que las raíces regeneradas a partir de callo de *S. gummosus* y *F. peninsulæ* también presentan crecimiento determinado (Shishkova *et al.*, 2007). Esto evidencia que el programa de desarrollo que origina este patrón de crecimiento está conservado entre diferentes tipos de raíces en cactáceas y no es exclusivo de la raíz primaria.

Por todas las características anteriormente mencionadas las cactáceas del Desierto de Sonora proporcionan un modelo ideal para el estudio de los mecanismos que conllevan al agotamiento del RAM.

En el presente trabajo se utilizará a *Pachycereus pringlei* como modelo para el estudio de los mecanismos moleculares que conllevan al agotamiento del RAM. *P. pringlei*, comúnmente conocida como cardón, es planta arborescente y columnar y pertenece a la tribu *Pachycereeae* que se ubica dentro de la sub-familia Cactoideae. Entre las

cactáceas el cardón es de las especies que alcanzan mayor altura habiéndose reportado especímenes de hasta 19 metros; en el panel A de la figura 9 se muestra una planta adulta de *P. pringlei*.

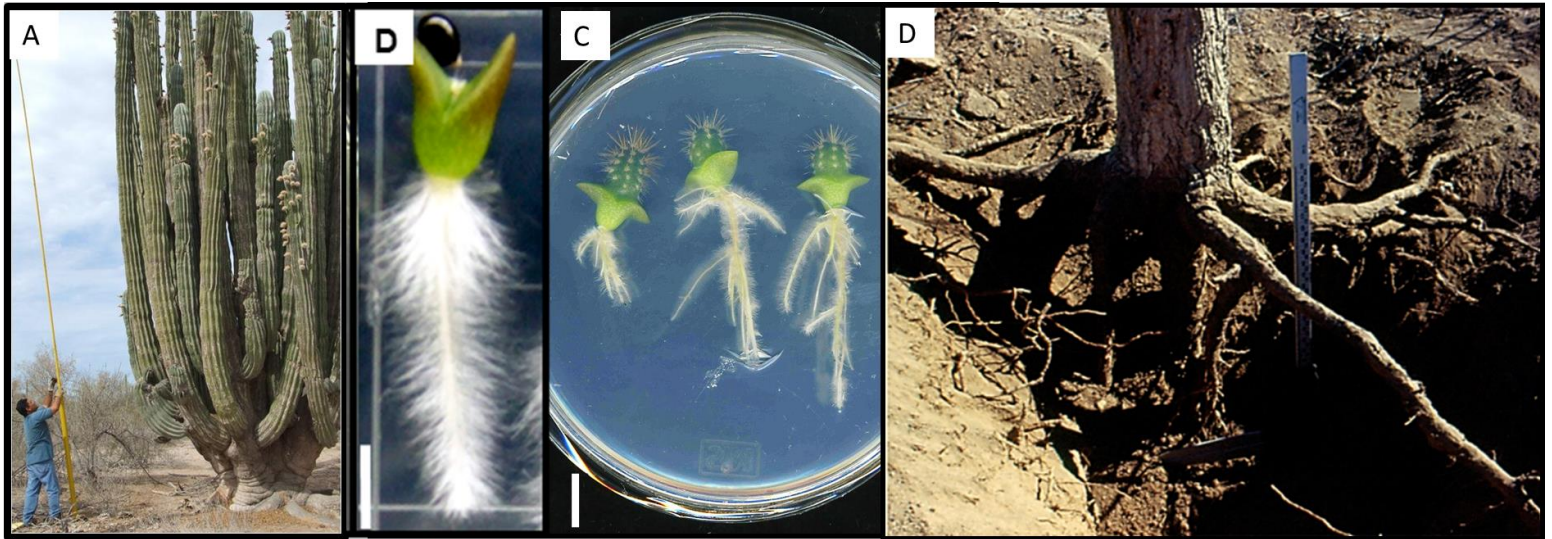


Fig. 9. *Pachycereus pringlei*. A) Planta adulta (Fotografía: Svetlana Shishkova). B) plántula en la que la raíz primaria ha llegado a la etapa terminal del desarrollo; barra, 5 mm. C) Plantas de *Pachycereus pringlei* de 3 meses, crecidas en medio LS 0.2X sin sacarosa; barra, 1 cm. D) Sistema radical de una planta adulta joven (Niklas et al., 2002).

Después de que las semillas germinan, la raíz primaria de las plántulas de *P. pringlei* crece alrededor de ocho días y alcanza una longitud de hasta 46 mm (Dubrovsky & Gomez-Lomelí, 2003). En la figura 9B se muestra una plántula de *P. pringlei* en la que la raíz primaria ha alcanzado la etapa terminal del desarrollo, es decir que el RAM se agotó, todas las células de la punta de la raíz se han diferenciado y el crecimiento de la raíz primaria ha llegado a su fin. Entre las especies estudiadas de cactáceas del Desierto de Sonora la raíz primaria de *P. pringlei* es la que presenta el periodo de crecimiento más prolongado. Esto nos da la ventaja de poder examinar con mayor detalle el proceso de transición que se da hasta el agotamiento del RAM, es por esta razón que elegimos trabajar con esta especie.

En la figura 9C se muestran plántulas de *P. pringlei* de 3 meses en las que el sistema radical es compacto y ramificado, este tipo de sistema radical se forma, como se explicaba antes, debido al crecimiento determinado de la raíz primaria y raíces laterales. Mientras que la planta va creciendo crece aparecen más raíces laterales (todas con crecimiento determinado) que van conformando el sistema radical. Dichas raíces presentaran crecimiento secundario, es decir que las raíces incrementan su diámetro debido a la formación de nuevas capas concéntricas de células. Con el crecimiento secundario los límites entre la raíz primaria y raíces laterales de órdenes subsecuentes que formarán parte de raíz principal se van perdiendo conforme va creciendo el sistema radical dando lugar a estructuras radicales más gruesas y alargadas como puede verse en la figura 9D donde se muestra el sistema radical de una planta adulta joven.



En estas cactáceas se llama raíz dominante a la raíz vertical que ancla la planta al suelo y que se ha formado como resultado de la fusión de raíces laterales de diferentes órdenes (Mauseth, 2009; citado por Dubrovsky & Shishkova, 2013). En la fotografía de la figura 9D la raíz dominante tiene alrededor de medio metro de profundidad en el sistema radical de una planta adulta de *Pachycereus pringlei* de aproximadamente 6 metros de altura.

### **1.7 Distribución del crecimiento determinado de la raíz primaria en la familia Cactaceae**

La familia Cactaceae está constituida por un amplio grupo de especies (de 1500 a 1800) en su mayoría altamente especializadas y adaptadas a climas áridos. Esta familia está integrada por cuatro subfamilias: Cactoideae, Opuntioideae, Maihuenioideae y Pereskioideae (Hunt *et al.*, 2006). La subfamilia Cactoideae, a la que pertenece *P. pringlei*, es la más grande y se divide en siete tribus, las cuales incluyen alrededor de 110 géneros (Hunt *et al.*, 2006).

Considerando el crecimiento determinado de la raíz primaria como un rasgo que ha permitido a las cactáceas establecerse en climas áridos, en Shishkova *et al.* (2013) se presenta un estudio acerca de la distribución de este rasgo a lo largo de la familia Cactaceae con el fin de conocer el grado de conservación de este carácter (Fig. 10).

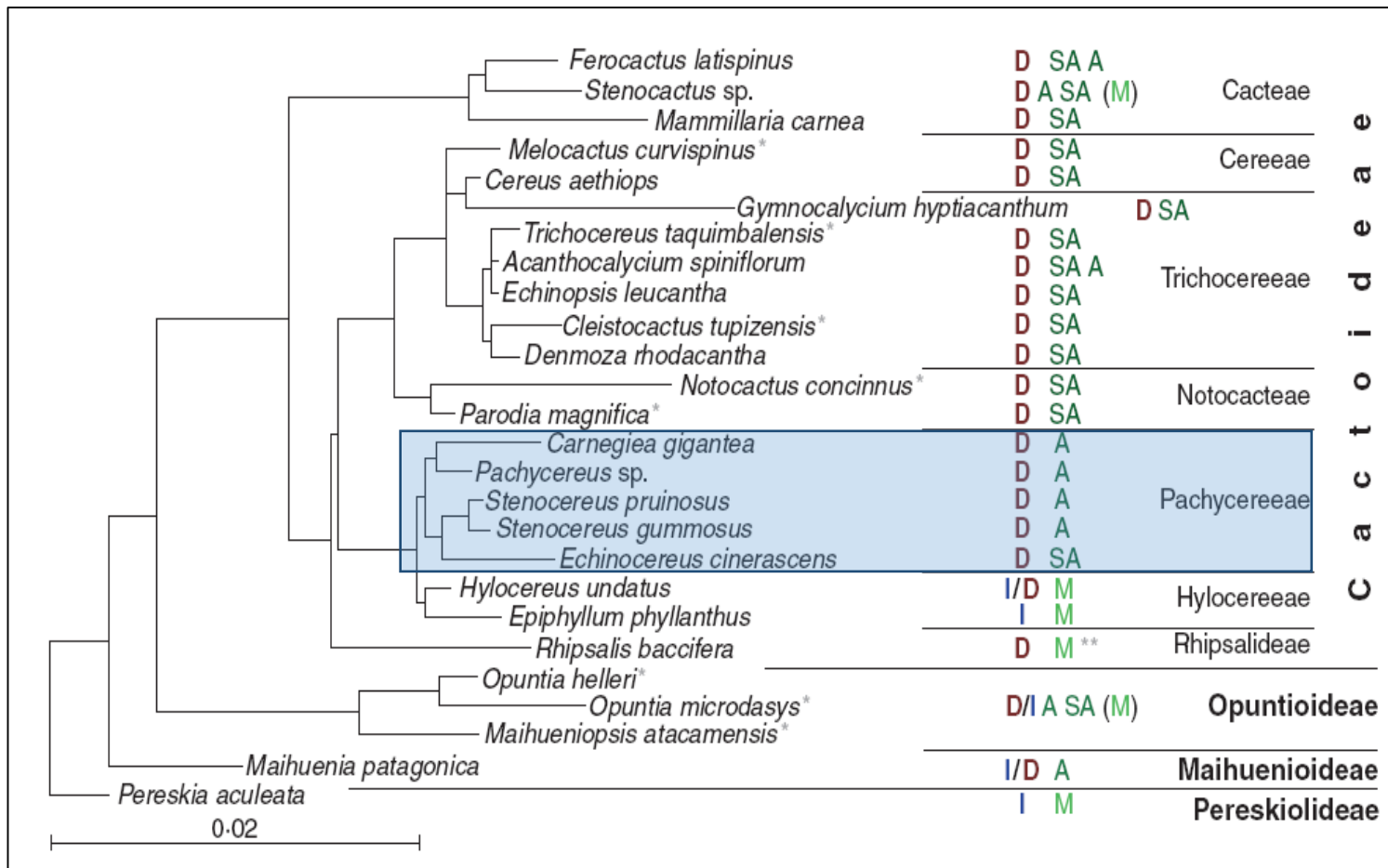


Fig. 10. Árbol filogenético de géneros de cactáceas analizados en Shishkova *et al.* (2013) en el que muestra el tipo de crecimiento de la raíz primaria: Determinado (D) o Indeterminado (I); y el hábitat de cada especie: Semiárido (SA), Árido (A) o Méjico (M). En azul se destaca la tribu a la que pertenece *P. pringlei*. Este árbol fue construido con base en las secuencias de los marcadores moleculares *trnL-trnF*, *trnK-matK* y *rpl16* que son comúnmente utilizados para análisis taxonómicos y que se encuentran en la base de datos GenBank.

\*Secuencias no disponibles para la especie estudiada, sino para otra especie del mismo género.

\*\* Habitat méjico de epifitas obligadas del género *Rhipsalis* que experimentan déficit hídrico.

En Shishkova *et al.* (2013), se realizó el análisis detallado del patrón de crecimiento de la raíz primaria de las especies pertenecientes a las siete tribus de la subfamilia Cactoideae (Fig. 10). De las cinco tribus cuyo hábitat se encuentra en regiones áridas y semiáridas todas las plántulas de las especies analizadas mostraron crecimiento determinado de la raíz primaria. También mostraron crecimiento determinado las especies de la tribu *Rhipsalideae*, las cuales experimentan déficit hídrico por ser epífitas de zonas húmedas. Solamente las plántulas pertenecientes a dos especies analizadas de la tribu Hylocereeae, que son epífitas de bosques húmedos, tuvieron crecimiento tanto determinado como indeterminado de la raíz primaria.

Es interesante que para una de estas dos especies, *Epiphyllum phyllanthus* (Shishkova *et al.* 2013), utilizando semillas de una planta cultivada, solamente se obtuvieron plántulas con crecimiento indeterminado de la raíz primaria, mientras que en el trabajo

reportado por Almeida *et al.* (2013), se germinaron semillas de varias plantas colectadas en la naturaleza y se encontraron plántulas con crecimiento determinado de la raíz primaria. Lo anterior demostró que todas las especies analizadas de la subfamilia Cactoideae que habitan zonas áridas o semiáridas presentan crecimiento determinado de la raíz primaria, aunque las especies epífitas del tribu Hylocereeae también pueden tener crecimiento indeterminado.

Para las tres familias restantes solamente se llevó a cabo un análisis preliminar del tipo de crecimiento de la raíz primaria. La mayoría de las plantas de las especies analizadas pertenecientes a la subfamilia Opuntioideae presentaron crecimiento determinado. Al mismo tiempo, las tres especies analizadas de la subfamilia Pereskioideae, que habitan zonas mésicas, y la mayoría de las plántulas de la Maihuenioideae mostraron crecimiento indeterminado de la raíz primaria.

Con base en los resultados del análisis del patrón de crecimiento de la raíz primaria de las 4 subfamilias, de manera preliminar se puede hipotetizar que el patrón de crecimiento determinado se fijó después de la separación de los linajes Cactoideae/Opuntioideae y Maihuenioideae/Pereskioideae.

La conservación del patrón de crecimiento determinado en diferentes especies de cactáceas indica que este tipo de crecimiento es controlado por algún mecanismo molecular que conlleva al agotamiento del RAM.

## **Mutantes con crecimiento determinado de la raíz identificadas en *A. thaliana***

### **1.8 Identificación de los factores de transcripción *SHORT ROOT* y *SCARECROW***

Existen algunas mutantes de *A. thaliana* que muestran características de agotamiento del meristemo y, como consecuencia tienen crecimiento determinado de la raíz primaria. Es por esto que este tipo de mutantes representan una herramienta valiosa para el estudio de los mecanismos del mantenimiento y agotamiento del RAM. Muchas de estas mutaciones fueron aisladas mediante un escrutinio de plantas con raíces cortas en comparación con las plantas de tipo silvestre, como es el caso de las mutantes *short root* (*shr*) y *scarecrow* (*scr*) (Fig. 11).

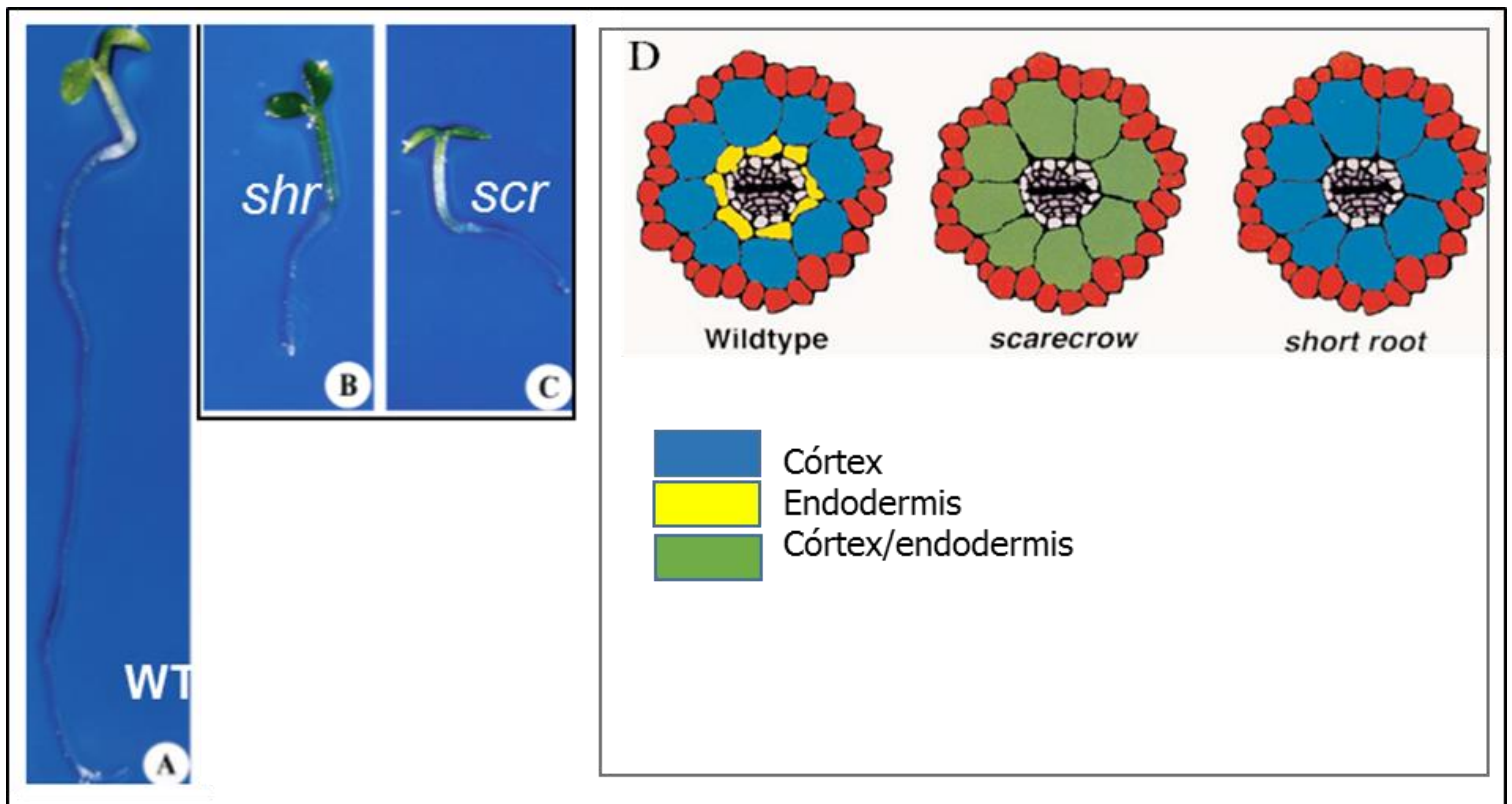


Fig. 11. Fenotipo de plántulas mutantes con crecimiento determinado de la raíz de *A. thaliana*. A) Planta de tipo silvestre (WT); B) mutante *short root* (*shr*) y C) mutante *scarecrow* (*scr*). Plantas a los 7 días después de la germinación (d.d.g.). A la misma escala de ampliación. (Modificado de Scheres *et al.*, 1995). D) Organización radial de la raíz de *A. thaliana* en la planta de tipo silvestre (WT) y mutante *scr* y *shr* (Scheres, Benfey, & Dolan, 2002).

En las mutantes *shr* la raíz detiene su crecimiento a los 5 días de haber emergido la radícula y el ápice de la raíz se cubre de pelos radicales. En la mutante *scr* ocurre lo mismo a los 8 días de haber emergido la radícula (Scheres *et al.*, 1995);(Benfey *et al.*, 1993).

Cuando se analizó a detalle la estructura de la raíz primaria de las mutantes *shr* y *scr* se encontró que ambas están afectadas también en el establecimiento del patrón radial. En estas mutantes no se lleva a cabo la división celular longitudinal de córtex/endodermis, por lo que en lugar de dos capas, una de córtex y otra de endodermis, se forma una sola capa celular. En la mutante *shr* esta capa celular presenta características de córtex (Benfey *et al.*, 1993) (Fig. 11 D), mientras que en la mutante *scr* esta capa celular presenta características tanto de córtex como de endodermis (Scheres *et al.*, 1995).

El análisis de las secuencias de estos genes mostró que las proteínas SHR y SCR son factores de transcripción de la familia GRAS. Estudios posteriores revelaron que *SHR* se expresa en el tejido vascular, y una vez traducida la proteína ésta se mueve desde el estele a la capa de células adyacente y entra al núcleo. En el centro quiescente SHR promueve la expresión de *SCR*. *SCR* se expresa en las células precursoras del CQ, en las células iniciales de córtex/endodermis y en la capa de endodermis. La proteína SCR

es requerida de manera autónoma por las células para establecer y mantener la identidad del CQ y las células que lo rodean (Sabatini *et al.*, 2003).

## **1.9 El papel de los factores de transcripción *PLETHORA* en *A. thaliana***

Los factores de transcripción PLT son específicos de plantas, se expresan en todos los tejidos en división y tienen un papel central en los procesos del desarrollo, por lo que podrían ser parte del mecanismo de control molecular que conlleva al agotamiento del RAM.

Las proteínas PLT pertenecen a la súper-familia AP2/ERF (APETALA2/ETHYLENE-RESPONSIVE ELEMENT BINDING FACTOR). Esta familia se caracteriza por contener el dominio AP2 de unión a DNA, anotado en la base de datos Pfam como PF00847 (pfam.xfam.org/). La súper-familia AP2/ERF se divide en dos clases de acuerdo al número de dominios AP2 que contienen, proteínas de la clase ERF-*like* que contienen un solo dominio AP2 y las de la clase denominada AP2 contienen dos dominios AP2 (Kim *et al.*, 2006).

Dentro de la clase AP2 se encuentra la subclase AIL (AINTEGUMENTA LIKE) que se caracteriza por presentar dos dominios AP2 de unión a DNA y señales de localización nuclear. Las proteínas PLT cumplen con dichas características por lo que se clasifican como miembros de la subclase AIL. En *A. thaliana* son ocho los miembros de esta subclase: ANT, AIL1, y los codificados por los genes PLT: PLT1, PLT2, PLT3/AIL6, BBM/PLT4, PLT5/AIL5 y PLT7/ AIL7 (revisado en Horstman *et al.*, 2013).

Los genes *PLETHORA* fueron encontrados usando la estrategia de "promoter trap" para identificar genes con expresión en el CQ (Aida *et al.*, 2004). En primer lugar se encontraron los genes que codifican para las proteínas PLT1 y PLT2 (Aida *et al.*, 2004), que presentan una similitud entre sí del 76% y una identidad del 80%. Posteriormente por homología se identificaron los genes *PLT3*, *BBM*, *PLT5* y *PLT7* (Galinha *et al.*, 2007).

### **1. 9.1 La función de los genes *PLT* en el RAM en *A. thaliana***

La función de los genes *PLT* se ha estudiado principalmente en *A. thaliana*. Los genes parcialmente redundantes *PLT1*, *PLT2*, *PLT3* y *BBM* son necesarios para el establecimiento del RAM y por ende para el desarrollo de la raíz. Las mutantes de pérdida de función de los genes *PLT* evidencian dichas funciones. Las mutantes simples tanto de *plt1* como de *plt2*, tienen raíz un poco más corta respecto a las plantas de tipo silvestre, presentan más células y filas de células en la columela. La mutante doble *plt1 plt2* presenta crecimiento determinado de la raíz primaria y raíces laterales (Fig. 12 B, E) (Aida *et al.*, 2004), mientras que la triple mutante *plt1 plt2 plt3* no tiene raíz.

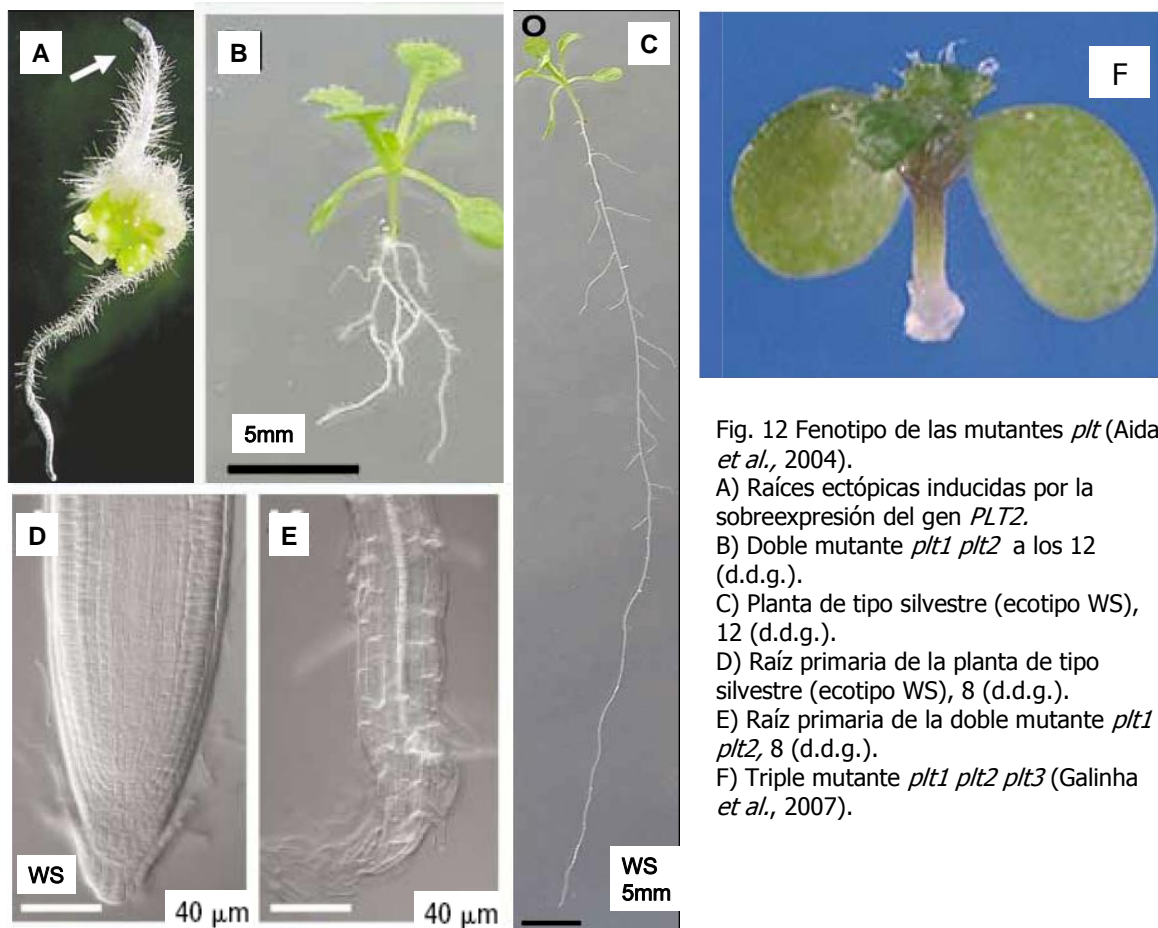


Fig. 12 Fenotipo de las mutantes *plt* (Aida *et al.*, 2004).  
 A) Raíces ectópicas inducidas por la sobreexpresión del gen *PLT2*.  
 B) Doble mutante *plt1 plt2* a los 12 (d.d.g.).  
 C) Planta de tipo silvestre (ecotipo WS), 12 (d.d.g.).  
 D) Raíz primaria de la planta de tipo silvestre (ecotipo WS), 8 (d.d.g.).  
 E) Raíz primaria de la doble mutante *plt1 plt2*, 8 (d.d.g.).  
 F) Triple mutante *plt1 plt2 plt3* (Galinha *et al.*, 2007).

La sobreexpresión inducida de *PLT2* promueve la formación de raíces ectópicas (Fig.12 A) (Aida *et al.*, 2004); es decir que en dichas plantas mutantes se induce la formación de RAM ectópicos. En la fig. 12A se muestra que las raíces ectópicas pueden aparecer en el brote, lo que indica que se estableció un RAM en lugar del SAM como consecuencia de la expresión ectópica de *PLT*. Lo anterior termina por probar la función de los genes *PLT* en el establecimiento del RAM.

Cuando se adicionan copias extras del gen *PLT2* bajo su propio promotor, se produce un meristemo más alargado, en el cuál el dominio de divisiones más frecuentes se mueve a la parte más basal (esto es, más cercano a la zona de elongación) del RAM (Mähönen *et al.*, 2014).

Una alta expresión de *PLT2* disminuye la tasa de división celular. Ejemplo de ello es lo que ocurre en el nicho de las células iniciales en la planta de tipo silvestre, en donde se observa una alta concentración de *PLT2* que correlaciona con una tasa de división celular muy baja. Por otro lado, un nivel medio de expresión de *PLT2* aumenta la tasa de división celular. Esto es lo que sucede en el resto de las células del RAM en la planta de tipo silvestre, donde las divisiones celulares se dan de manera frecuente. Se ha demostrado además que la expresión de *PLT2* tiene la capacidad de inhibir la expansión celular y la diferenciación (Mähönen *et al.*, 2014).

Durante la etapa embrionaria los genes *PLT* también juegan un papel muy importante. Las mutantes dobles y triples de pérdida de función de los genes *PLT1*, *PLT2*, *PLT3* y *BBM* muestran anormalidades embrionarias. El desarrollo de las mutantes *plt2 bbm* no llega más allá de la etapa embrionaria (revisado en Horstma *et al.*, 2013). Además, el hecho de que la sobre-expresión de *BBM* o *PLT5* induzca la formación ectópica de embriones en meristemo y cotiledones, constituye una evidencia de la importancia de los genes *PLT* en la embriogénesis (Bouttiler *et al.*, 2002; Passarinho *et al.*, 2008). Lo anterior sugiere que durante el desarrollo temprano del embrión los genes *BBM* y *PLT5* actúan como reguladores maestros, lo cual refleja el papel de los genes *PLT* en el mantenimiento de las células madre.

### **1.9.2 Gradientes de expresión de los genes *PLT* en el RAM de *A. thaliana***

Los niveles de los transcritos y las proteínas *PLT1*, *PLT2*, *PLT3* y *BBM/PLT4* forman gradientes a lo largo del ápice de la raíz primaria de *A. thaliana*; en la figura 13 se muestra la actividad en la raíz primaria de los promotores de los genes *PLT* y los genes *PLT* fusionados a proteínas fluorescentes. El nicho de células iniciales está definido por la presencia de altos niveles de dichos factores de transcripción. La formación del gradiente de proteínas *PLT* regula el mantenimiento de las células iniciales, así como el límite del meristemo (Galinha *et al.*, 2007). Se sugiere que la disminución de la concentración de *PLT* a lo largo del gradiente determina la transición hacia la diferenciación (Mähönen *et al.*, 2014). Esta correlación entre los niveles de proteínas *PLT* con los estados de desarrollo define las zonas de la raíz; esta es la razón por la cual se considera a las proteínas *PLT* reguladores maestros dependientes de la dosis en el desarrollo de la raíz (Galinha *et al.*, 2007).

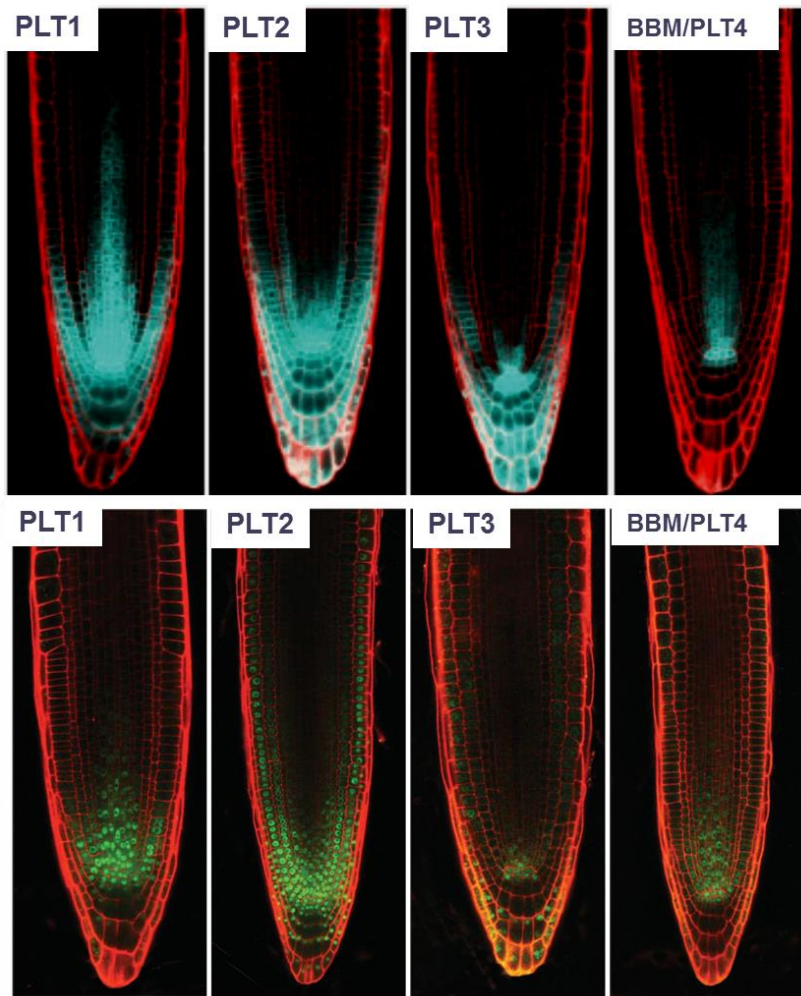


Fig. 13 Panel superior, actividad de los promotores de los genes *PLT*, secuencia codificadora de la proteína cian fluorescente (CFP) fusionada al promotor completo de los genes *PLT1*, *PLT2*, *PLT3*, *BBM/PLT4*. Panel inferior, proteína fluorescente YFP fusionada a la secuencia genómica de *PLT1*, *PLT2*, *PLT3*, *BBM/PLT4* (Galinha *et al.*, 2007).

La transcripción de los genes *PLT* es estimulada por auxinas y requiere de los factores de respuesta a auxinas ARF5, ARF7 y ARF19, que se unen a los elementos de respuesta a auxinas (Aida *et al.*, 2004). El gradiente de expresión de las proteínas PLT es similar al gradiente que muestra la actividad del promotor DR5 inducible por auxina. Sin embargo, aunque el gradiente de expresión de *PLT* está influenciado por la distribución de las auxinas en el RAM, este gradiente no representa una lectura directa del gradiente de auxinas. Prueba de ello son algunas mutantes en las que está disminuido el gradiente de respuesta a auxinas pero no está alterado el gradiente de *PLT2* como son *aux1*, *ein2* y *gnom* (Fischer *et al.*, 2006). La inducción de los *PLT* mediante la vía de auxinas no es de respuesta inmediata y requiere de un elevado nivel de auxinas (Mähönen *et al.*, 2014).

Existe evidencia que indica que *PLT2* se transcribe en la zona en la que se encuentra el nicho de células iniciales, pero el transcrito de *PLT2* o la proteína PLT2 migran hacia las células aledañas (Mähönen *et al.*, 2014). Otro de los factores que se suman a la



formación del gradiente de proteínas es la dilución de la concentración de dichas proteínas durante el crecimiento por distribución de célula a célula.

La amplitud del gradiente de PLT depende directamente de la estabilidad de las mismas proteínas. Al fusionar el promotor de *PLT2* a la secuencia codificante de una proteína estable como la Histona y una proteína fluorescente se observa un aumento en la amplitud del gradiente; mientras que al fusionar promotor de *PLT2* a la secuencia codificante de una proteína lábil como la CiclinaB1;1 la amplitud del gradiente disminuye. De lo anterior se deduce que la vida media de las proteínas PLT es un factor que controla la amplitud del gradiente (Mähönen *et al.*, 2014).

### 1.9.3 Control de la expresión de *PLT*

Los péptidos de la familia RGF (Root Growth Factor) son péptidos sulfatados de 13 amino ácidos que regulan la expresión de *PLT1* y *PLT2* tanto a nivel transcripcional como a nivel post-transcripcional. Los RGFs tienen un papel redundante en la regulación de la actividad del meristemo. En las plantas de *A. thaliana* de tipo silvestre, los RGFs se distribuyen a lo largo del RAM y actúan directamente en células individuales para definir el nivel de expresión de *PLT* (Mähönen *et al.*, 2014). Se ha sugerido que una de las formas en las que los RGFs logran este control es mediante la estabilización de las proteínas PLT: como se mencionó antes, la estabilidad de las proteínas PLT define la amplitud del gradiente de las mismas.

Se sabe que tanto la división como la expansión celular están influenciadas por los niveles de auxinas (Sabatini *et al.*, 1999); las auxinas influyen dichos procesos de manera inmediata e independiente del control de *PLT*. Esto permite que exista una respuesta inmediata ante cambios ambientales como sucede con los tropismos. Por otra parte, la respuesta tardía a cambios en auxinas que influyen la distribución de *PLT*, confiere estabilidad a los procesos del desarrollo al no alterar de manera inmediata el gradiente de *PLT* (Mähönen *et al.*, 2014).

En cuanto al control de *PLT* a nivel epigenético se sabe que el complejo GCN5 histona acetil-transferasa incrementa la expresión de *PLT1* y *PLT2* durante el desarrollo post-embriónico. Este complejo regula además la expresión de las ciclinas CycB1;1 CycB1;3 en el RAM (Kornet & Scheres, 2009). Las modificaciones en la cromatina tienen un importante papel en control de la transcripción. *GCN5* está implicado en procesos de desarrollo, por lo que la disrupción de este gen genera efectos pleiotrópicos en la planta. En la parte aérea, este gen regula el funcionamiento del SAM controlando la expresión de *WUS* (Bertrand *et al.*, 2003), mientras que en la raíz regula el funcionamiento del RAM a través de los genes *PLT* (Kornet & Scheres, 2009).

### 1.9.4 Los genes *PLT* en el SAM

Los genes *ANT*, *PLT3*, *PLT5* y *PLT7* se expresan en el brote. De forma similar a lo que sucede en la raíz, en la parte aérea *ANT*, *PLT3*, *PLT5* y *PLT7* tienen diferentes dominios de expresión, pero estos se sobrelapan. Estos genes se expresan en las

capas epidérmica y subepidérmica de los meristemas, tanto vegetativos como inflorescentes (Krizek 2009; Prasad *et al.*, 2011).

*ANT*, *PLT3* y *PLT7* son necesarios para el mantenimiento del SAM. *ANT* y *PLT3* tienen papeles parcialmente redundantes en aspectos del desarrollo floral y del brote, incluyendo el crecimiento de órganos e iniciación del meristemo floral (Krizek *et al.*, 2009). *PLT7* guarda una alta homología con *PLT3*, por lo que se esperaba que participara también en dichos procesos. En la triple mutante *ant plt3 plt7* el SAM primario se agota después de producir entre tres y cinco hojas. En este caso se sabe que estos genes no son totalmente redundantes, pues tienen diferentes dominios de expresión dentro del SAM y exhiben diferentes interacciones genéticas con miembros de dos de las vías requeridas en la actividad meristemática: *WUS* y *STM* (*shoot meristem less*). Sin embargo la acción conjunta de estos tres genes es requerida para promover la división celular y reprimir la diferenciación, lo que mantiene la estabilidad del SAM (Mudunkotge & Krizek, 2012). Estos tres genes (*ANT*, *PLT3* y *PLT7*) juegan en el SAM el mismo papel que *PLT1*, *PLT2* y *BBM* en el RAM al mantener la estabilidad del meristemo y controlar la tasa de diferenciación celular. Esta es una clara evidencia de la existencia de un módulo regulatorio para el mantenimiento de células madre que es común en el SAM y RAM. Sin embargo, una de las diferencias entre ambos sistemas es que en la raíz los genes *PLT* participan en el establecimiento del RAM durante la embriogénesis (Aida *et al.* 2004), mientras que en la parte aérea los genes *PLT* no son requeridos sino hasta después de la germinación para mantener el funcionamiento del SAM (Mudunkotge & Krizek, 2012).

### 1.9.5 La participación de genes *PLT* en filotaxis y rizotaxis

La filotaxis es la disposición que presentan las hojas alrededor del tallo y es característica de cada especie. La expresión de los genes *PLT3*, *PLT5* y *PLT7* en la zona central del SAM es necesaria para estabilizar el proceso de filotaxis espiral característico de *Arabidopsis*. La triple mutante *plt3 plt5 plt7* presenta alteraciones en filotaxis, es decir que se altera en sitio en que debería ocurrir la iniciación de órganos laterales. La evidencia sugiere que *PLT* controla la abundancia de auxinas al promover la biosíntesis de auxinas en la zona central del SAM. Específicamente *PLT* regula la expresión de los genes *YUCCA1* y *YUCCA4* que participan en la biosíntesis de auxinas y que mediante esta vía es que se estabiliza la filotaxis (Pinon *et al.*, 2013).

Este mismo grupo de genes (*PLT3*, *PLT5* y *PLT7*) en la raíz se expresa en las células del periciclo que dan origen a las células fundadoras de los primordios de raíces laterales. Los dominios de expresión de estos genes se sobrelapan en los primordios de raíces laterales. El estudio de las mutantes de pérdida de función de estos genes indica que dirigen la rizotaxis (la disposición o acomodo de las raíces). En la triple mutante *plt3 plt5 plt7* se forman clusters de raíces, a diferencia de las plantas de tipo silvestre en que típicamente no se forman primordios a menos de 300 µm uno de otro (Hofhuis *et al.*, 2013).

En la triple mutante *plt3 plt5 plt7* la emergencia y morfología de los primordios de raíces laterales también está severamente afectada. En algunas de las líneas de

complementación con *PLT5* o *PLT7* se restaura la morfología y emergencia de los primordios, pero la rizotaxis no se restablece por completo. La acción de estos genes es necesaria para estabilizar la rizotaxis y la emergencia de raíces laterales, pero controlan ambos procesos de forma independiente. Por otro lado, los genes *PLT1*, *PLT2* y *BBM/PLT4* se expresan cuando los primordios ya se han formado (Hofhuis *et al.*, 2013). Los datos aquí señalados revelan que la actividad local conjunta de *PLT3*, *PLT5* y *PLT7* define la posición espacial de nuevos órganos laterales en el brote y en la raíz.

### 1.9.6 Los genes blanco de *PLT*

Muchas de las interacciones génicas que sostienen los genes *PLT* se han enumerado en este texto. Sin embargo aún se conoce poco acerca de los blancos sobre los cuales actúan directamente los genes *PLT*. Únicamente se han identificado los genes blanco de BBM. En Passarinho *et al.*, 2008 se analizaron por microarreglos los genes que en plántulas de *Arabidopsis* de 8 días de edad mostraron expresión diferencial tras inducir la expresión de la proteína BBM de *Brassica napus* por 8 horas. Se encontraron 2449 genes expresados aproximadamente la mitad se indujeron y la otra mitad se reprimieron. Se estudiaron 43 genes cuya expresión aumentaba al menos 4 veces, la mayoría se expresan en los embriones somáticos inducidos por BBM. Entre estos genes se encontró al mismo BBM, lo que indica que este gen autorregula su expresión mediante retroalimentación positiva. El resto de los genes participan en procesos de transcripción, ubiquitinación o codifican proteínas localizadas en la pared celular o involucradas en su modificación.

ADF9 (*ACTIN DEPOLYMERIZING FACTOR 9*) es otro de los genes preferencialmente expresados. Este gen codifica una cofilina que se une a F-actina y G-actina separando filamentos de actina y aumentando su despolarización (Carlier *et al.*, 1997). Al inducirse la expresión de BBM y usando un marcador de actina (GFP:FIMBRIN ACTIN DOMAIN 2) se observó que los haces de actina cambiaron completamente su orientación después de 1.5 minutos (Passarinho *et al.*, 2008). Estos datos vinculan directamente a BBM con la reorganización del citoesqueleto que es necesaria durante la proliferación celular. Si bien este puede ser un blanco directo de BBM, el número de genes que son activados o inhibidos por BBM indica que existe una amplia red de genes que participan en la maquinaria celular para el mantenimiento de células madre que no se ha caracterizado aún.

Para concluir la parte introductoria cabe hacer notar que al comparar el sistema radical de la doble mutante *plt1 plt2* con el de las cactáceas del Desierto de Sonora, estos son muy similares ya que presentan crecimiento determinado de la raíz primaria y de raíces laterales (Fig. 9C y 12 B). Por lo anterior y considerando las funciones descritas para estos genes en *A. thaliana* sería interesante analizar el rol de los homólogos de estos genes en cactáceas con crecimiento determinado.

## 2. Antecedentes

El único grupo de angiospermas en el que se ha reportado crecimiento determinado de la raíz *primaria* en plantas de tipo silvestre y en todas las condiciones de crecimiento son las cactáceas del Desierto de Sonora (Dubrovsky, 1997). Sin embargo, hasta la fecha no se cuenta con la secuencia del genoma de ninguna especie de la familia Cactaceae.

Al no existir suficiente información publicada sobre genes de cactáceas, mucho menos de su análisis funcional, en nuestro grupo de trabajo se siguieron diferentes estrategias (descritas en los siguientes dos apartados del texto) para obtener secuencias de genes de cactáceas. Cabe señalar que para ninguna de las especies que se utilizaron existían secuencias reportadas de genes. La búsqueda de se enfocó en encontrar grupos de genes involucrados en el crecimiento determinado de la raíz en cactáceas.

### 2.1 Generación de bibliotecas sustractivas de expresión génica

Inicialmente, se construyeron bibliotecas sustractivas de los genes que se expresan preferencialmente en las etapas inicial (cuando el RAM aún existe) y terminal (cuando el RAM ya se ha agotado) del desarrollo determinado de la raíz primaria de la pitaya agria *S. gummosus* (Shishkova *et al.*, 2013) (véase fig.7). De la biblioteca de la etapa inicial se secuenciaron 2000 clonas correspondientes a fragmentos de cDNAs. Las secuencias obtenidas se ensamblaron en 689 "unigenes", entre los cuales se encontraron solamente tres posibles factores de transcripción. Ninguno de los tres relacionado con los genes de los que se ha reportado que participan en el mantenimiento del RAM.

### 2.2 mRNA-seq: transcriptomas de puntas de raíces de *P. pringlei*

El desarrollo de las técnicas de secuenciación de nueva generación, o secuenciación masiva, así como del ensamblaje *de novo* de millones de las secuencias cortas producidas, empezó a hacer posible el obtener datos del transcriptoma de distintas especies sin tener la secuencia de su genoma.

En el presente trabajo se utilizó el transcriptoma de puntas de raíces primarias de *P. pringlei*. Estos datos se obtuvieron mediante el método de secuenciación masiva llamada mRNA-seq. El mRNA-seq es una técnica de secuenciación de nueva generación que incluye la síntesis de cDNA a partir de transcritos de mRNA poliadenilados y permite mapear las lecturas cortas sobre un genoma secuenciado o ensamblarlas *de novo* en los *contigs*, si el genoma no está secuenciado, y posteriormente cuantificar la frecuencia numérica de secuencias que forman parte de un *contig* dentro de cada una de las muestras, es decir, estimar la expresión diferencial de los genes en las muestras correspondientes.

La muestra biológica utilizada para generar esta información comprendía 1 mm de la punta de raíces primarias de las etapas inicial y terminal del desarrollo de la raíz (Fig. 14).

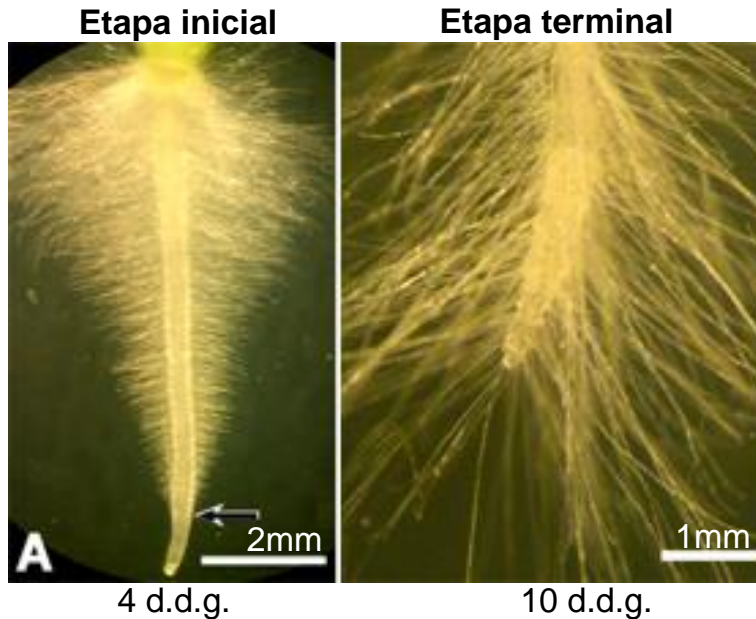


Fig. 14. Raíz primaria de *P. pringlei*. A) Raíz en la etapa inicial del desarrollo, la flecha señala los pelos radicales más cercanos a la punta de la raíz. B) Raíz en la etapa terminal del desarrollo (Dubrovsky y Gómez-Lomelí, 2003).

La secuenciación en el Illumina Genome Analyzer (GA) II, ensamblaje y anotación de *contigs* y la estimación de la expresión diferencial se realizaron como parte de un proyecto de colaboración entre nuestro laboratorio, el Genomic Center de Universidad de California-Davis y la compañía CLC Bio (actualmente CLC Bio – Qiagen).

En un primer ensamblaje de esta secuenciación las aproximadamente 20 millones de secuencias de 85 nucleótidos se ensamblaron en 31,224 *contigs* con longitud mayor a 300 nt usando la función “*de novo assembler*” del programa CLC Genomic Workbench ([www.clcbio.com](http://www.clcbio.com)). En ese momento tanto las técnicas de secuenciación de nueva generación, como los métodos de ensamblaje *de novo* apenas empezaban a desarrollarse, por lo que era de esperar que existieran muchos errores y que hasta 10% de los *contigs* pudieran ser quiméricos.

Los *contigs* fueron anotados por BLASTx usando la base datos de proteínas RefSeq y utilizamos el prefijo *Pp* para denominar a los posibles homólogos de *P. pringlei*. Las lecturas mRNA-seq de cada etapa mapeadas con el mismo programa CLC Genomic Workbench (GW) sobre los *contigs* ensamblados, permitieron estimar los niveles virtuales de expresión génica. Se contó el número total de lecturas que mapearon a cada *contig* y este número se normalizó usando el método de RPKMs (Reads Per Kilobase per Million Mapped Reads, Mortazavi *et al.* 2008) para contender con sesgos

debidos a la variación de tamaño entre los diferentes *contigs* y la cantidad de lecturas mapeadas.

En el primer ensamblaje se encontraron tres posibles homólogos a los genes *PLT*; al analizar sus niveles virtuales de expresión, se encontró que estos tienen un nivel de acumulación de 8 a 11 veces mayor en la etapa inicial respecto a la etapa terminal (Fig. 15; López-Valle, 2011, Tesis de licenciatura).

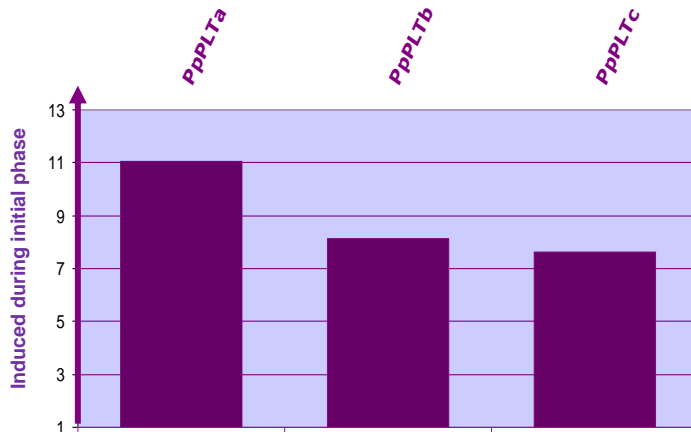


Fig. 15. Expresión virtual de los genes *PpPLTa*, *PpPLTb* y *PpPLTc* analizados por el programa CLC GW. En el eje de las ordenadas, se indica el número de veces de diferencia de valores RPKM en la etapa inicial respecto a la etapa terminal López-Valle, 2011.

Posteriormente, aprovechando el rápido avance en programación para el análisis de datos de RNA-seq, se realizó un segundo ensamblaje de las mismas secuencias de 85 nucleótidos, en el cual se descartaron las secuencias que estuvieran repetidas. Esta vez se utilizó un "plug in" del programa CLC GW, específicamente,  $\beta$  assembler", el cual permite modular diferentes parámetros, para realizar el ensamblaje, por ejemplo, el tamaño de palabra (word size) para evitar varios de los errores posibles en el ensamblaje. Como resultado de este segundo ensamblaje se generaron 26,631 *contigs* con longitud mayor a 300 nt. Es decir, menos *contigs*, pero con un mayor tamaño promedio y un número considerablemente menor de *contigs* quiméricos. En este trabajo de tesis se utilizó este segundo ensamblaje de secuencias.

### 3. Justificación

De los reportes que existen sobre la caracterización de las mutantes de *A. thaliana* que muestran crecimiento determinado de raíz sabemos que los factores de transcripción PLT1, PLT2 y PLT3, son importantes en el mantenimiento del RAM (Aida *et al.*, 2004; Galinha *et al.*, 2007). Se ha propuesto que estos factores de transcripción funcionan como reguladores maestros del desarrollo al modular la expresión de grandes grupos de genes blanco. Sin embargo, hasta el momento existen muy pocos reportes sobre el análisis funcional de los genes *PLT* en otras especies vegetales.

El análisis de los genes *PLT* en *P. pringlei*, una especie con crecimiento determinado de la raíz, ayudará a entender los mecanismos de este tipo de crecimiento. Las cactáceas desérticas son el único grupo conocido de angiospermas en las que se ha observado crecimiento determinado de la raíz primaria como parte de su programa de desarrollo. Aún no se conoce el mecanismo que conduce al agotamiento del RAM en cactáceas.

El examinar los niveles de transcritos de *PLT* abrirá la posibilidad de analizar el papel que juegan los genes *PLT* en las cactáceas. Comparar el comportamiento que se ha reportado para estos genes en *A. thaliana* contra el de una planta de otro grupo taxonómico aclarará la participación de estos genes en el desarrollo de la raíz primaria. Además, contribuirá al conocimiento de los mecanismos del desarrollo de la raíz en diferentes especies vegetales y determinará su grado de conservación.

## 4. Hipótesis

En *Pachycereus pringlei* el RAM se agota a los pocos días de haber emergido la raíz primaria. Se sabe que en *Arabidopsis thaliana* los genes *PLT* son importantes para el mantenimiento del RAM. Por lo tanto proponemos como hipótesis que: Los genes *PLT* pueden estar involucrados en el mecanismo de regulación que conlleva al agotamiento del RAM de la raíz primaria de *P. pringlei*.

## 5. Objetivo general

Analizar a nivel molecular los genes homólogos de los genes *PLETHORA* durante el desarrollo de la raíz primaria de *P. pringlei*.

### 5.1 Objetivos particulares:

- Identificar los contigs de los transcritos de genes que codifican para los posibles homólogos de los genes *PLETHORA*.
- Determinar la integridad de la secuencia codificante y la presencia o ausencia de secuencias no traducidas en los extremos 5' y 3' de cada uno de los transcritos ensamblados *PpPLT*.
- Identificar los posibles ortólogos de *PpPLT* a las proteínas homólogas de *A. thaliana* y de otras especies vegetales.
- Realizar una comparación evolutiva de los genes *PLT* entre *P. pringlei* y otras especies vegetales.
- Analizar la expresión de los genes *PpPLT* seleccionados en el desarrollo de la raíz primaria de *P. pringlei*.



## 6. Material y Métodos

### Glosario de definiciones bioinformáticas

**Alineamiento:** Método de comparación de cadenas de texto que se utiliza para encontrar regiones de homología entre secuencias de DNA, RNA o proteínas, al generar una matriz de  $n \times m$  columnas ( $n$ =longitud de la secuencia 1;  $m$ =longitud de la secuencia 2), en la cual se pueden insertar espacios o guiones para representar inserciones o deleciones de aminoácidos o bases con el fin de que las regiones similares queden empalmadas.

**BioEdit:** Programa para edición de alineamientos y análisis de secuencias.

**BLAST:** (*Basic Local Alignment Search Tool*) Programa bioinformático que busca secuencias similares (*hits*) a una secuencia problema (*query*) en una base de datos y calcula la significación estadística de las coincidencias (*matches*) entre ellas.

**BLASTn:** Algoritmo de BLAST que interroga a una base de datos de nucleótidos al utilizar una secuencia *query* también de nucleótidos.

**BLASTx:** Algoritmo de BLAST en el cual se interroga una base de datos de proteínas al utilizar la secuencia proteica que se obtiene al traducir en los seis marcos de lectura a la secuencia *query* que está en nucleótidos.

**Bootstrap:** Método estadístico para evaluar la confiabilidad de un árbol filogenético. A partir de cada secuencia se extraen  $n$  nucleótidos o aminoácidos y se realizan sustituciones aleatorias en esas subsecuencias, las cuales se utilizan para obtener un nuevo árbol. Este procedimiento se repite varias veces y se cuantifica el número de veces que una rama del árbol original aparece en el total de árboles generados mediante este método. La topología del árbol, o una fracción de esta, es más robusta conforme los valores de soporte de rama obtenidos por el método de *bootstrap* son más altos.

**Clustal Omega:** Programa bioinformático para realizar alineamientos múltiples de secuencias.

**Contig:** Durante el ensamblaje de genomas o transcriptomas, la secuencia que se obtiene a partir del empalme de dos o más lecturas.

**E-value:** Es uno de los parámetros que caracteriza a un alineamiento entre dos secuencias y, de manera general, representa a la probabilidad de encontrar por azar a la secuencia *hit* (o cualquier otra con la misma calificación en el alineamiento) al utilizar la secuencia *query*.

**GenBank:** Base de datos de secuencias nucleotídicas y de aminoácidos perteneciente al NIH (National Institutes of Health) de acceso público.

**Máxima verosimilitud (Maximum likelihood):** Método de inferencia filogenética. Asigna valores de probabilidad a las topologías posibles para un conjunto  $n$  de secuencias de acuerdo con modelos de sustitución variables para calcular la distancia entre dos secuencias. El resultado de este algoritmo es la topología con mayor probabilidad de ser correcta de acuerdo con los modelos de sustitución empleados. Es un método exhaustivo y, por lo tanto, computacionalmente demandante.

**Modelo de sustitución JTT (Jones-Taylor-Thornton):** Modelo empírico de sustitución de aminoácidos que está basado en la probabilidad de sustitución cuantificada en un conjunto de proteínas por Jones, Taylor y Thornton (1992).

**Neighbor Joining:** Método de reconstrucción filogenética que parte de una matriz de distancia entre secuencias y agrupa a las secuencias de modo tal que minimiza la longitud de las ramas. El resultado es un árbol no enraizado.

**Pfam:** Base de datos con una gran colección de familias proteicas cada una representada por alineamientos múltiples en los que se incluyen dominios y familias comunes.

**Phytozome:** Base de datos con acceso a los genomas actualmente de 54 especies vegetales que han sido secuenciados y anotados.

**RefSeq:** (Reference Sequence) es una colección de secuencias de DNA genómico, transcritos y proteínas no redundantes.

## 6.1 Análisis filogenéticos preliminares de genes *PLT*

### 6.1.1 Identificación de *contigs* con similitud a *PLT*

Usando el programa CLC Genomic Workbench ([www.clcbio.com](http://www.clcbio.com)), se realizó un BLASTx del gen *PLT2* de *A. thaliana* contra la base de datos de las secuencias del segundo ensamblaje antes descrito para encontrar *contigs* con similitud a *PLT*. Usando la función BLASTx ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)) con la base de datos de proteínas RefSeq de GenBank se identificaron los marcos de lectura de los *contigs* y los dominios conservados presentes en dichas secuencias traducidas.

### 6.1.2 Identificación de posibles proteínas ortólogas a los *contigs PpPLT*

Se seleccionaron 20 secuencias de proteínas anotadas como posibles ortólogas de AIL y PLT en la base de datos Refseq de GenBank; Las secuencias utilizadas pertenecen a las siguientes especies: *Cucumis sativus*, *Fragaria vesca*, *Glycine max*, *Vitis vinifera*, *Solanum lycopersicum*, *Medicago truncatula* y *Zea mays*.

Usando el programa Geneious R7 ([www.geneious.com](http://www.geneious.com)) se hizo un alineamiento múltiple en el que se incluyeron las 20 secuencias seleccionadas, las 6 secuencias de PLT de *A. thaliana*, y los 6 *contigs* de *P. pringlei* con similitud a PLT. Se construyó un árbol filogenético con el programa CLC Genomic Workbench. Se utilizó el algoritmo de Neighbor Joining como un método para generar una filogenia preliminar.

De acuerdo a los resultados del árbol filogenético se asignó un probable ortólogo de *A. thaliana* a cada uno de los *contigs*. Posteriormente, con la función BLASTx del programa CLC Genomic Workbench, se alinearon cada uno de los *contigs* con su

posible homólogo más cercano para comparar el grado de conservación de cada proteína en general y de los dominios AP2 en particular.

## **6.2 Análisis filogenético de genes *AIL***

### **6.2.1 Búsqueda en bases de datos de proteínas *AIL***

Las secuencias de proteínas *AIL* se obtuvieron en la base de datos de Phytozome versión 9 (<http://www.phytozome.com>) usando la herramienta tBLASTx, usando como base las secuencias de proteínas *PLT* de *Arabidopsis thaliana* [*PLT1* AT3G20840; *PLT2*: AT1G51190; *PLT3*: AT5G10510; *PLT4*/*BBM*: AT5G17430; *PLT5*: AT5G57390; *PLT7*: AT5G65510]. El tBLASTx se realizó usando la matriz BLOSUM62 para cada una de las especies de embriofitas con genoma completamente secuenciado y que están disponibles en la base de datos de Phytozome. Los genes *PLT* putativos se seleccionaron de acuerdo con los siguientes criterios: un E-value menor a  $1 \times 10^{-10}$  en comparación con la secuencia base, el marco de lectura abierto de la secuencia debía contener al menos 100 aminoácidos, y la proteína traducida debía contener dos dominios AP2. Las secuencias utilizadas se enlistan en el Anexo II (parte1). La nomenclatura usada para las especies incluye las primeras 3 letras del género y las primeras 3 letras de la especie.

### **6.2.2 Alineamiento de secuencias y reconstrucción filogenética de proteínas *AIL***

Las 309 secuencias seleccionadas de las bases de datos se alinearon usando la herramienta Clustal Omega para alineamientos de múltiples secuencias (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) y el resultado de los alineamientos fue editado manualmente en BioEdit ([www.mbio.ncsu.edu/bioedit/bioedit.html](http://www.mbio.ncsu.edu/bioedit/bioedit.html)) . Se fijó como valor de cobertura mínimo el 65 % (i.e., las secuencias con más de 35 % de gaps fueron excluidos). Para identificar las secuencias como proteínas putativas *PLT* o *AIL* se construyó un árbol filogenético con las 309 secuencias usando el método de Neighbor Joining. Este es un método que permite construir de manera eficiente una filogenia al usar una gran cantidad de datos. El árbol consenso fue inferido de 1,000 réplicas de *bootstrap*.

### **6.2.3 Alineamiento de secuencias y reconstrucción filogenética de proteínas *PLT***

A partir de la filogenia generada de las proteínas *AIL* se identificaron las proteínas *PLT* putativas, se extrajeron sus secuencias y se realinearon. El listado de secuencias utilizadas puede consultarse en el Anexo II (parte 2). En este caso el valor mínimo de cobertura se fijó en el 85 %. Se construyó un nuevo árbol filogenético usando el método maximum likelihood, con el modelo JTT matrix-based y una prueba de *bootstrap* con 1,000 réplicas. Este método de inferencia de filogenias es más exhaustivo que Neighbor Joining y computacionalmente es más demandante, por esa razón se utilizó en este análisis final. Para la construcción de los árboles se empleó el programa MEGA 6.06 (<http://www.megasoftware.net/>).

### 6.3 Germinación de semillas de *Pachycereus pringlei* y su cultivo *in vitro*

Se utilizaron semillas de cactáceas de la especie *Pachycereus pringlei* (S. Watson) Britton & Rose. Las semillas de *P. pringlei* se esterilizaron durante 20 minutos con una solución de hipoclorito comercial al 60 % con 0.02 % de TritonX-100, se realizaron 4 lavados de 10 minutos con agua destilada estéril, y se sembraron en cajas de Petri con medio de Linsmaier-Skoog (LS, PhytoTechnology Laboratories) al 0.2 X, sin sacarosa y 1.2 % de bacto-agar. Las cajas de Petri se colocaron de manera vertical y se cultivaron en una incubadora con condiciones controladas de temperatura a 28 °C y con un fotoperíodo de 12 horas de luz y 12 horas de oscuridad.

### 6.4 Colecta de muestras biológicas para caracterizar la expresión de *PpPLT*

Se colectaron las puntas de raíces primarias, de aproximadamente 1mm de largo, de plantas de *P. pringlei* de cuatro diferentes etapas del desarrollo (Fig. 16).

La etapa inicial del desarrollo se dividió en dos etapas:

- Etapa inicial 1: plantas de 1 a 2 (d.d.g.). La longitud de la raíz es de aproximadamente de 3 a 6 mm, no presenta pelos radicales cerca de la punta (Fig. 16 A).
- Etapa inicial 2: plantas de 3 a 5 (d.d.g.). El meristemo y la zona de elongación siguen presentes y no hay pelos radicales cerca de la punta de la raíz (Fig. 16 B).
- Etapa intermedia: está formada por plantas que empiezan a tener pelos radicales más cerca de la punta de la raíz. Estas plantas tienen aproximadamente de 6 a 8 días de haber germinado (Fig. 16 C).
- Etapa terminal: formada por plantas con el meristemo agotado completamente. Presentan pelos radicales largos en la punta de la raíz (Fig. 16 D).

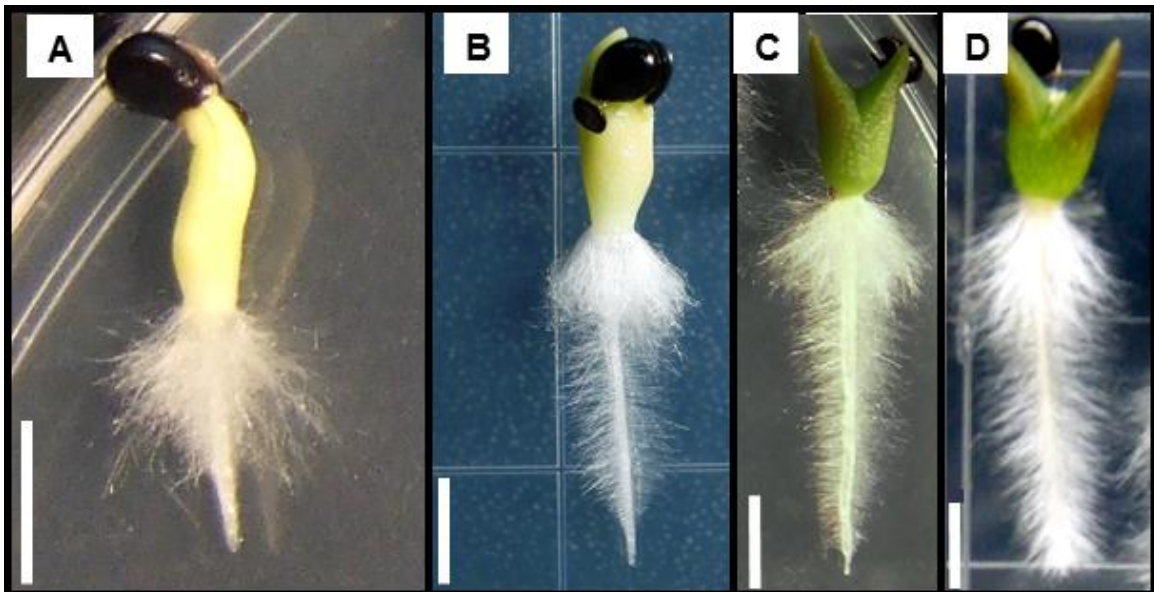


Fig. 16. Etapas del desarrollo de la raíz primaria de *P. pringlei*. A) Etapa inicial 1; B) Etapa inicial 2; C) Etapa intermedia; D) Etapa terminal. Fotografías de plantas a diferentes ampliaciones. La barra incluida en cada imagen es de 5 mm. En el texto que precede a esta imagen se describen las características de las muestras biológicas incluidas en cada etapa.

La edad de plántulas que corresponde a cada etapa o sub-etapa del desarrollo de la raíz primaria de *P. pringlei* es aproximada, porque, como se demostró anteriormente (Rodríguez-Rodríguez *et al.*, 2003), en diferentes individuos cada etapa tiene una duración distinta.

Para seleccionar y clasificar con más facilidad la sub-etapa en la que se encontraba la raíz primaria de cada plántula individual, diariamente se marcó la longitud de la raíz con plumones de diferentes colores. Además, en la etapa inicial la raíz primaria crece rápido, en la etapa intermedia crece más despacio y en la etapa terminal se detiene el crecimiento. Este tipo de monitoreo permite distinguir la etapa en la que se encuentra la raíz. Las puntas de las raíces se depositaron en críoviales que se sumergieron en nitrógeno líquido y se almacenaron posteriormente a -75°C hasta su procesamiento.

### **6.5 Extracción de RNA total**

Se extrajo el RNA total de las muestras colectadas usando el reactivo TRizol® (Invitrogen) bajo las especificaciones del proveedor. Se determinó la concentración del RNA aislado mediante espectrofotometría, considerando el valor de densidad óptica en la longitud de onda de 260 nm usando un NanoDrop® (Thermo Fisher Scientific, Wilmington). Se analizaron las razones de valores de densidad óptica 260/280, 260/230 y por electroforesis el patrón de bandas de RNA ribosomal 28S y 18S de cada preparación de RNA total, en un gel de agarosa al 1 % para estimar si su calidad era aceptable para el ensayo de RT-qPCR.

### **6.6 Preparación de geles de agarosa y electroforesis de RNA y DNA**

Los geles nativos para el análisis las muestras de RNA así como productos de amplificación por RACE-PCR se prepararon con TAE 1X, 1 % de agarosa. Los geles de electroforesis para el análisis de los productos de amplificación de oligonucleótidos para RT-qPCR se prepararon con TAE 1 X agarosa 2 %. Los geles de electroforesis se tiñeron con el colorante GelRed (para la detección de RNA o DNA) de la marca Biotium. Una vez fundida la agarosa se añade 1 µl de GelRed 10,000 X por cada 10 ml de TAE 1 X. Las muestras de RNA total se incluyeron en el gel, y se resolvieron por electroforesis a 80 Volts en buffer TAE 1X.

### **6.7 Tratamiento de RNA con DNasa**

El RNA total de cada una de las muestras se sometió a una reacción de digestión de DNA genómico usando a la enzima DNasa I recombinante (Roche, Suiza); 1 µl por cada 50 µg de RNA. La reacción se incubó a 37 °C por 20 min, tras lo cual se enfrió en hielo. Finalmente, se detuvo la reacción incubando a 75 °C por 10 min.

### **6.8 Síntesis de cDNA a partir de RNA (Retro-transcripción)**

El RNA total tratado con DNasa I se sometió a una reacción de retro-transcripción usando un oligonucleótido oligo-(dT)<sub>18</sub> para generar la primera hebra de cDNA en el molde de los mRNAs poli-adenilados. Se utilizó el kit "RevertAid™ H Minus First Strand cDNA synthesis" (Fermentas/ Life Technologies) o el kit "SuperScript II" (Clontech).

## **6.9 RT-qPCR (PCR cuantitativa con transcriptasa reversa)**

La RT-qPCR es una variante de la PCR utilizada para amplificar y simultáneamente cuantificar el producto de la amplificación de cDNA. Para ello emplea, del mismo modo que la PCR convencional (punto final), un par de oligonucleótidos específicos, dNTPs, un buffer de reacción, un templado (cDNA en este caso) y una DNA polimerasa termoestable. A dicha mezcla se le adiciona una sustancia marcada con un fluoróforo que, en un termociclador con sensores para medir fluorescencia tras excitar el fluoróforo a la longitud de onda apropiada, permita medir la tasa de generación de un producto específico luego de cada ciclo de amplificación.

### **6.9.1 Diseño de oligonucleótidos para RT-qPCR**

Los oligonucleótidos empleados para RT-qPCR deben cumplir con ciertas características: amplificar regiones de entre 75 y 150 pb (incluyendo la región de los oligonucleótidos), deben formar el menor número posible de dímeros y estructuras secundarias, esto tanto para el gen de interés como para los genes normalizadores. Al pertenecer los genes *PLT* a una amplia familia de genes redundantes y con dominios conservados, se usó la región no codificante de la secuencia de los mensajeros para diseñar los oligonucleótidos y así evitar la amplificación de productos inespecíficos.

### **6.9.2 Prueba de oligonucleótidos en PCR (Reacción en cadena de la polimerasa) de punto final**

Los oligonucleótidos se probaron en PCR de punto final para estandarizar las reacciones, *i.e.* probar las temperaturas de alineamiento y número de ciclos requeridos para lograr la amplificación deseada.

El montaje de reacciones con la enzima TaqPol de ThermoScientific se describe a continuación:

Se utilizaron 1  $\mu$ l de cDNA (50 ng), 2  $\mu$ l de amortiguador 10 X (100 mM de Tris a pH 8.3, 500 mM de KCl), 1.2  $\mu$ l de MgCl<sub>2</sub> (15 mM), 0.2  $\mu$ l de mezcla de dNTPs (1.25 mM de cada dNTP), 0.2  $\mu$ l de Taq polimerasa (5 U/ $\mu$ l), 14.2  $\mu$ l de dH<sub>2</sub>O, 0.8  $\mu$ l de cada oligonucleótido a 10  $\mu$ M, para un volúmen total de 20  $\mu$ l.

Se usó el siguiente programa en el equipo C1000 Thermal Cycler de BioRad para amplificar los fragmentos de DNA deseados:

Ciclo de desnaturalización de DNA: 94 °C por 1 min.

Ciclo de alineación de oligonucleótidos y de amplificación de DNA: 94 °C por 30 s, 58 °C 30 s, 72 °C por 1 min. 35 ciclos.

Ciclo de extensión de DNA: 72°C 5 durante 5 min. seguidos de 4 °C.

Se analizó el patrón de fragmentos de DNA por electroforesis.

### **6.9.3 Ensayo de RT-qPCR)**

Las reacciones se realizaron en el equipo LightCycler® Nano de Roche usando aproximadamente 100 ng de cDNA como templado. Se utilizó el reactivo 'Maxima SYBR Green qPCR Master Mix' de Thermo Scientific de acuerdo con las especificaciones del proveedor, escalando las reacciones a un volúmen total de 10  $\mu$ l.

La detección se hizo en la etapa de extensión, para la obtención del valor de C<sub>q</sub> (ciclo de cuantificación) de cada reacción para realizar los cálculos de la abundancia de cDNA en cada muestra.

El programa de amplificación usado fue:

95 °C 10 min.

45 ciclos: 95 °C 15 segundos, 62 °C 15 segundos y 72 °C 30 segundos.

60 °C a 97 °C (Curva de disociación).

En el Anexo III pueden consultarse los detalles de la estandarización de estas reacciones.

#### **6.9.4 Normalización de la expresión de genes blanco**

Se normalizaron los niveles de expresión para cada una de las muestras biológicas usando el modelo matemático para la cuantificación relativa en RT-qPCR publicado por Pfaffl, 2001.

En este método la tasa de expresión de cada gen se calcula mediante la siguiente ecuación,  $2^{-\Delta Ct}$ , donde  $\Delta Ct$  es el cambio en la expresión del gen en la muestra respecto al control en la misma muestra.

Para calcular la expresión relativa de cada gen blanco, la tasa de expresión se normaliza dividiéndola entre el factor de normalización, el cual está dado por la media geométrica de las tasas de expresión de los genes de referencia usados en la muestra.

Como genes de referencia se usaron *PpCLA*, *PpAGL31* y *PpPP2A*, que guardan un alto grado de similitud con los genes de *Arabidopsis At4G24550* (subunidad del complejo adaptador de clatrina), *At5G65050* (proteína MADS-box AGL31) y *AT1G69960* (fosfatasa alcalina 2A) que fueron descritos como genes con expresión estable durante el desarrollo en dicha especie (Czechowski *et al.*, 2005).

#### **6.10 Amplificación de los extremos de cDNA por RACE-PCR**

Para realizar una caracterización más detallada del papel de los genes *PLT* en el desarrollo de la raíz, se pretende en un futuro realizar la hibridación *in situ* de dichos genes. Los genes *PLT* pertenecen a una súper-familia que presenta un alto grado de similitud entre sus miembros. Por ello es necesario contar con la secuencia completa de los transcritos, incluyendo las secuencias 5' y/o 3' no traducidas, que son las regiones no conservadas entre estos genes y que pueden usarse para diseñar las sondas de hibridación para cada uno de los genes *PLT*. Con el fin de obtener las secuencias completas de dichos transcritos, se empleó la estrategia RACE-PCR (amplificación de los extremos de cDNA por PCR). Este método implica la adición de una secuencia conocida a un extremo del mensajero, misma que se usa para poder amplificar el mensajero desde su parte interna conocida y el extremo modificado en una reacción de PCR.

Para cada extremo 5' o 3' las reacciones se realizan por separado usando cDNA modificado del extremo que desea obtenerse. Se realizan dos rondas de amplificación. En la primera ronda de amplificación se usan oligonucleótidos complementarios al extremo modificado y un oligonucleótido dentro de la región conocida del gen de interés (Figura 17). En la segunda ronda se realiza un PCR anidado, es decir, se usan oligonucleótidos más internos que se unen al producto amplificado en la primera

ronda de amplificación, esto con la idea de obtener productos específicos en la segunda reacción.

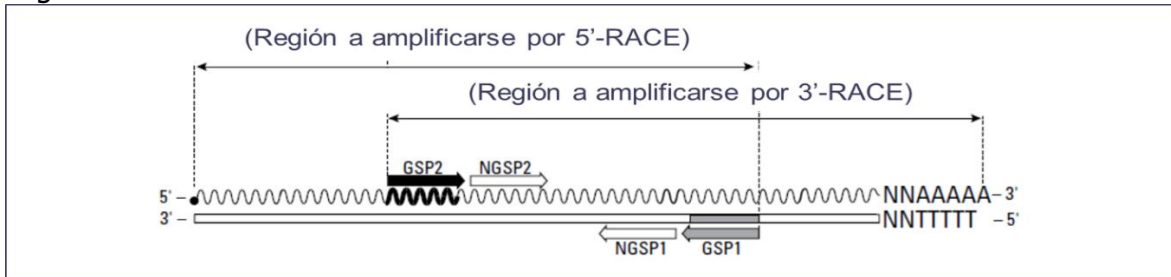


Fig. 17. Esquema de posición de oligonucleótidos usados en RACE-PCR para cada gen de interés. La línea representa la primera hebra de cDNA. GSP, oligonucleótido específico al gen de interés; NGSP, oligonucleótido anidado al gen de interés. Modificado del manual de "SMARTer RACE cDNA amplification" (Clontech, 2009).

### 6.10.1 Diseño de oligonucleótidos para RACE-PCR

Por cada extremo a amplificar de cada uno de los genes de interés debían diseñarse 2 oligonucleótidos con la misma direccionalidad usando como molde la secuencia conocida del gen y cumpliendo con las siguientes características:

- de 23 a 28 nucleótidos
- 50- 70% GC
- $T_m \geq 65^\circ\text{C}$ , preferentemente  $T_m > 70^\circ\text{C}$
- No deben ser complementarios a ninguno de los otros oligonucleótidos usados en estas reacciones (Long universal, short universal)

Se diseñaron oligonucleótidos para todos los genes *PpPLT*. Se sintetizaron solo los siguientes oligonucleótidos:

RACE3plt19\_854F 29 CAT GGA GTG TCT AGC TCT TCT GGG GTC GT  
RACE3plt31\_3518 28 CCG CTG TTC AAG CAG CTT CTC CCA TTA C

PLT15race5\_243R:  
GCC AGG AAA GCA ACC TCT CTC TCT CCC A  
PLT19race5\_282F:  
CCA TGT TGG TGG TGC CTT GTT ACA C  
PLT21race5\_428R:  
CGC CTC CTC TTC AGT AGC AAA AGT G  
PLT13race5\_203R:  
CCT CAA TCT CCT TCT CGT AGT CAC TAA TGG G  
PLT28race5\_163R:  
GAG CTT CTC CGA TTG GTG GTG GGG GT  
PLT31race5\_2346:  
GGT GGT GAG ATG GAG GAA GAG CAG TGG G

PLT15Nes5R1 28: CTA GTG GGA GCT ATG GTG GCG GTG GTG G  
PLT19Nes5R1 26: TCC ACT ACT CTT CCT CCT AAT AGC CG  
PLT31Nes5R1 26: CAG TTG CTC GCT GCT GCT GGT GCC AT  
PLT15Nes3F1 28: AGA CTT GGC GAC CCT TCC TGG ACT GAC C



En la figura 18 se muestra una representación gráfica de la ubicación de los oligonucleótidos utilizados para cada uno de los genes de interés y en la Tabla 1 se indica el tamaño de los amplicones esperados para las reacciones de RACE-PCR que se utilizan dichos oligonucleótidos.

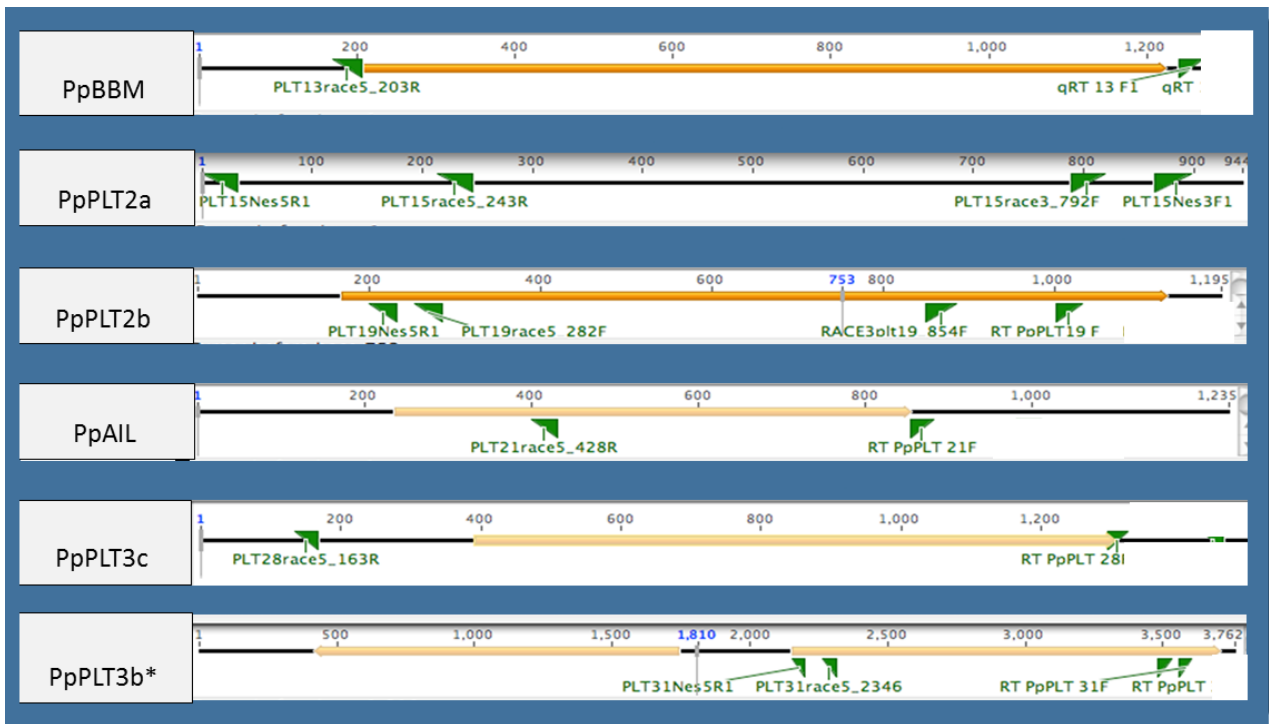


Fig.18. Representación gráfica de la ubicación de los oligonucleótidos utilizados en las reacciones de RACE-PCR para cada uno de los genes de interés. En amarillo se muestra el marco de lectura abierto de los contigs. Los triángulos verdes representan a los oligonucleótidos y su direccionalidad. \* *Contig* quimérico en el que se encontró un segundo marco de lectura abierto (barra amarilla del lado izq.).

Tabla 1. Longitud mínima esperada de los amplicones para las reacciones de RACE-PCR 5' y 3'

	reacción primaria 5'	Reacción secundaria 5'	Reacción primaria 3'	Reacción secundaria 3'
<i>PpPLT2a</i>	+243 pb	+30 pb	+153 pb	+77 pb
<i>PpBMM</i>	+203 pb			
<i>PpPLT2b</i>	+282 pb	+229 pb	+342 pb	+190 pb
<i>PpAIL</i>	+428 pb			
<i>PpPLT3c</i>	+128 pb			
<i>PpPLT3b*</i>	2 kb o +143 pb en caso	+26 pb	+272 pb	+191 pb

	de ser quimérico			
--	------------------	--	--	--

### 6.10.2 Síntesis de poblaciones de cDNAs con extremo 5' modificado

Para amplificar la secuencia faltante del UTR 5' de los mRNAs el primer paso es la generación de una población de cDNAs usando como molde dichos mRNAs y adicionándoles una secuencia conocida en el extremo 5'.

Para ello se utilizó la tecnología SMART (*Switching Mechanism At 5' End of RNA Transcript*), en la que se sintetiza cDNA con una modificación al extremo 5'. La enzima *SMARTScribe Reverse Transcriptase* sintetiza cDNA tomando como templado RNA. Al llegar al extremo 5' del templado adiciona una secuencia conocida de nucleótidos. Esta secuencia es complementaria a una parte del oligonucleótido *SMARTer IIA*. Cuando el oligo SMARTer IIA se ha unido a la secuencia complementaria la enzima, sigue la reacción de síntesis de DNA, ahora usando como templado dicho oligonucleótido.

### 6.10.3 Síntesis de población de cDNAs con extremo 3' modificado

Para obtener esta población se llevó a cabo una síntesis de cDNA estándar utilizando un oligonucleótido dT modificado para añadir una secuencia en el extremo 3'. Esta secuencia es complementaria a la del oligonucleótido *SMARTer IIA*, permitiendo que en reacciones de PCR sucesivas se amplifique el cDNA desde el extremo 3'.

### 6.10.4 Síntesis de oligonucleótidos

Para el RACE PCR fue necesario sintetizar los siguientes oligonucleótidos reportados en el kit "SMARTer RACE cDNA amplification" de Clontech:

raceUniversal: CTA ATA CGA CTC ACT ATA GGG CAA GCA GTG GTA TCA ACG CAGAGT

NestedUniver: AAG CAG TGG TAT CAA CGC AGA GT

ShortUniver: CTA ATA CGA CTC ACT ATA GGG C

Los oligonucleótidos se mandaron sintetizar en la Unidad de Síntesis del Instituto de Biotecnología de la UNAM.

### 6.10.5 Síntesis de población de cDNAs con extremo 5' modificado

Estas reacciones se realizaron siguiendo el protocolo de SMARTer RACE cDNA amplification de Clontech, utilizando algunos reactivos del kit SMARTer PCR cDNA synthesis y la enzima SMARTScribe Reverse Transcriptase de Clontech.

Se usó como templado 1µg de RNA total de raíces de *P. pringlei* y los siguientes oligonucleótidos:

- 5'-CDS Primer A (12 µM)  
5'-(T)<sub>25</sub>V N-3' \*  
(N = A, C, G, o T; V = A, G, o C) (10 µM)  
\* Sintetizado en la Unidad de Síntesis del IBT.
- SMARTer IIA (12 µM)  
5'-AAGCAGTGGTATCAACGCAGAGTACXXXX-3'  
Kit Clontech.

### **6.10.6 Ensayo de amplificación de extremos de cDNA por PCR (RACE PCR)**

Para cada extremo 5' o 3' las reacciones se realizaron por separado usando cDNA con el extremo modificado que desea obtenerse. Se usó la enzima de alta fidelidad Advantage2 de Clontech siguiendo las especificaciones del proveedor.

Para las reacciones primarias se usó el siguiente programa de amplificación (recomendado para Tms entre 60 y 70 °C)

25 ciclos:

94 °C 30 segundos

68 °C 30 segundos

72 °C 4 min.

En las reacciones secundarias se usaron los siguientes programas de amplificación:

Para el caso en el que las Tms de los oligonucleótidos de los genes de interés es mayor a la temperatura del oligo NUP ( $T_m = 57$  °C) se usaron súper ciclos:

- 95 °C 1 min
- 12 ciclos: 94 °C 30 segundos, 62 °C 30 segundos, 72 °C 4 minutos  
94 °C 30 segundos, 62 °C 30 segundos, 72 °C 4 minutos  
94 °C 30 segundos, 53 °C 30 segundos, 72 °C 4 minutos
- Seguimiento de 4 °C

Para el caso en el que las Tms de los oligonucleótidos de los genes de interés fueron menores a 57 °C se usó el siguiente programa de amplificación "touch down":

95 °C 1 min.

Touchdown de 8 ciclos:

95 °C 30 segundos, 60 °C -1 °C por ciclo 30 segundos, 72 °C 4 minutos.

28 ciclos:

95 °C 30 segundos, 53 °C 30 segundos, 72 °C 4 minutos.

72 °C 30 segundos, seguido de 4 °C.

### **Reacción primaria de RACE PCR 5'**

Usando como templado la población de cDNAs con extremo 5' modificado se realizó una reacción de PCR con los oligos:

- Universal Primer Mix:
  - Long (0.4 μM):  
5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'
  - Short (2 μM): 5'-CTAATACGACTCACTATAGGGC-3'
- Oligonucleótido específico 3' del gen de interés

### **Reacción secundaria de RACE PCR (anidada) 5'**

Usando como templado 1 μl del producto de la reacción primaria de RACE PCR 5' y los oligos:

- NestedUniver (10 μM)  
Oligonucleótido anidado del gen de interés

### **Reacción primaria de RACE PCR 3'**

Usando como templado la población de cDNAs con extremo 3' modificado se realizó una reacción de PCR con los oligos:

- Universal Primer Mix (ver arriba)
- Oligonucleótido específico 5' del gen de interés

### **Reacción secundaria de RACE PCR (anidada) 3'**

Usando como templado 1  $\mu$ l del producto de la reacción primaria de RACE PCR 5' y los oligos:

- Nested Universal Primer A (ver arriba)
- Oligonucleótido anidado 5' del gen de interés

## 7. Resultados y discusión

### 7.1 Análisis de secuencias obtenidas a partir del ensamblaje realizado *in silico*

Mediante el método de secuenciación masiva mRNA-seq y el ensamblaje *de novo* (anterior a este trabajo de tesis) se obtuvieron aproximadamente 31,000 contigs (transcritos) de puntas de raíces primarias de *P. pringlei*. La muestra biológica analizada comprendía 1 mm de la punta de raíces primarias de las etapas inicial y terminal del desarrollo de la raíz (Fig. 14). En este primer ensamblaje del transcriptoma se encontraron tres *contigs* con similitud a los genes *PLT*, sin embargo, estos contenían secuencias químéricas.

En el segundo ensamblaje de las mismas lecturas de mRNA-seq, las secuencias de 85 nucleótidos de los cDNAs obtenidos a partir de transcritos poli-adenilados se ensamblaron en ~26,000 contigs, ~5,000 menos que en el primer ensamblaje. En este segundo ensamblaje encontramos seis *contigs* de posibles factores de transcripción de la familia AP2 con un alto grado de similitud a PLT. Dichos *contigs* fueron identificados mediante BLASTx usando como *query* la proteína PLT2 de *A. thaliana*. Las secuencias de nucleótidos de los *contigs* pueden consultarse en el Anexo I.

Los seis *contigs* encontrados se analizaron por BLASTx en la base de datos de RefSeq para confirmar la posible homología con proteínas PLT. En la tabla 2 se muestra el resultado.

Tabla 2. *Contigs* de *P. pringlei* del segundo ensamblaje con similitud a PLT.

Contig	Descripción del primer hit de BLAST (NCBI)	Primer hit de BLAST (NCBI)	Longitud (pb)	E-value
Contig A	PREDICTED: AP2-like ethylene-responsive transcription factor PLT2-like [Vitis vinifera]	XP_002285539	944	2.16E-105
Contig B	PREDICTED: AP2-like ethylene-responsive transcription factor PLT2-like [Glycine max]	XP_003539421	1195	1.05E-105
Contig C	AP2 domain-containing transcription factor [Populus trichocarpa]	XP_002311259	1315	1.89E-119
Contig D	DNA binding protein, putative [Ricinus communis]	XP_002525305	1506	1.25E-85
Contig E	PREDICTED: AP2-like ethylene-responsive transcription factor ALL6-like [Vitis vinifera]	XP_003635541	*3762	3.41E-136
Contig F	PREDICTED: AP2-like ethylene-responsive transcription factor ALL6-like [Vitis vinifera]	XP_003635541	1235	9.57E-110

\*Posible *contig* químérico.

Se identificaron los marcos de lectura abiertos de dichos contigs. Bajo dichos marcos de lectura se tradujeron las secuencias nucleotídicas para generar las secuencias de aminoácidos correspondientes, que pueden consultarse en el Anexo I.

En el análisis de dominios conservados con la base de datos Pfam, se encontró que cada uno de los *contigs* contiene los dos dominios AP2, que están presentes en la subclase de proteínas AIL. Se determinó además, que partes de las secuencias de los *contigs* corresponden a las regiones no traducidas (UTRs). En la figura 19 se muestra una representación gráfica de las regiones codificantes, UTRs y dominios AP2 de cada *contig*.

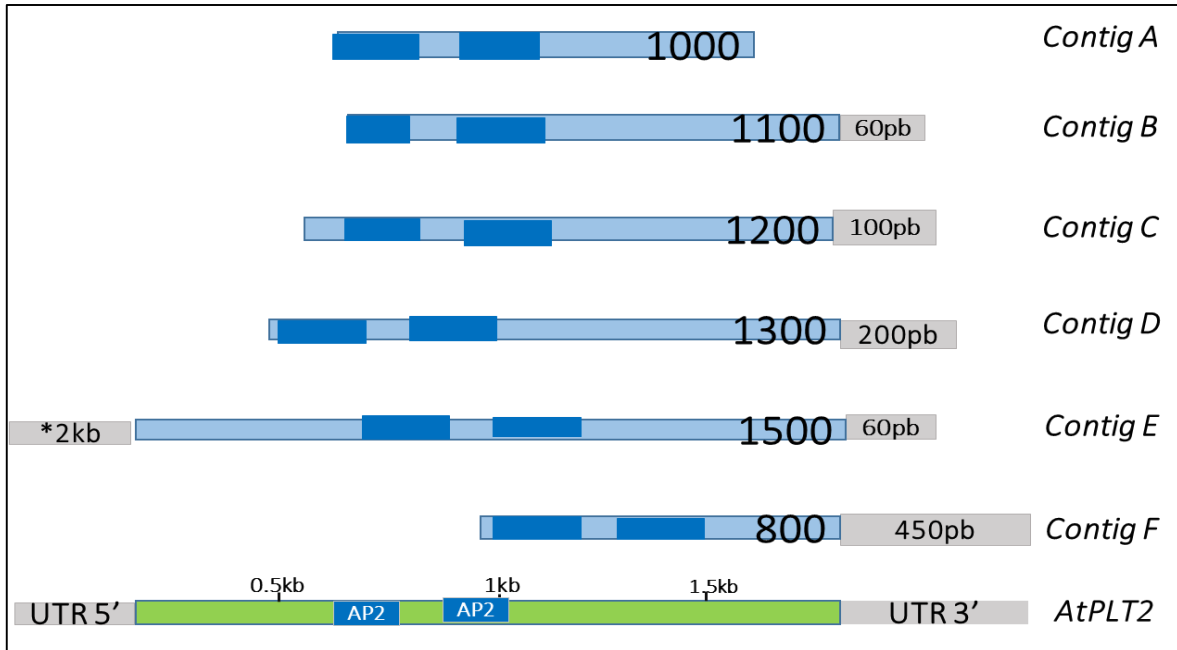


Fig. 19. Integridad de secuencias codificantes y regiones 5' y 3' no traducidas en los *contigs PpPLT*. En azul se muestra la longitud (en pb) de la secuencia codificante contenida en cada *PpPLT*. En azul oscuro se muestra la localización de los dominios AP2 dentro del marco de lectura. En gris, la longitud de las secuencias no traducidas en ambos extremos del transcrito. En verde, representación gráfica de la secuencia del gen *PLT2* de *A. thaliana*. \*Región UTR 5' dentro de la cual se encontró un segundo marco de lectura abierto en sentido opuesto, en el *contig* quimérico.

Las secuencias codificantes de los *contigs PpPLT* tienen de 800 a 1500 pb, mientras que en *A. thaliana* los marcos de lectura abiertos tienen ~1800pb. El *Contig E* es el único que contiene una metionina inicial en su secuencia, lo cual nos indica que el resto de los *contigs* contienen secuencias truncas. En la región 3' no traducida (3' UTR) las secuencias *PpPLT* contienen entre 60 a 450 pb, a excepción del *Contig A* que no cuenta con los extremos no codificantes. Únicamente para el *Contig E* se encontró la región 5' UTR, aunque este *contig* parece ser quimérico.

El rápido desarrollo de la programación permitió mejorar mucho los resultados para el segundo ensamblaje, en el cuál se han encontrado muy pocos *contigs* quiméricos. Sin embargo, este fue el caso del *Contig E* que contiene un marco de lectura abierto completo con alto grado de similitud a la proteína PLT3; dentro del mismo *contig* río arriba se encontró otro marco de lectura abierto de ~1.5kb que muestra similitud con una proteína hipotética (Fig. 20).

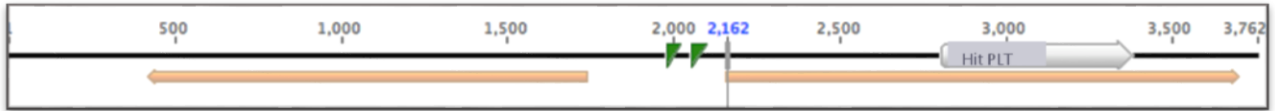


Fig. 20. Dos marcos de lectura abiertos encontrados por BLASTx en el *Contig E*. La línea negra representa la secuencia de nucleótidos, las flechas amarillas representan los marcos de lectura abiertos encontrados. La flecha gris muestra la región del marco de lectura abierto que tiene similitud con PLT.

Es posible que se trate de dos mRNAs unidos en un mismo *contig* en un error de ensamblaje debido a la presencia de secuencias parecidas (complementarias) en sus secuencias no traducidas 5'. Los análisis del *Contig E* se realizaron basándose solamente en la secuencia del ORF que guarda similitud con los genes PLT. Sin embargo, este segundo marco de lectura puede afectar los valores de la expresión virtual, por lo que subsecuentemente en el texto se marcará con un asterisco.

La subfamilia de genes *PLT* en *A. thaliana* contiene genes redundantes que presentan dominios altamente conservados y un alto grado de homología entre sí. Para poder atribuir a cada uno de los *contigs* un posible ortólogo de *A. thaliana*, se compararon las secuencias traducidas de los *contigs* de *P. pringlei* con las de proteínas PLT y AIL anotadas en la base de datos RefSeq, estas secuencias pueden consultarse en el Anexo I. En la figura 21 se muestra la representación gráfica del alineamiento de dichas secuencias.

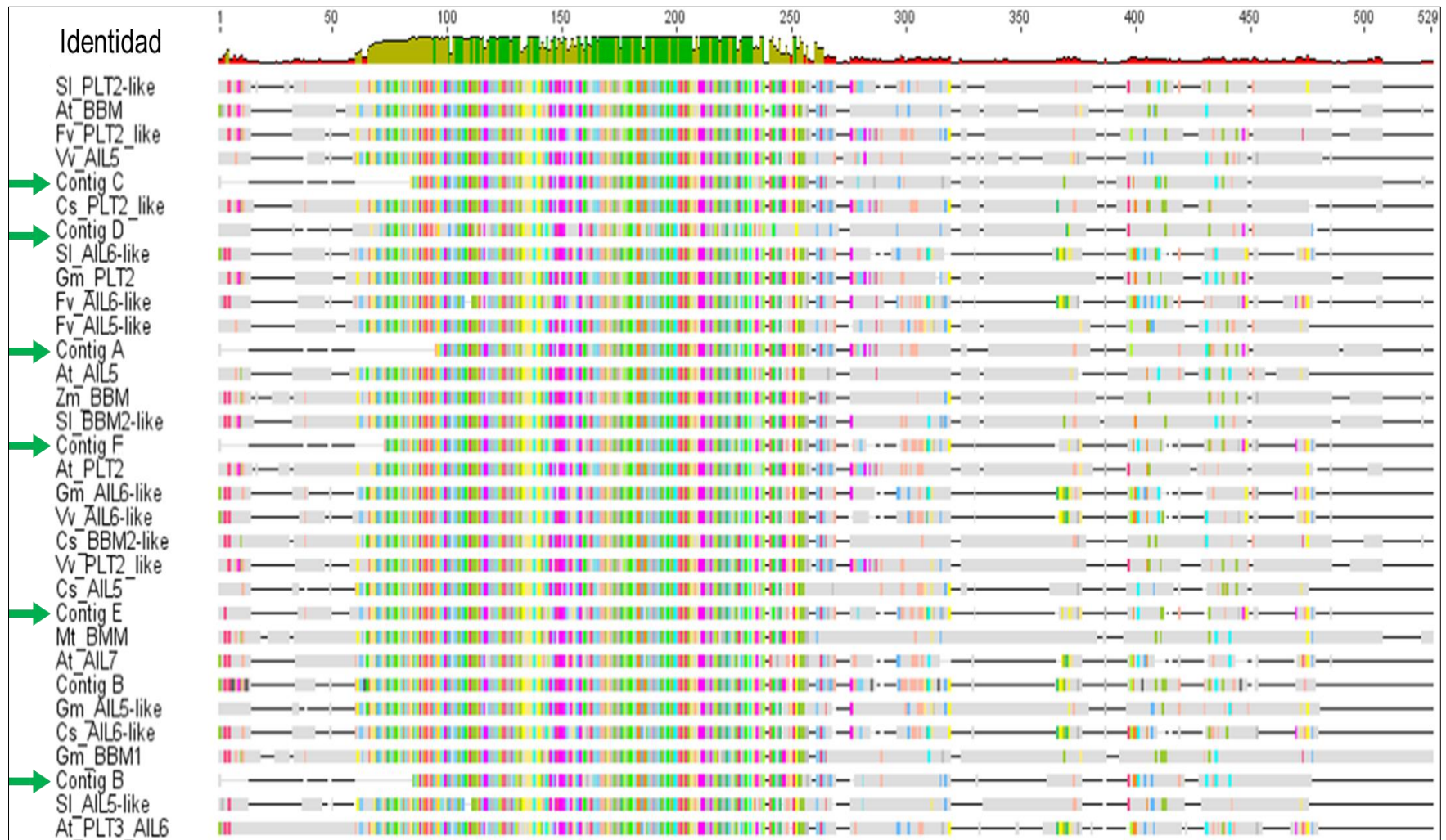




Fig. 21. Alineamiento de secuencias de proteínas con alto grado de similitud a PLT. Las flechas verdes muestran las secuencias traducidas de los *contigs* de *P. pringlei*. Cs (*Cucumis sativus*), Fv (*Fragaria vesca*), Gm (*Glycine max*), Vv (*Vitis vinifera*), Sl (*Solanum lycopersicum*), Mt (*Medicago truncatula*) y Zm (*Zea mays*). Gráfico y alineamiento generado en Geneious7. En la parte superior se representa la secuencia consenso, en la que en verde se destacan los aminoácidos conservados y en café los medianamente conservados. En el resto de las secuencias se resaltan los aminoácidos conservados de acuerdo al código de colores de aminoácidos del programa. Las líneas negras representan gaps en el alineamiento. La numeración en la parte superior indica la longitud de la secuencia consenso.

Los dominios AP2 se encuentran dentro de la región más conservada del alineamiento. Los dominios AP2 de los *contigs* de *P. pringlei* muestran alta similitud con los del resto de las proteínas PLT.

A partir del alineamiento antes presentado se generó un árbol filogenético usando el programa *CLC Genomic Workbench6* y con el algoritmo de *Neighbor Joining* (Fig. 22).

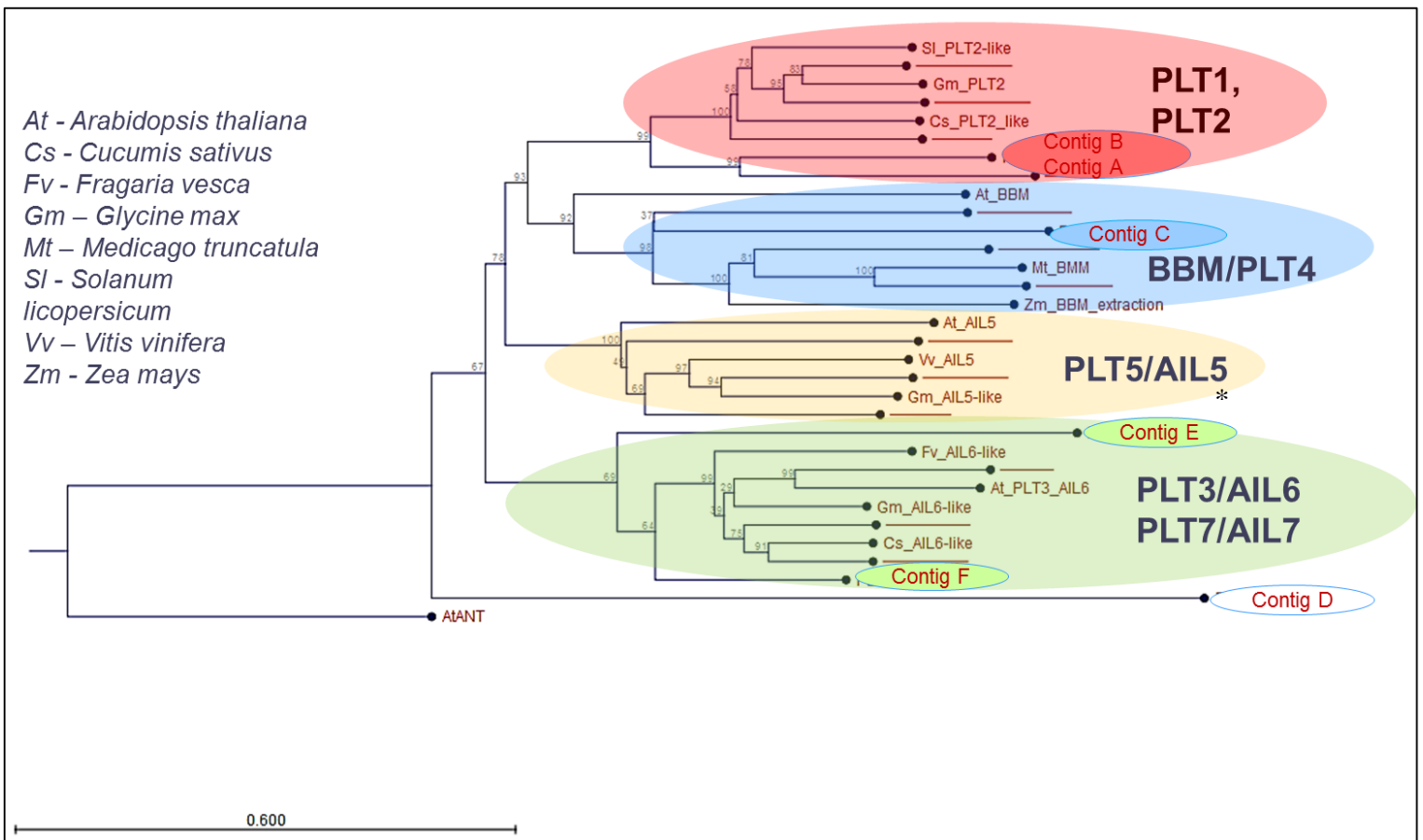


Fig. 22. Árbol filogenético de secuencias de proteínas con alto grado de similitud a PLT y de los *contigs* de *P. pringlei*. \*Posible *contig* quimérico.

En el árbol generado se puede apreciar que las proteínas PLT1 y PLT2 de *A. thaliana* tienen un alto grado de homología, por lo tanto estas dos proteínas de *A. thaliana* se ubican en un mismo nodo. Dos de las secuencias traducidas de cactáceas se agrupan

en esta misma rama. Esto sugiere que las secuencias codificadas por los *Contigs* A y B son los potenciales ortólogos de PLT1 y/o PLT2 de *P. pringlei*.

La secuencia traducida del *Contig C* de *P. pringlei* se ubica en el nodo que agrupa a las posibles proteínas ortólogas de BBM, por lo que el gen correspondiente podría ser el ortólogo de *BBM*. Las tres secuencias restantes (*Contigs* D, E y F) de cactáceas son más cercanas a PLT3/AIL6 y PLT7. Los *Contigs* E y F se ubican dentro de esta rama, el *Contig D* queda muy cercano a esta rama aunque no se agrupa con las demás secuencias. Considerando estos resultados, los *contigs* se nombraron de acuerdo a su posible homólogo más cercano de *A. thaliana* y usando el prefijo *Pp* (*P. pringlei*) como se muestra en la Tabla 3.

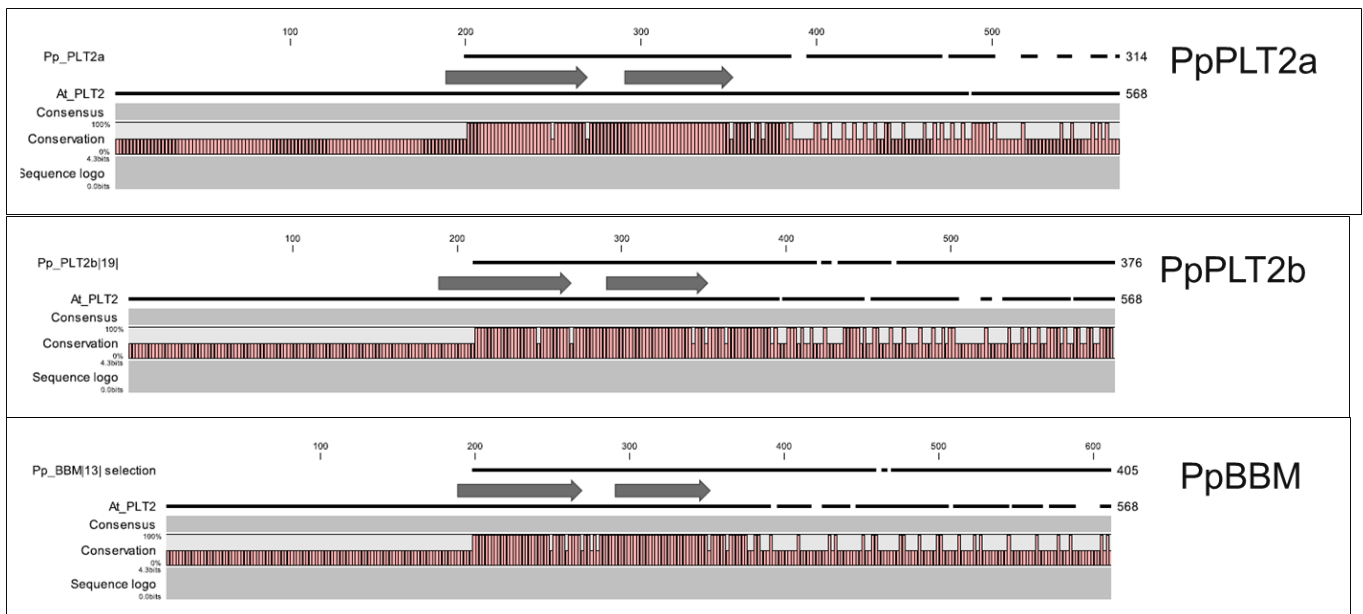
Tabla 3. Nombres asignados a los *contigs* de *P. pringlei* de acuerdo a su similitud con PLT.

Contig	ID	Longitud (pb)	Secuencia codificante(aa)	Identidad**	Similitud**
<i>Contig A</i>	PpPLT2a	944	314	68% (214/314 aa)	77% (245/315 aa)
<i>Contig B</i>	PpPLT2b	1195	376	63% (242/387 aa)	71% (275/387 aa)
<i>Contig C</i>	PpBBM	1315	405	89% (164/185 aa)	96% (178/185 aa)
<i>Contig D</i>	PpPLT3a	1506	432	60% (124/206 aa)	74% (153/206 aa)
<i>Contig E*</i>	PpPLT3b	3762*	511	66% (219/334 aa)	74% (249/334 aa)
<i>Contig F</i>	PpPLT3c	1235	283	66% (209/319 aa)	72% (230/319 aa)

\*Posible *contig* quimérico.

\*\* Respecto a su posible ortólogo más cercano de *A. thaliana*.

Para examinar el grado de conservación de los dominios AP2 de unión a DNA, las secuencias traducidas de los *contigs* se compararon con la secuencia de aminoácidos de PLT2 de *A. thaliana* (Figura 23).



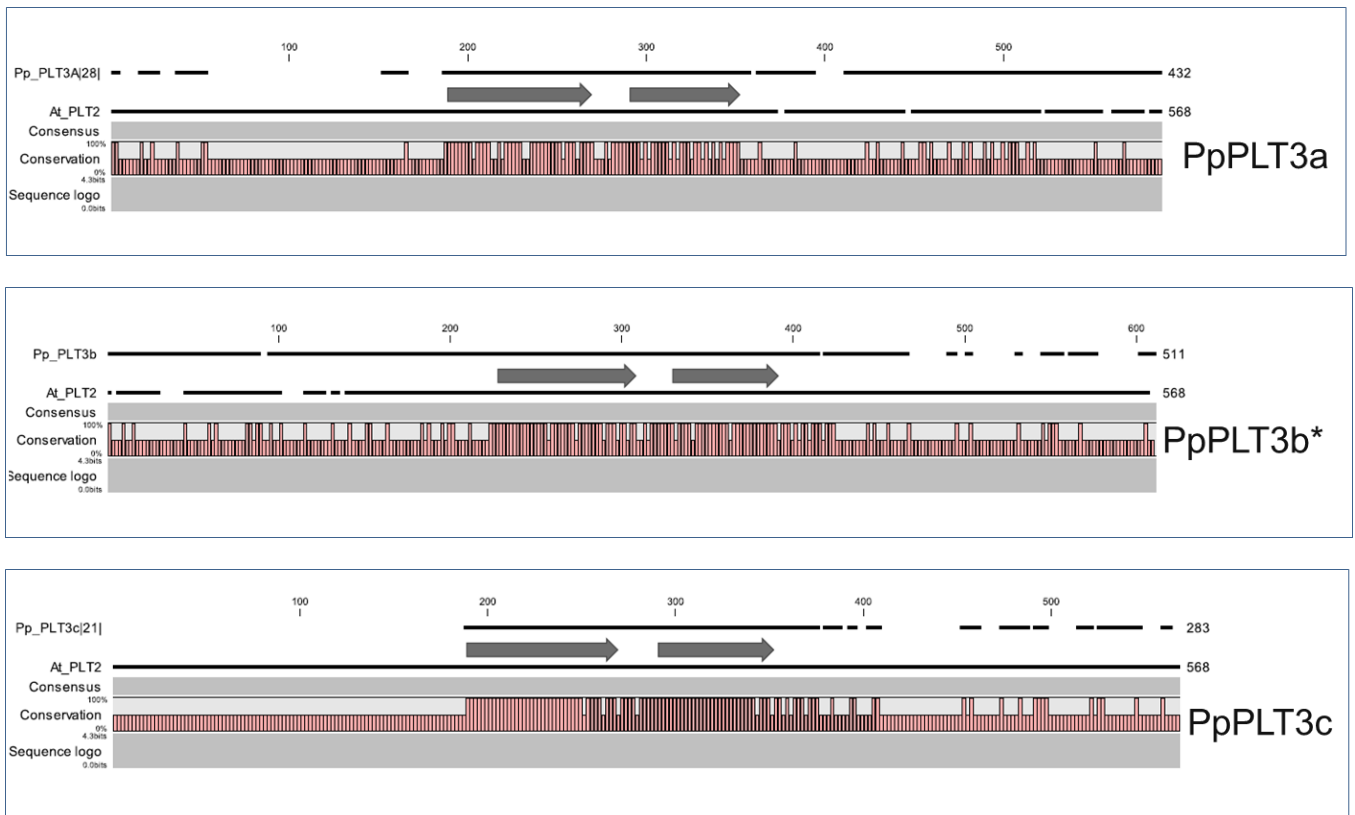


Fig. 23. Alineamiento de cada una de las secuencias traducidas de los *contigs* de *P. pringlei* Vs. PLT2 de *A. thaliana* en la región más conservada que incluye los dominios AP2. La línea negra superior representa el *contig* de *P. pringlei*, la discontinuidad en las líneas negras representa los gaps en el alineamiento. Las flechas representan los dominios AP2 de *A. thaliana*. Las barras altas representan los aminoácidos conservados. Las barras rosas bajas representan los aminoácidos no conservados entre las dos proteínas. La numeración en la parte superior, corresponde a la longitud de la secuencia consenso. Los números en parte derecha de los esquemas hacen referencia a la longitud de la secuencia de aminoácidos de los *contigs* traducidos y de PLT2 de *A. thaliana*. \*Posible *contig* quimérico.

Como puede observarse, los dominios AP2 están altamente conservados en todos los *contigs*, siendo PpPLT3a el menos conservado. En el caso de PpPLT2a, PpPLT2b y PpBBM falta una pequeña porción de la secuencia de uno de los dominios AP2.

En la figura 24 se compara el tamaño de los posibles ortólogos a proteínas PLT de diferentes especies vegetales contra las PpPLT.

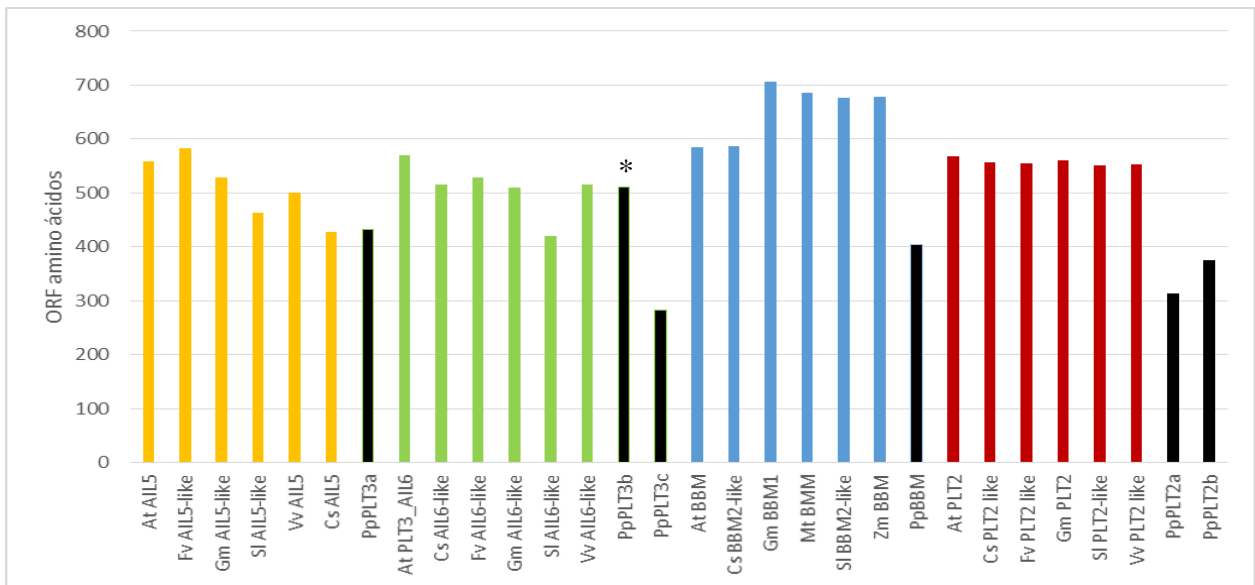


Fig. 24. Longitudes de las secuencias de aminoácidos de las proteínas PLT. En negro, secuencias traducidas de los *contigs* de *P. pringlei*. Agrupados en color amarillo los posibles ortólogos de AIL5/PLT5, en color verde AIL6/PLT6, en color azul BBM y en color rojo PLT1/2 de *Cs* (*Cucumis sativus*), *Fv* (*Fragaria vesca*), *Gm* (*Glycine max*), *Vv* (*Vitis vinifera*), *Sl* (*Solanum lycopersicum*), *Mt* (*Medicago truncatula*) y *Zm* (*Zea mays*). \*Se usó solo el marco de lectura con similitud a PLT del posible *contig* quimérico (ver fig. 16).

Los *contigs* traducidos PpPLT3c, PpBBM, PpPLT2a y PpPLT2b son considerablemente más cortos que sus posibles ortólogos. Esto aunado a que las secuencias de los *contigs* no contienen secuencias codificantes completas nos indica que las secuencias que obtuvimos están truncadas.

Tras el análisis preliminar de los *contigs*, encontramos que el segundo ensamblaje, que se realizó a partir de un número no muy grande de lecturas de longitud media y de un solo extremo ("single reads" o lecturas únicas), lo cual resultó en secuencias incompletas de los cDNAs secuenciados.

Lo anterior podría ser resultado del alto nivel de similitud de los dominios AP2 y del gran número de genes pertenecientes a la súper-familia génica *AP2*, que en *A. thaliana* consiste de 144 miembros (Sakuma *et al.*, 2002). Para evitar el ensamblaje de *contigs* quiméricos en el segundo ensamblaje, fue inevitable obtener *contigs* incompletos para una gran familia de proteínas con dominios conservados. La solución para mejorar nuestra estrategia de ensamblaje sería utilizar mayor número de lecturas más largas y, sobre todo, pareadas.

## 7.2 Análisis filogenético

Para corroborar los resultados del análisis filogenético preliminar, se realizó un análisis extensivo en el que se incluyeron todas las secuencias de proteínas PLT disponibles en bases de datos. Con este fin se usaron las secuencias de los dominios de AP2 de *A.*

*thaliana* para identificar a los posibles homólogos de PLT en las especies vegetales con genoma secuenciado que se encuentran en la base de datos de Phytozome versión 9 (<http://www.phytozome.com>).

Usando la herramienta BLASTx se identificaron 309 proteínas putativas AIL, el subgrupo que contiene a las proteínas PLT y que se encuentra dentro de la superfamilia de factores de transcripción AP2, en 35 genomas (Anexo II). La nomenclatura usada para las especies incluye las primeras 3 letras del género y las primeras 3 letras de la especie, por lo que en este caso las seis secuencias de *P. pringlei* están anotadas como Pacpri1-6.

Las secuencias traducidas se alinearon usando la herramienta Clustal Omega y se construyó un árbol filogenético con el método de *Neighbor joining*, para agrupar las secuencias de acuerdo a su similitud. En la fig. 25 se muestra el árbol consenso resultante de 1,000 réplicas de *bootstrap*.

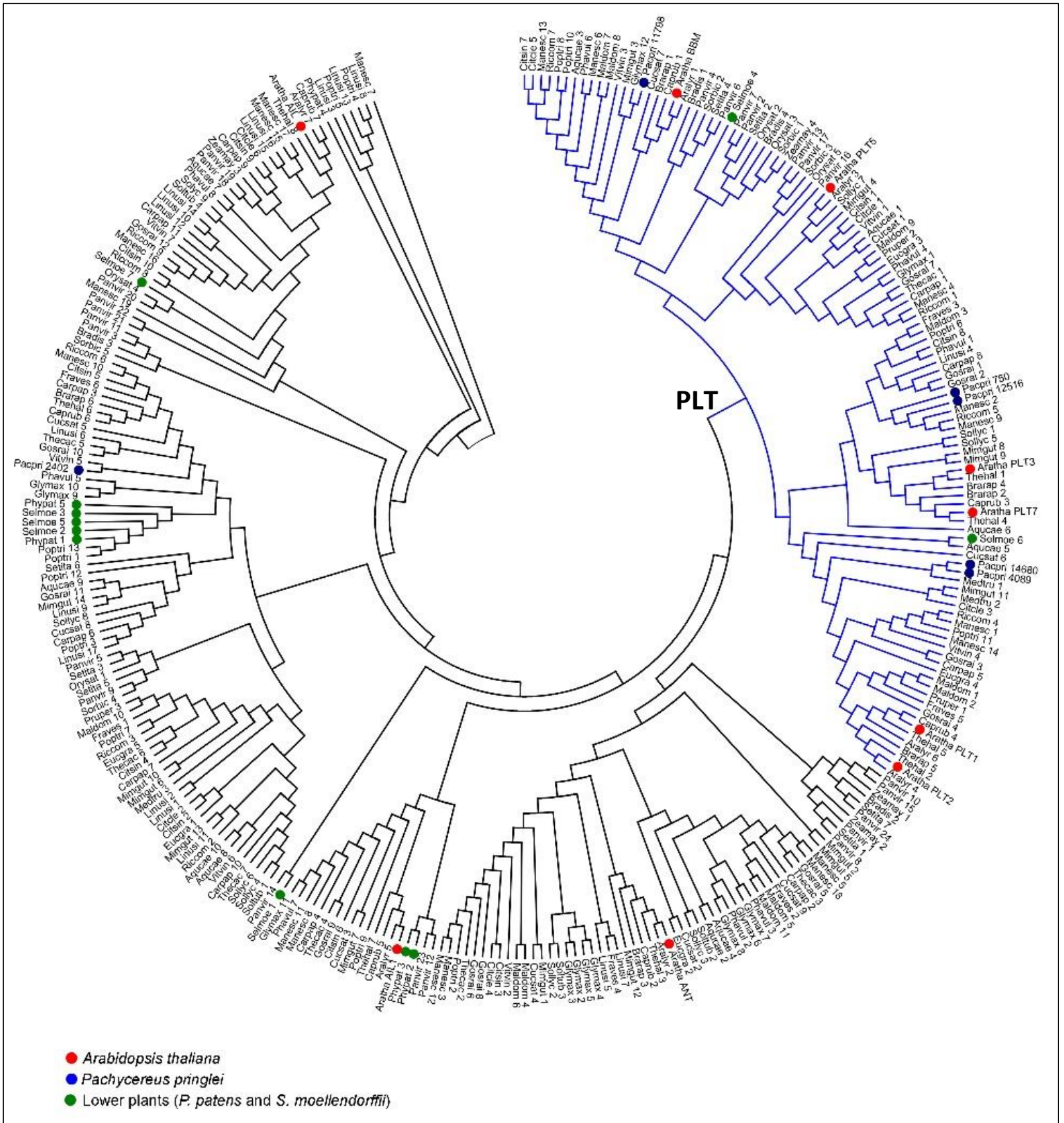


Fig. 25. Árbol filogenético de la subclase de las proteínas AIL pertenecientes a la súper-familia AP2/ERF. Se usó el método de Neighbor joining, para la construcción de este árbol. Se incluyeron 309 secuencias de proteínas de 36 especies vegetales obtenidas de la base de datos de Phytozome. Todas las secuencias

contienen dos dominios AP2. Las ramas azules del árbol corresponden a las proteínas PLT ya que este clado incluye a las proteínas PLT antes descritas en *A. thaliana*.

En el árbol filogenético que se muestra en esta figura pueden distinguirse dos grandes clados: el de las proteínas PLT (cuyas ramas están marcadas con líneas azules) y el clado de las proteínas AIL conformado por el resto de las secuencias del árbol. El *contig* antes denominado PLT3a se ubicó dentro del clado de proteínas AIL en este análisis, por lo que se renombró como PpAIL, nombre con el que se identificará en adelante. De acuerdo a la topología de este árbol pudieron identificarse 114 proteínas PLT de distintas especies vegetales (cuyas ramas están marcadas con líneas azules, en la fig. 25). Las secuencias de las proteínas PLT putativas se realinearon; a partir de esas secuencias se construyó un árbol filogenético usando el método *máximum likelihood*. Este método de inferencia de filogenias evalúa hipótesis acerca de la historia evolutiva en términos de los modelos propuestos, lo que en este caso nos da una mayor certeza de que las filogenias inferidas sean correctas. (Fig. 26).

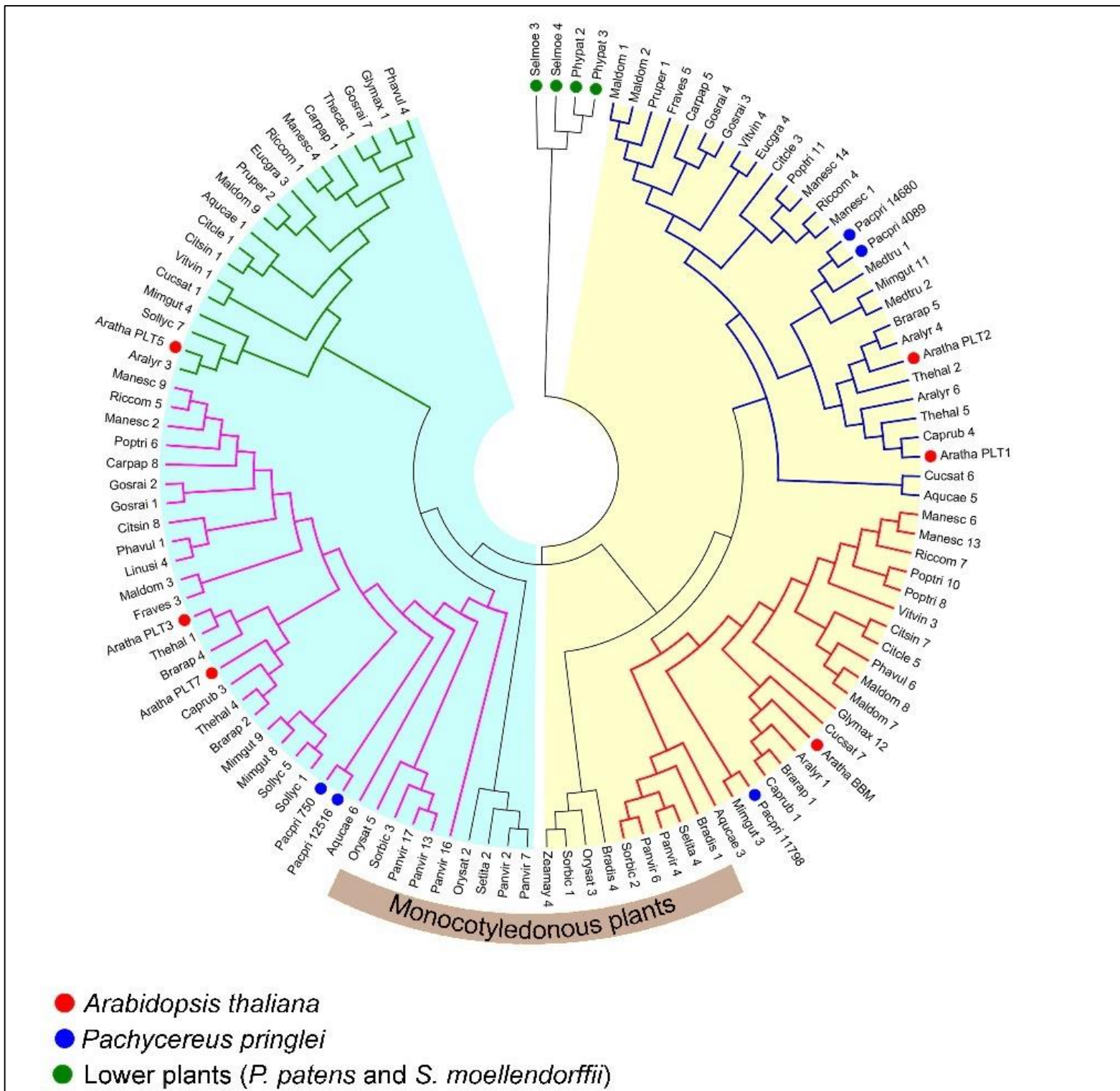


Fig. 26. Árbol filogenético de proteínas PLT. La reconstrucción filogenética se hizo por el método *maximum-likelihood*. La nomenclatura aquí usada para las especies incluye a las primeras 3 letras del género y las primeras 3 letras de la especie. La lista de secuencias y nombres desabreviados usados en la composición de esta figura se encuentran en el Anexo II (parte II). PLT1 y PLT2 se encuentran representados por las ramas azules; PLT3 y PLT7 por las ramas rosadas; BBM y PLT4 por las ramas rojas y PLT5 por las ramas verdes. Las ramas grises corresponden a las proteínas de plantas monocotiledóneas y caen dentro de los clados PLT3 y PLT7 y BBM/PLT4.



Para armar la historia filogenética de proteínas PLT únicamente se utilizaron las secuencias con más del 85% de cobertura en el alineamiento de partida, es decir se excluyeron del alineamiento aquellas secuencias con más de 15% de gaps. Como grupo externo, para enraizar el árbol se incluyeron las secuencias de proteínas putativas AIL del musgo *Physcomitrella patens* y de la licophyta *Selaginella moellendorffii*.

De acuerdo con la topología del árbol filogenético de proteínas PLT (Fig. 26), se clasificaron a las proteínas PLT de plantas dicotiledóneas en cuatro grupos: PLT1 y PLT2; PLT3 y PLT7; PLT5; y BBM/PLT4. Siguiendo esta clasificación, se graficó la longitud de proteínas pertenecientes a cada uno de estos grupos, con el fin de compararlos con la longitud de las secuencias de aminoácidos de los *contigs* de *P. pringlei* (Fig. 27).

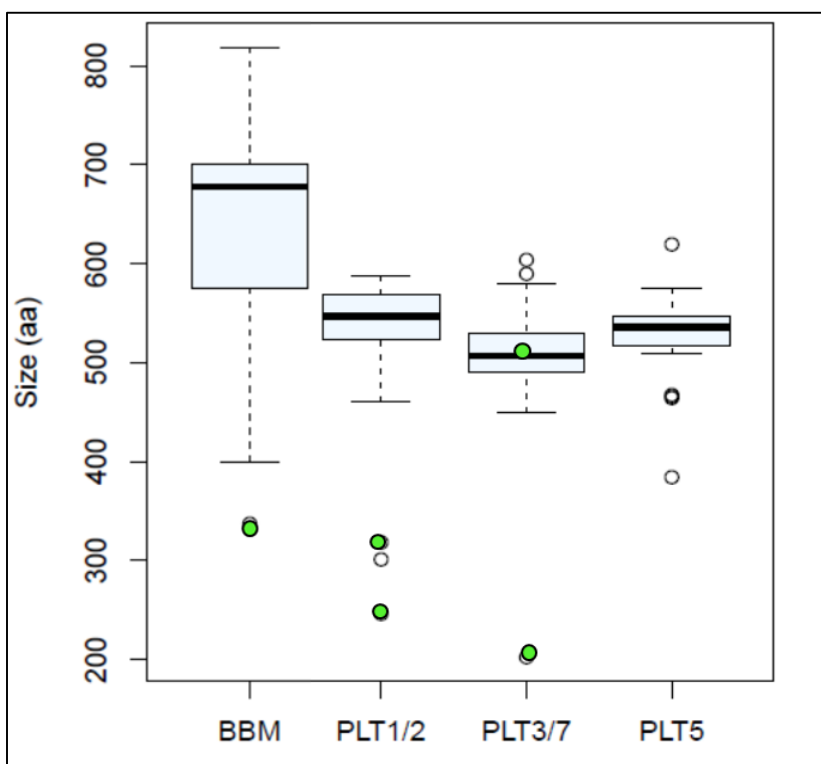


Fig. 27. Longitud de proteínas PLT en diferentes especies dicotiledóneas. Las proteínas se clasificaron como PLT1 y PLT2, PLT3 y PLT7, PLT5, BBM/PLT4 que pertenecen a las ramas separadas. Los círculos verdes, muestran la longitud de los *contigs* de *P. pringlei*. Los círculos blancos representan proteínas de tamaño atípico de otras especies.

Las secuencias de cuatro de las PpPLTs con las que contamos son claramente más cortas que las posibles proteínas homólogas de otras especies vegetales (Fig. 27). Estas cuatro secuencias se encuentran incompletas, ya que ninguna contiene una metionina río arriba de los dominios AP2 que pueda tomarse como metionina inicial. Considerando lo anterior suponemos que el tamaño más corto de las secuencias de *P. pringlei* se debe a que estas secuencias están incompletas y que muy probablemente la longitud de la secuencia completa sea similar a la de otras especies vegetales.

### 7.2.1 Clados A y B de *PLT*

La topología resultante del árbol de proteínas PLT indica que existen dos clases principales de proteínas PLT que nombramos Clado A y Clado B. En el Clado A (sombreado en color amarillo, fig. 26) se agrupan las secuencias con similitud a PLT2, PLT1 y BBM/PLT4. El Clado B (sombreado en color azul, fig. 26) está conformado por las secuencias con similitud a PLT3, PLT5 y PLT7.

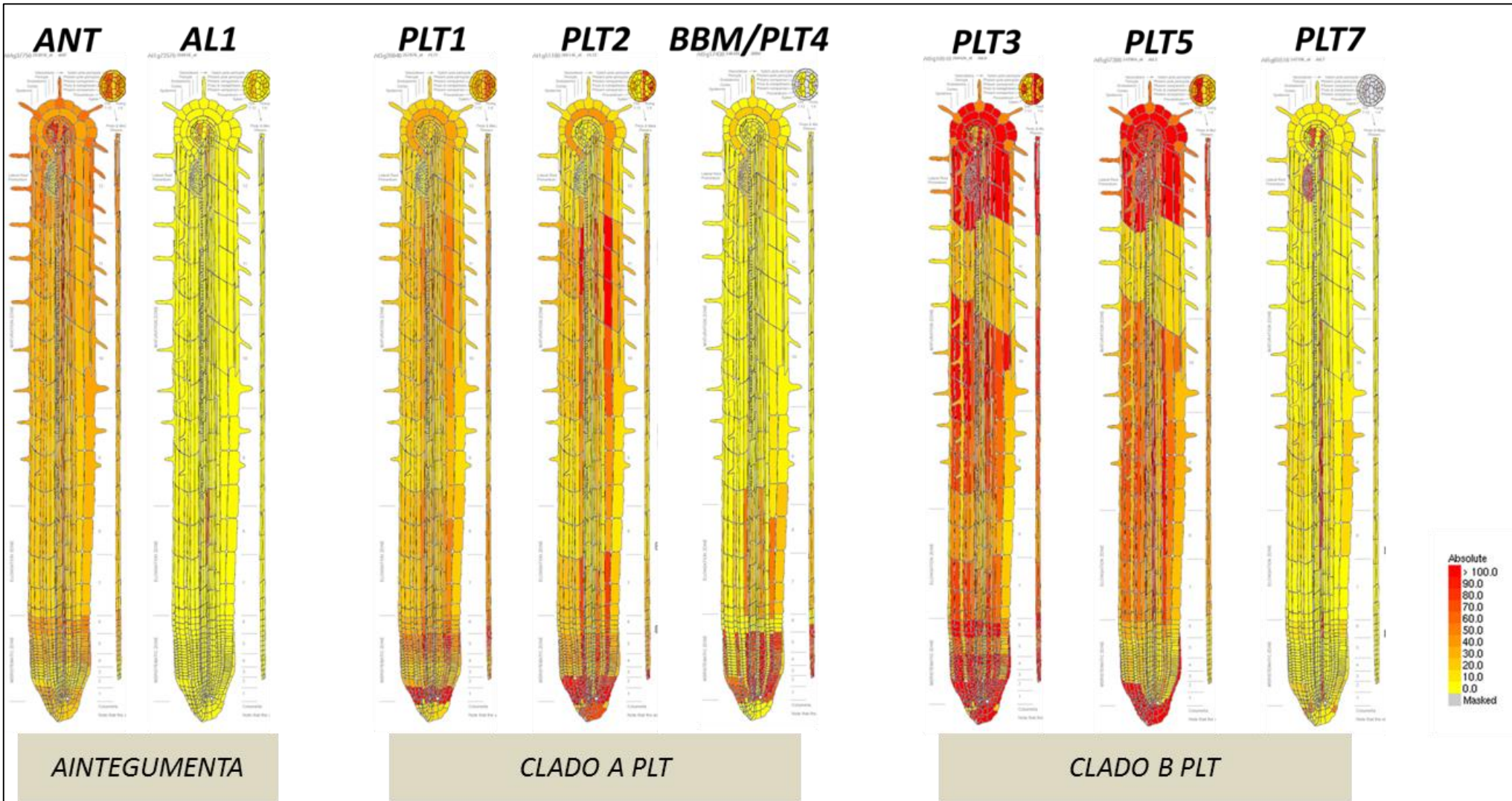
En la raíz de *A. thaliana* el patrón de expresión de los mRNA que codifican proteínas PLT (Fig. 28, consultado en la base de datos eFP Browser <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>) muestra que los genes pertenecientes al Clado A se expresan principalmente en la zona meristemática, incluyendo el CQ. En contraste, los transcritos génicos *PLT5* y *PLT7* del Clado B, tienen poca expresión en RAM, *PLT5* se expresa en la cofia y zona de diferenciación y *PLT3* se expresa en toda la raíz, sin mostrar especificidad en el RAM (Fig. 28 en la pág. 55).

En la zona aérea de la planta, los genes del Clado B se expresan en el SAM en diferentes dominios que se sobrelapan y que son necesarios en el establecimiento de la filotaxis (Prasad *et al.*, 2011). En la raíz, los genes del Clado B se expresan en las células del periciclo que darán origen a los primordios de raíces laterales y que participan en el proceso de rizotaxis (Hofhuis *et al.*, 2013).

Dos de los genes del Clado A, *PLT1* y *PLT2*, son esenciales para el funcionamiento del RAM (Aida *et al.*, 2004); mientras que el otro gen, *BBM*, tiene un papel más importante en el establecimiento del patrón basal durante la embriogénesis (Bouttiler *et al.*, 2002). Del Clado B, *PLT5* es la proteína que se necesita en el establecimiento del patrón basal (Bouttiler *et al.*, 2002); y *PLT3* junto con el gen *ANT* son los que tienen el papel más fuerte en el mantenimiento del SAM (Krizek *et al.*, 2009). La importancia de *PLT3* para el desarrollo de la raíz se ilustra con el hecho de que, como se mencionó anteriormente, la mutante triple en genes *PLT1*, *PLT2* y *PLT3* no desarrolla la raíz.

La especificidad de los roles que juegan en *A. thaliana* los genes en la filotaxis y en la rizotaxis *Vs.* el mantenimiento RAM y SAM concuerdan con la separación de los Clados A y B encontrada en la filogenia. Esto nos indica que las funciones que cumplen los genes *PLT* están altamente conservadas en raíz y parte aérea, por lo que la subfuncionalización y compartimentalización debió surgir por eventos de duplicación durante la evolución.

Fig. 28. Expresión de los genes *PLT* en las puntas de la raíz de *A. thaliana*. Los genes de Clado A (*PLT1*, *PLT2* y *BBM*, ver fig. 24) se expresan principalmente en la zona meristemática. El patrón de expresión de *PLT3* muestra que está presente a lo largo de la raíz, tanto en la zona meristemática, como en la zona de elongación. *PLT5* se expresa principalmente en las células de la cofia y menos aún en la zona de elongación, mientras que para *PLT7* los valores de expresión en la punta de la raíz son mínimos. Los gráficos fueron generados usando el eFP Browser (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>); Los valores de expresión corresponden a resultados de los microarreglos publicados por Brady *et al.* (2007), en los que se utilizaron plántulas de 5 a 7 d.d.g.



## 7.2.2 Proteínas PLT en plantas monocotiledóneas y dicotiledóneas

La disponibilidad de los genomas de pastos nos permitió identificar 18 secuencias putativas de PLTs de 6 especies de plantas monocotiledóneas. Tras un extensivo análisis filogenético, encontramos que la rama de proteínas PLT5 está formada exclusivamente por secuencias de plantas dicotiledóneas y las PLT de monocotiledóneas se distribuyen en 4 diferentes ramas (Fig. 26).

La topología del árbol de proteínas PLT (Fig. 26) indica que *BBM*, *PLT3* y *PLT7* se establecieron antes de la separación que dió origen a las plantas monocotiledóneas y dicotiledóneas. Además, indica que en las dicotiledóneas ocurrieron duplicaciones de segmentos que originaron, por un lado, a la rama PLT5; y por otro lado, a la rama PLT1 y PLT2. Por su parte, en las monocotiledóneas también ocurrieron duplicaciones que originaron a las ramas que contienen exclusivamente secuencias de monocotiledóneas.

En previos análisis (Kim *et al.*, 2006; Long *et al.*, 2007) se ha reportado que las secuencias de las proteínas PLT de monocotiledóneas y dicotiledóneas se agrupan en clados separados. Más tarde, en Prasad *et al.*, 2011 (Fig. 29) se reportó que las PLT de monocotiledóneas se agrupan con las PLT de dicotiledóneas en uno de los subclados, el de PLT5, por lo que se infirió que el ancestro común de ambos grupos tendría genes similares a *PLT5*; sin embargo, estos análisis se llevaron a cabo con pocas secuencias de proteínas monocotiledóneas. El número considerablemente más grande de las proteínas incluidas en nuestro análisis nos permite generar la hipótesis alternativa que se mencionó antes.

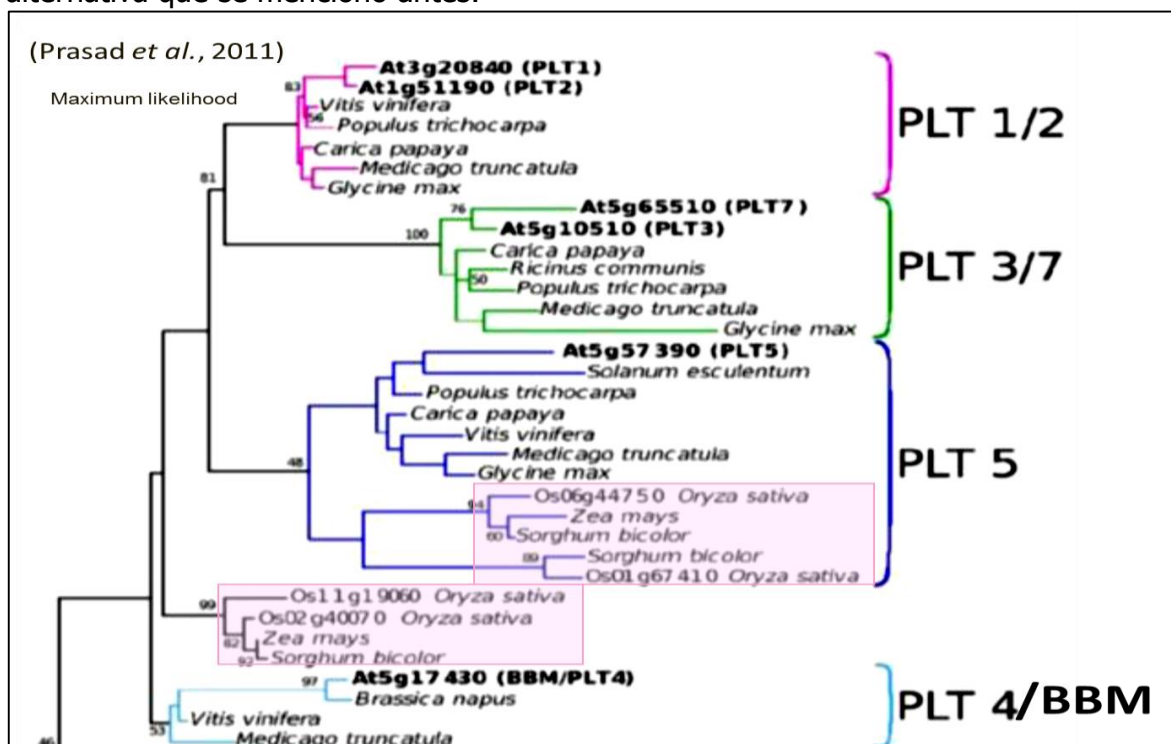


Fig. 29 Análisis filogenético de las proteínas PLT-like, imagen modificada de Prasad *et al.*, 2011. En los recuadros rosas se resaltan las secuencias de especies monocotiledóneas.

## Caracterización de la expresión de los genes *PpPLT*

### 7.3 Expresión Virtual de los *contigs*

Las lecturas de mRNA-seq de las etapas inicial y terminal del desarrollo de la raíz primaria de *P. pringlei* (ver fig. 14) fueron mapeadas sobre los *contigs* ensamblados, para estimar niveles virtuales de expresión génica. Se contó el número total de lecturas que mapearon a cada *contig* y este número se normalizó usando el método de RPKMs (Reads Per Kilobase per Million Mapped Reads). En la siguiente gráfica se presentan los niveles de expresión virtual de los mRNAs correspondientes a los *contigs PpPLT* de acuerdo al segundo ensamblaje (Fig. 30).

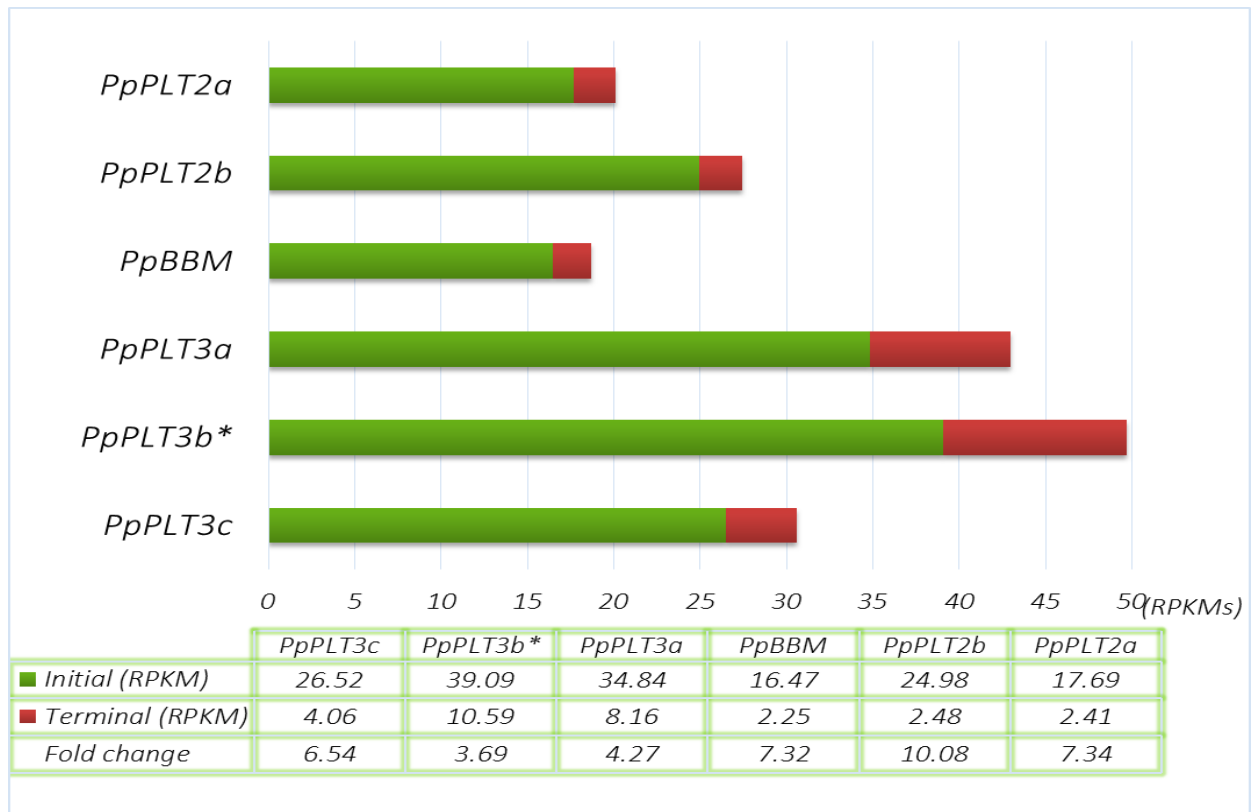


Fig. 30. Expresión virtual de los transcritos del segundo ensamblaje con similitud a *PLT* analizados por el programa *CLC Genomic Workbench*. En la gráfica se presentan los valores de RPKMs de los diferentes *contigs PpPLT*. En verde, los valores de RPKMs correspondientes a la muestra de la etapa inicial del desarrollo de la raíz primaria; en rojo, los correspondientes a la etapa terminal. *Fold change*: número de veces en que difieren los valores de RPKM entre la etapa inicial respecto a la etapa terminal. \*Posible *contig* quimérico.

De acuerdo a la diferencia entre los valores RPKM, los seis *contigs* con similitud a *PLT* se encuentran inducidos en la etapa inicial del desarrollo de la raíz primaria. Es notable también que los valores RPKM de *PpPLT2a*, *PpPLT2b* y *PpBBM* en la etapa terminal son menores a 3 RPKM, es decir, la abundancia de estos transcritos en esta etapa es muy baja. Por esta razón, aunque es claro que los tres transcritos se expresan más fuertemente en la etapa inicial, no se puede hacer una estimación adecuada de

diferencia en abundancia de transcritos entre las distintas etapas de desarrollo de la raíz. Para los otros tres transcritos *PpPLT3a*, *PpPLT3b* y *PpPLT3c* el nivel de expresión virtual en la etapa inicial llega a ser de hasta aproximadamente 6.5 veces más que en la etapa terminal. La disminución en los niveles de expresión de los seis genes en la etapa en la que el meristemo ya no está presente, sugiere que los genes *PLT* son necesarios para el mantenimiento de RAM en *P. pringlei*, aunque aún es necesario corroborar estos datos; por esta razón decidimos cuantificar la abundancia de los transcritos por RT-qPCR.

## 7.4 Caracterización de la expresión de los genes *PpPLT* por RT-qPCR

Mediante RT-qPCR se cuantificó la abundancia relativa del transcrito correspondiente a genes *PpPLT* para las cuatro etapas del desarrollo de la raíz primaria de *P. pringlei* que se definieron en la sección de Material y métodos 6.3, ver fig. 16. Se realizaron dos réplicas biológicas para este experimento y dos réplicas técnicas. Las reacciones de cada muestra biológica para los genes blanco se hicieron con la misma cantidad de templado utilizada en las reacciones para los genes normalizadores.

### 7.4.1 Diseño de oligonucleótidos para RT-qPCR

Debido al alto grado de homología (similitud del 63%) que presentan los genes de la subfamilia *PLT* en *A. thaliana*, para el análisis de la expresión de los genes *PpPLT* por RT-qPCR es necesario usar las regiones menos conservadas. Por ello, para el diseño de oligonucleótidos nos basamos en las regiones no traducidas de cada transcrito.

Se diseñaron oligonucleótidos que aparean con la región 3' no traducida de *PpBBM*, *PpAIL* y *PpPLT3c*. Para *PpPLT2b* y *PpPLT3b* se diseñaron oligonucleótidos sentido que aparean con la región codificante y reversos que aparean con la región 3' no traducida.

Tabla 4. Oligonucleótidos para el análisis de la expresión de *PpPLT* por RT-qPCR.

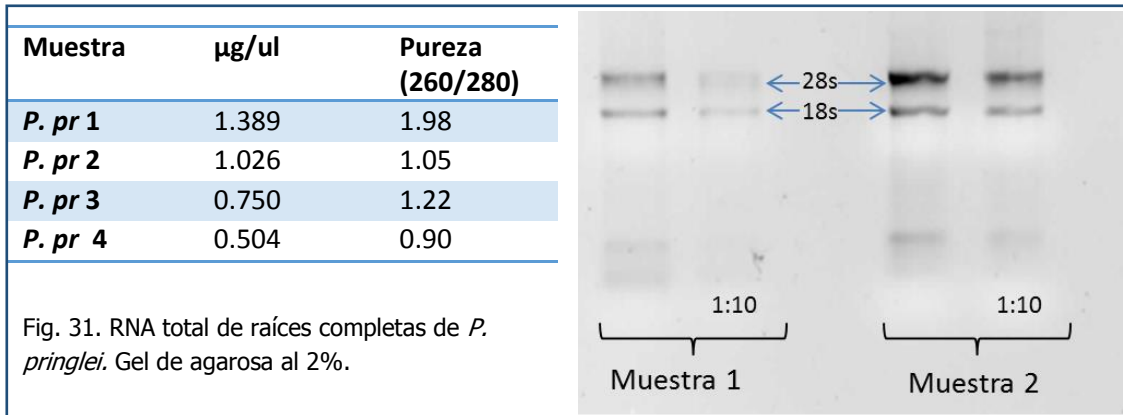
<i>PpPLT</i>	Secuencia
<i>PpBBM</i> reverse:	AAGGCAAAGGCTAGTCGCAC
<i>PpBBM</i> forward:	AGGAAAGGGAAAAGATCGATGC
<i>PpPLT2b</i> forward:	GGTGAAGGTAGATTATGACATGCC
<i>PpPLT2b</i> reverse:	CAACGTGGACCATCGCCAATC
<i>PpPLT3c</i> forward:	CGGAGTTAGGAAGGGCGTAG
<i>PpPLT3c</i> reverse:	GGTCAGTGAAAGAACCATTAGG
<i>PpAIL</i> forward:	TGCTCAAATGGCGCCTTTCC
<i>PpAIL</i> reverse:	CACACACATCCTTAAGTTTCTCC
<i>PpPLT3b</i> forward:	ACAATCTTGGCTGCAATGGTGG
<i>PpPLT3b</i> reverse:	TCTAGCTTCTTGGTTCTTCGGC

Para evitar amplificar productos de DNA inespecíficos, se comprobó que no existían áreas extensas de similitud de cada par de oligonucleótidos con otros *contigs* de *P.*

*pringlei*, mediante un ensayo BLASTn de cada uno de los amplicones contra la base de datos de *contigs*.

#### 7.4.2 Prueba de oligonucleótidos para RT-qPCR

Para realizar estas pruebas se purificó RNA total de raíces completas de 90 plantas de *P. pringlei* de entre 3 y 4 días de edad (etapa inicial 2, ver fig. 16). El RNA se extrajo con TRIzol, las concentraciones obtenidas se muestran en la Tabla 4. Las muestras de mayor concentración se analizaron por electroforesis, 1 µl de cada muestra de RNA con sus respectivas diluciones (1:10), determinar la calidad del RNA total (Fig. 31).



En el gel puede verse que en ambas muestras la banda del 28S es de mayor intensidad que la de 18S, comprobando así la integridad del RNA.

El RNA total de la muestra 1 se trató con DNasa I, y se sintetizó cDNA a partir de este. El cDNA se usó como templado en reacciones de PCR de punto final para probar los oligonucleótidos diseñados. Los productos de reacción se corrieron en un gel de agarosa 2 % (Fig. 32). Como control positivo se utilizaron los genes *PpAGL31* y *PpCLA*. Ambos genes se usaron como genes de referencia y previamente se utilizaron para amplificar los fragmentos de genes correspondientes en *S. gummosus* (Shishkova *et al.*, 2013).

Se amplificaron productos únicos del tamaño esperado para cada par de oligonucleótidos, lo que indica que estos aparentemente funcionan de manera correcta en el ensayo.

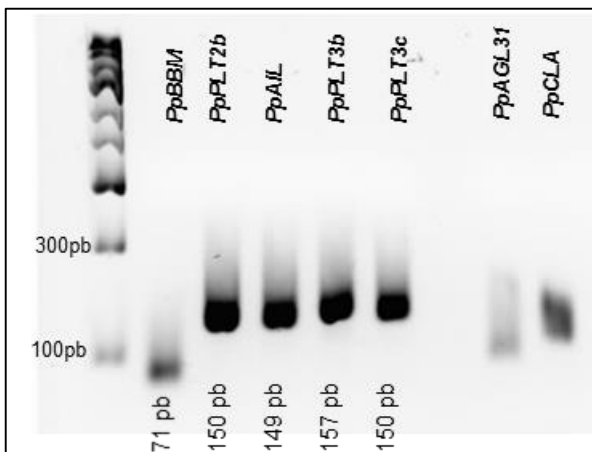


Fig. 32. Prueba de oligonucleótidos en PCR punto final. Los productos de PCR de los genes *PpBBM*, *PpPLT2b*, *PpAIL*, *PpPLT3b* y *PpPLT3c* se analizaron por electroforesis. Primer carril, marcador de peso molecular. En la figura se indican los tamaños esperados de las bandas. Control positivo: *PpAGL31* y *PpCLA*.

Los oligonucleótidos usados en reacciones de RT-qPCR deben cumplir además con ciertas características específicas: no deben generar dímeros y deben tener una eficiencia de reacción que permita calcular la cantidad de sustrato inicial (Pfaffl, 2004). Dichos parámetros se analizaron para cada par de oligos, los análisis se describen en el anexo III. En la tabla 5 se resumen los resultados de las curvas de eficiencia de oligonucleótidos y sus curvas de disociación correspondientes.

Tabla 5. Eficiencia de los oligonucleótidos diseñados para RT-qPCRs y presencia o ausencia de dímeros de oligos en las reacciones de amplificación.

**Genes de referencia:**

	Mejor eficiencia	repeticiones	Dimeros	Cq menores de 30
<i>PP2A: fosfatasa alcalina</i>	1.997	3	No	Sí
<i>CLA:</i>	2.019	3	No	Sí
<i>MAD: AGL 31</i>	1.978	2	a baja dilución	En 1:10 y 1:100

**Genes blanco:**

<i>PpPLT13: BBM</i>	2.305	2	Sí	En 1:10
<i>PpPLT2b</i>	1.995	3	No	Sí
<i>PpPLT21: PpPLT3c</i>	2.039	2	No	Sí
<i>PpPLT28: PpAIL</i>	1.9	4	NO	Sí
<i>PpPLT31: PpPLT3b</i>	1.903	3	No	Sí

En una reacción ideal de RT-qPCR la cantidad de producto se duplica con cada ciclo, y la eficiencia de los oligonucleótidos es de 2.0. El rango de eficiencia aceptable para los oligonucleótidos es de entre 1.85 y 2.15. De acuerdo con los resultados pueden usarse los oligonucleótidos diseñados y anotados anteriormente en nuestro laboratorio para los genes de referencia *PpCLA*, *PpAGL31* y *PpPP2A*. Para el gen *PpBBM* los oligonucleótidos probados no se pueden usar para cuantificar este gen, porque se generan dímeros, y esto interfiere con la cuantificación por medio del qPCR. Los oligonucleótidos *PpPLT2a*, *PpAIL* y *PpPLT3b* tienen una eficiencia óptima, y por lo tanto estos dos pares de oligonucleótidos para los genes *PpPLT* pueden utilizarse en reacciones de RT-qPCR.



### 7.4.3 Extracción de RNA total de muestras biológicas de puntas de la raíz primaria de *P. pringlei* en cuatro etapas del desarrollo para RT-qPCR

Se procesaron las muestras correspondientes a dos repeticiones biológicas y cuatro etapas del desarrollo de la raíz primaria (Tabla 6).

Tabla 6. Número de puntas de raíces que se incluyeron en cada muestra.

Etapa	Inicial 1	Inicial 2	Intermedia	Terminal
Muestra biológica A	345	395	470	515
Muestra biológica B	292	409	458	586

Se extrajo el RNA total de las muestras usando TRIzol. Se midió la pureza y concentración de cada muestra (Fig. 33).

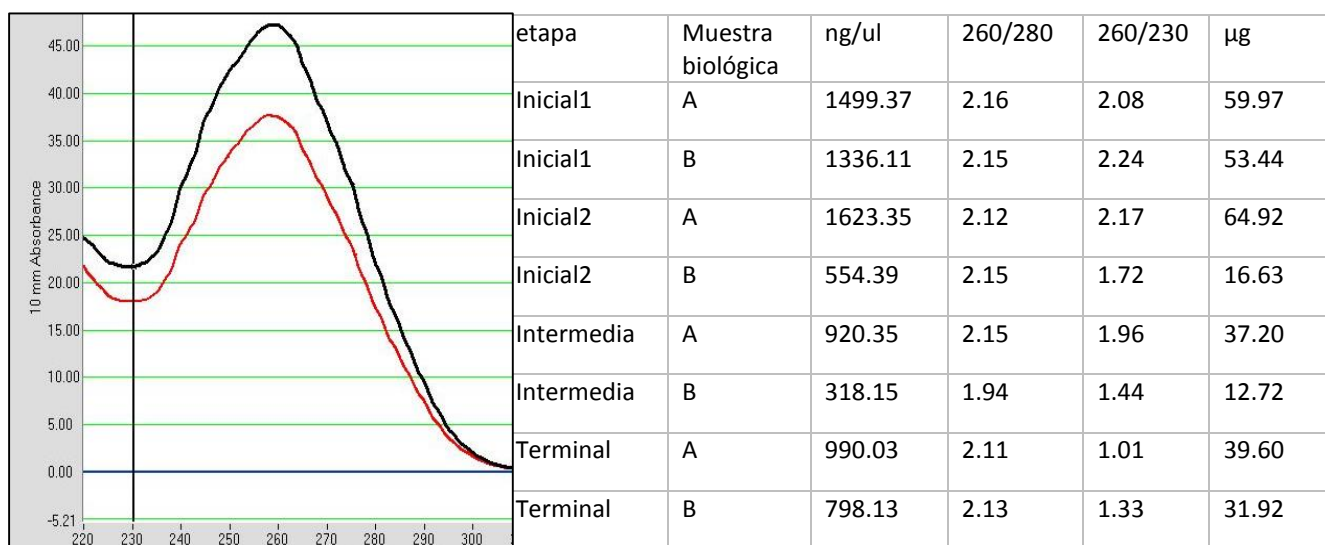


Fig. 33. RNA extraído de las puntas de raíces primarias de las diferentes etapas del desarrollo de *P. pringlei*. Panel izquierdo, curvas de absorbancia; se muestran como ejemplo las curvas de absorbancia de las etapas inicial 1 e inicial 2 de la muestra biológica B. Panel derecho, concentración de RNA total obtenido y cocientes de absorbancia de las muestras.

La cantidad y pureza del RNA fue ideal para realizar el análisis por RT-qPCR. La tasa de absorbancia 260/280 se usa como una medida para determinar la pureza del RNA obtenido. Un valor de  $\sim 2$  se acepta como RNA no contaminado con proteínas, valores menores a 1.8 indican la presencia de proteínas en la muestra. Para las muestras obtenidas, el índice es mayor a 2, lo que indica que no hay contaminantes (proteínas) que interfieran en reacciones posteriores.

La tasa de absorbancia 260/230, una medida secundaria que indica el enriquecimiento de ácidos nucleicos en la muestra; arrojó valores adecuados para la mitad de las muestras. Cuando el índice es menor a 2, como en las muestras de las etapas iniciales y terminales, puede indicar la presencia de cantidades más elevadas de polisacáridos,

que es común para RNA purificado de plantas y sobre todo de las especies de cactáceas.

Para evaluar la integridad del RNA, las muestras se analizaron por electroforesis en un gel de agarosa (1 %) (Fig. 34). La relación visual entre las bandas correspondientes al RNA ribosomal de la subunidad 28S y del 18S indican que la integridad del RNA contenido en las muestras es la esperada para poder realizar el ensayo de RT-qPCR.

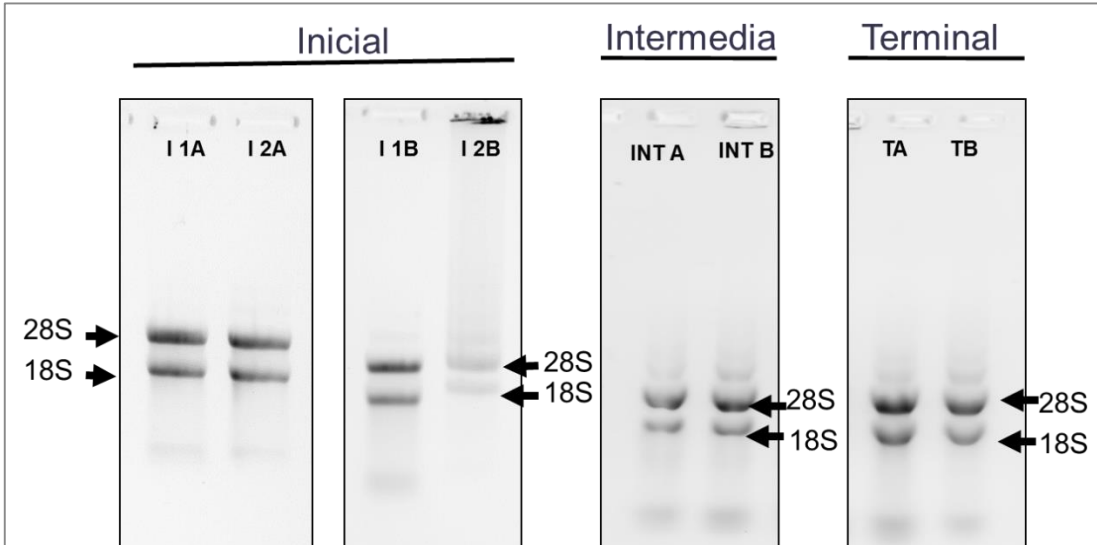


Fig. 34. Electroforesis en gel de agarosa (1%) de las muestras de RNA que se extrajeron a partir de la punta de la raíz primaria de *P. pringlei*. I1: etapa inicial 1. I2: etapa inicial 2. INT: etapa intermedia. T: etapa terminal. Las letras A y B corresponden a diferentes muestras biológicas. Flechas negras, RNA ribosomal de la subunidad 28S y 18S.

#### 7.4.5 PCR en Tiempo Real (RT-qPCR)

Como primera parte de los experimentos de RT-qPCR, se obtuvieron los valores de Cq amplificando los fragmentos de los tres genes de referencia y así se estimó la cantidad de cDNA en cada muestra biológica. Se utilizaron las dos muestras biológicas de las cuatro etapas del desarrollo de la raíz primaria; se hicieron tres réplicas técnicas de cada una de las reacciones. En la figura 35 se muestran los datos de Cqs obtenidos en estas reacciones. Cabe recordar que un menor valor de Cq indica una mayor concentración de cDNA en la muestra; y que la relación entre el Cq y la concentración de cDNA es exponencial, ya que el producto de la reacción se duplica en cada ciclo.

Los datos aquí mostrados sirven para estimar sí la cantidad de cDNA en las muestras es suficiente para poder cuantificar la abundancia de transcritos de manera confiable y en el caso de genes normalizadores servir como control interno de la cantidad de cDNA en reacciones de interés. Los Cqs deben ser menores a 35 para entrar dentro del rango cuantificable, y esta condición se cumplió para todos los genes y todas las muestras.

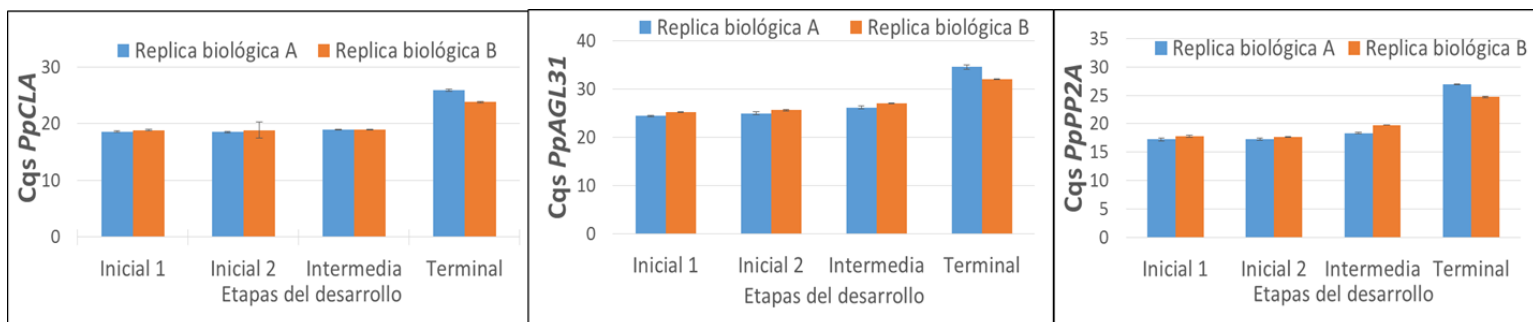


Fig. 35. Cqs de las reacciones de RT-qPCR para los genes de referencia *PpCLA*, *PpAGL31* y *PpPP2A*. Cada barra corresponde al Cq promedio de tres réplicas técnicas para cada una de las muestras biológicas  $\pm$  Desviación Estándar.

En cada una de las gráficas se observa el mismo comportamiento: en las primeras tres etapas el Cq y en consecuencia la concentración de cDNA en las muestras es casi constante, y en la última etapa el Cq es más alto. Un Cq es considerable más alto en la etapa terminal indica que la concentración de cDNA es menor en las muestras de esta etapa.

Los valores de Cqs promedio para los tres genes de referencia entran en el rango esperado para genes de referencia, es decir menor a 27 ciclos; excepto por el gen *PpAGL31* en la etapa terminal, que es de 33.25 ciclos, aunque entra dentro del rango confiable para su cuantificación.

Las diferencias en los valores de Cqs alcanzados por cada gen indican que *PpPP2A* es el mensajero más abundante de los tres genes de referencia, mientras que el gen *PpAGL31* tiene más baja expresión en puntas de la raíz.

Utilizando los datos presentados en la figura 35, se llevó a cabo la corrección de la cantidad de cDNA en cada una de las muestras biológicas. Para ello se calcula el factor de normalización, dado por la media geométrica de los Cqs de los genes de referencia de cada muestra biológica, en este caso se usaron los tres genes de referencia *PpCLA*, *PpAGL31* y *PpPP2A*.

Se realizó el análisis de RT-qPCR para los cuatro genes blanco seleccionados (*PpPLT2a*, *PpAIL*, *PpPLT3b* y *PpPLT3c*) en las cuatro etapas del desarrollo, con dos réplicas biológicas y tres réplicas técnicas de cada reacción. En la figura 36 se muestran los Cqs de estas reacciones.

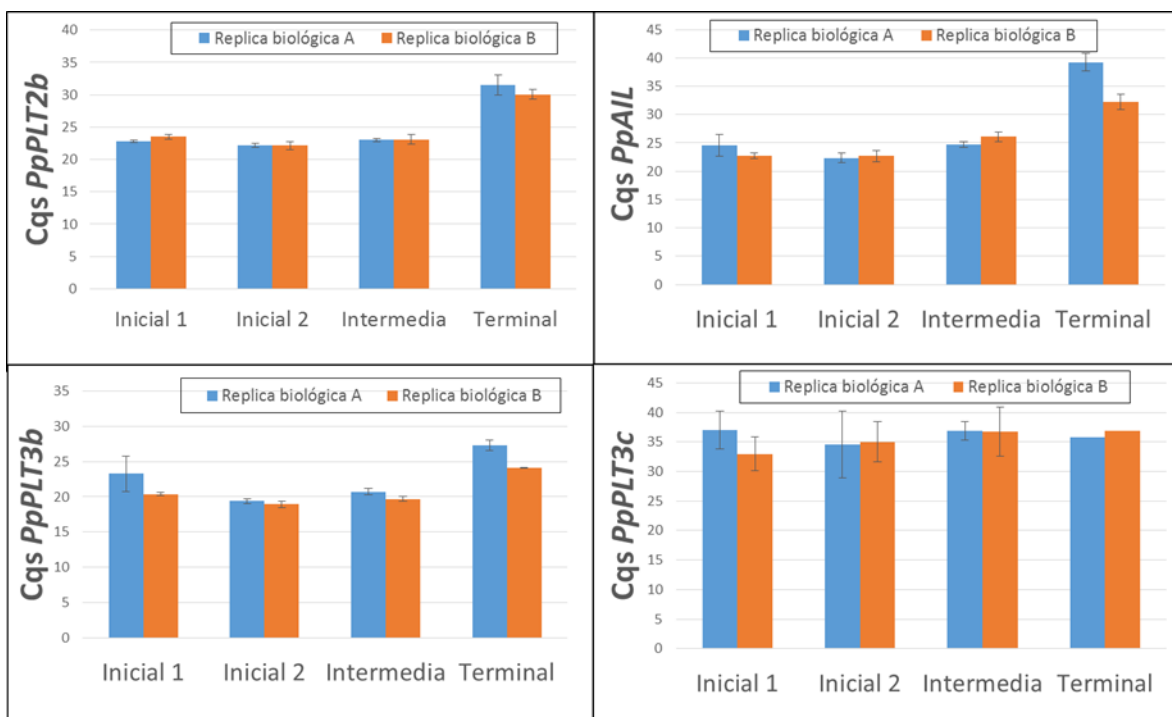


Fig. 36. Cq's de las reacciones de RT-qPCR para los genes blanco *PpPLT2b*, *PpAIL*, *PpPLT3b* y *PpPLT3c*. Cada barra corresponde al Cq promedio  $\pm$  Desviación Estándar de tres réplicas técnicas para cada una de las muestras biológicas.

Para los genes *PpPLT2b*, *PpAIL* y *PpPLT3b* los valores de Cq caen dentro del rango en que pueden ser medidos, excepto por la etapa terminal del gen *PpAIL* que sobre pasa los 35 ciclos. Este último resultado indica que su expresión para esta etapa en particular es muy baja.

Los valores de Cq obtenidos en las reacciones para el gen *PpPLT3c* fueron muy elevados, superan el rango de 35 ciclos; por lo que no puede cuantificarse la expresión de este gen.

Los datos presentados en la fig. 36 se normalizaron usando el método  $\Delta Ct$  y se corrigieron con el factor de normalización calculado con tres genes de referencia para cada muestra biológica. Los datos se presentan en razón del valor de la expresión relativa para la etapa inicial, con valor de 1, para cada gen de interés (figura 37).

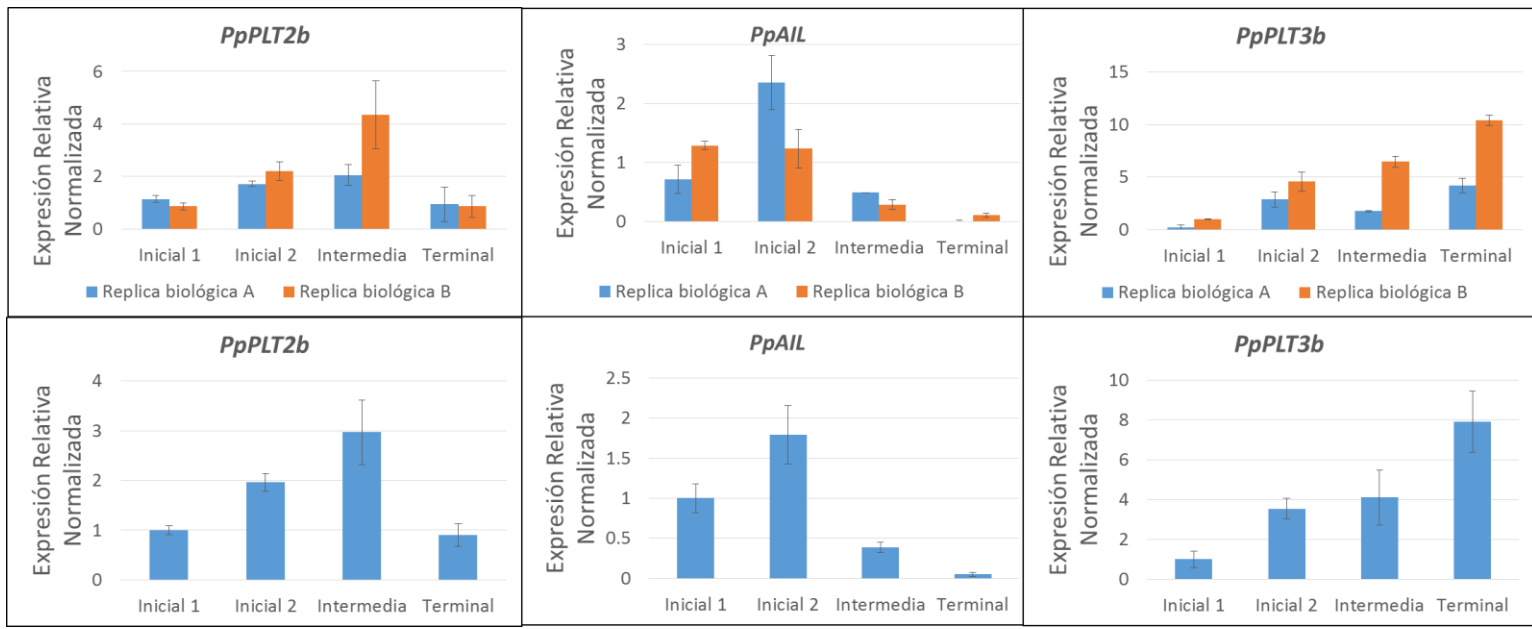


Fig. 37. Expresión relativa normalizada de los genes *PpPLT2b*, *PpAIL* y *PpPLT3b*. Datos de RT-qPCR normalizados usando tres genes de referencia (*PpCLA*, *PpAGL31* y *PpPP2A*). Se muestran los datos de tres réplicas técnicas. En los paneles superiores se graficaron los datos de las etapas del desarrollo para las dos réplicas biológicas utilizadas; En azul, la réplica biológica A; En anaranjado, la réplica biológica B. Se muestra promedio  $\pm$  la desviación estándar. En los paneles inferiores se muestra la expresión relativa promedio de ambas réplicas biológicas (azul). Se muestra el error estándar.

El gen *PpPLT2b* fue el único perteneciente al clado A (ver fig. 28) de los genes *PLT* cuya expresión fue eficientemente medida por RT-qPCR. Este gen presenta alta similitud tanto con *PLT1* como con *PLT2* de *A. thaliana*, que son necesarios para el mantenimiento del RAM. Por lo anterior se esperaba encontrar un decremento en la expresión en la etapa terminal en la que ya no hay un RAM y en la que no sería ya necesaria esta acción de mantenimiento de la actividad meristemática. El resultado obtenido fue consistente con esta teoría porque de la etapa intermedia a la terminal la expresión decayó a la mitad.

En este patrón de expresión puede observarse que, después de las etapas iniciales del desarrollo, hay un aumento en la expresión que llega a su máximo para la etapa intermedia. Es durante la etapa intermedia que hay una transición de las células no diferenciadas hacia células diferenciadas y que se pierde la actividad meristemática. Parece que hay un mecanismo que detona este aumento en la expresión de *PpPLT2b* mientras en la raíz existe un meristemo funcional.

Para *PpAIL* se ve una mayor expresión en las etapas iniciales y un decaimiento hacia las etapas intermedia y terminal, en esta última la expresión es apenas cuantificable (Fig. 37, segundo panel superior). El gen *PpAIL* guarda una mayor similitud con el gen *AtAIL*, sin embargo, la función de este gen no ha sido aún descrita. El otro gen con el que tiene una alta similitud es *AtANT*. El papel principal del gen *AtANT* es el del establecimiento de meristemo primario en la parte aérea de la planta. Debido a la alta

similitud con *AtANT*, el gen *PpAIL* podría cumplir la función de establecer el RAM en la raíz primaria durante las primeras etapas del desarrollo, lo cual es consistente con la caída de su expresión en la etapa terminal.

*PpPLT3b* muestra un aumento desde la etapa inicial hasta la etapa terminal del desarrollo de la raíz primaria del cardón. Este gen tiene mayor similitud con genes que tienen muy baja expresión en el RAM, pero se expresan a lo largo de la raíz; el aumento en la expresión de *PpPLT3b* podría explicarse por la diferenciación que ocurre en las células de la punta de la raíz, ya que en *A. thaliana* *PLT3* y *PLT7* no se expresan en el RAM, pero sí a lo largo de la parte de la raíz que contiene células diferenciadas.

## 7.5 RACE-PCR

### 7.5.4 Síntesis de población de cDNAs con extremo modificado

Con el fin de obtener las secuencias completas de dichos transcritos, se empleó la estrategia RACE-PCR (amplificación de los extremos de cDNA por PCR). Este método implica la adición de una secuencia conocida a un extremo del mensajero, misma que se usa para poder amplificar el mensajero desde su parte interna conocida y el extremo modificado en una reacción de PCR.

Para generar las poblaciones de cDNA con extremos modificados se utilizó RNA de raíces completas de *P. pringlei* como templado. El RNA se expuso a una temperatura de 37°C por dos horas para probar su estabilidad. En un gel de electroforesis se cargaron 500 ng del RNA antes y después del tratamiento (Fig. 38)

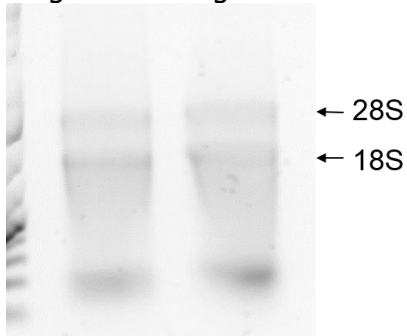


Fig. 38. RNA total de raíces de *P. pringlei*. Gel de electroforesis agarosa 1%. Carril der. RNA total expuesto a 37°C por 2 horas, carril izq., RNA sin tratar.

En la imagen que se muestra (figura 38) pueden observarse las bandas que corresponden a la subunidad 28S y 18S rRNA. Ambas son igualmente tenues, incluso la banda de 18S rRNA se distingue un poco más intensa que la banda 28S rRNA, lo que indica que la integridad del RNA total no es buena. Además, puede observarse un barrido y una banda de menor tamaño que aumenta en la muestra expuesta a 37 °C. Tanto el barrido como la banda de menor tamaño indican que hay degradación del RNA.

Pese a ello, y debido a la escasez de material biológico, se realizaron las reacciones con esta muestra de RNA como templado para generar la población de cDNAs con extremo 5' modificado y la población de cDNAs con extremo 3' modificado.

### 7.5.5 RACE-PCR 5'

Se realizó la reacción primaria para los seis genes de interés con la enzima Advantage2 de Clontech. Se corrió el producto de las reacciones en un gel de electroforesis como se ha descrito previamente. La cantidad de producto resultante de esta amplificación fue baja, ya que no se observaron bandas en el gel. Sin embargo se usó el producto de esta reacción como templado en las reacciones secundarias de los genes: *PpPLT2a*, *PpPLT2b*, *PpPLT3b*.

Como controles positivos de la reacción se usó un par de oligonucleótidos para el gen *PpPLT2a* que amplifica 150 pb de la región UTR 3'; oligonucleótidos para el mismo gen que amplifican una región de 1.3 kb cuando se usa como templado cDNA y 2.5 kb cuando se amplifica DNA (debido a los intrones que contiene el gen).

Para asegurarnos de que no hubiera productos inespecíficos debidos al uso de los oligonucleótidos universales, como control negativo, se realizó una reacción en la que no se incluyeron oligonucleótidos específicos de los genes de interés, sino únicamente los oligonucleótidos universales (*long*, *short* y *nested*).

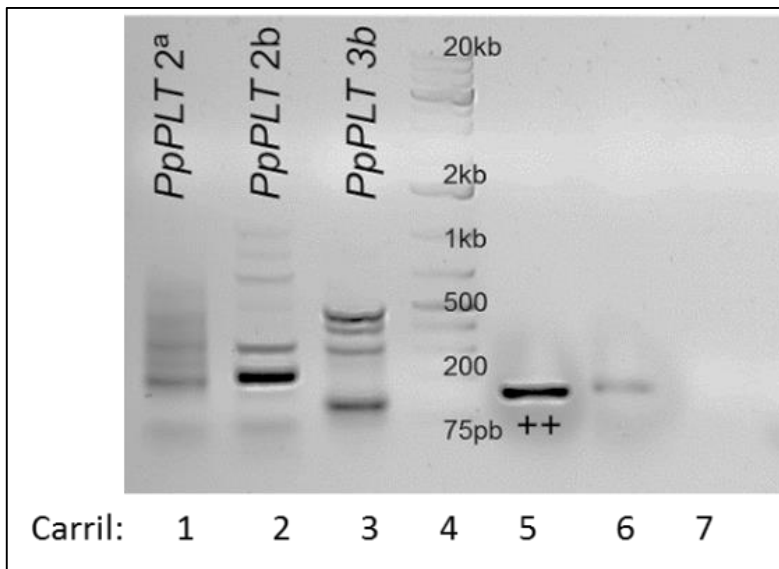


Fig. 39. Productos de la reacción secundaria 5'-RACE-PCR. Carril 1: *PpPLT2a*. Carril 2: *PpPLT2b*. Carril 3: *PpPLT3b*. Carril 4: marcador molecular. Carril 5: control positivo *PpPLT2a* 150pb. Carril 6: Control positivo con oligonucleótidos para *PpPLT2a*, banda esperada 1.3kb. Carril 7: Control negativo sin templado.

Para los tres cDNAs en la reacción anidada de PCR se obtuvieron varias bandas (fig. 39), todas menores a 1 kb. Sólo para el gen *PpPLT2b* se observa una banda mayoritaria, pero parece ser menor al tamaño mínimo esperado (229 pb). Por el control negativo sabemos que estas bandas no son generadas por el uso de los oligonucleótidos universales. El control positivo de amplificación del fragmento de 150

pb del gen *PpPLT2a* indica que todos los reactivos funcionan bien. Sin embargo, en el control positivo para *PpPLT2a* de 1.3 kb no hubo amplificación de la banda esperada.

Esto podría indicar que la integridad del cDNA no es suficiente para que se amplifique una región de un transcrito de mucho mayor tamaño que 150 pb y podría ser resultado de la baja calidad del RNA total usado para generar la población de cDNA con extremo 5' modificado. Al mismo tiempo, la baja calidad de la muestra utilizada puede provocar que se amplifiquen bandas provenientes de cDNA de transcritos incompletos de los genes de interés.

### 7.5.6 RACE-PCR 3'

En este caso se hizo también la reacción primaria para los seis genes de interés con la enzima de alta fidelidad *Advantage2* de Clontech. Se analizó el producto de las reacciones en un gel de agarosa y no se observaron bandas en el gel (foto no mostrada).

Como la enzima *Advantage2* ya había caducado, estas reacciones se repitieron con la enzima recién comprada *TaqPol* de ThermoScientific, que no es de alta fidelidad. Con esta enzima sí se amplificaron bandas de bajo peso molecular en la reacción primaria (Figura 40).

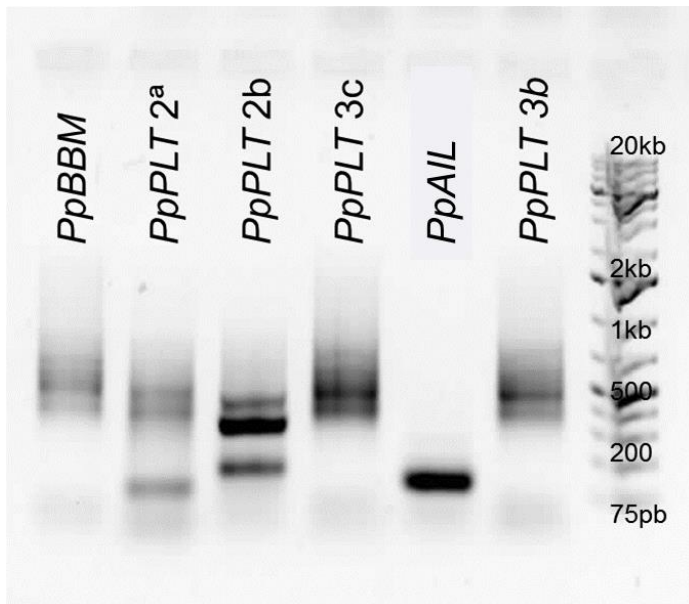


Fig. 40. Productos de la reacción primaria de 3'-RACE-PCR con la enzima *TaqPol* de ThermoScientific.

Para el gen *PpAIL* se amplificó una banda mayoritaria y pocas bandas adicionales. La secuencia del *contig* de ese gen contiene 200 pb de la región UTR 3', por lo que esperamos una banda mayor a 200 pb. El producto de amplificación es de menor tamaño, por lo que podría tratarse de un producto inespecífico.



Para el gen *PpPLT2b* esperábamos amplificar una banda de más de 200 pb; se observa una banda mayoritaria de aprox. 300 pb (Figura 44), pero también se observa un barrido. Para el resto de los genes se observa un barrido en el gel, lo que indica que se trata de productos inespecíficos. Es por esto que para la reacción secundaria no se usaron estos productos de reacción sino los generados con la enzima *Advantage2*, aunque no alcanzaban a visualizarse en un gel.

Se hicieron las reacciones secundarias para los genes: *PpPLT2a*, *PpPLT2b*, *PpPLT3b* con la enzima *Advantage2* y se corrieron en un gel de electroforesis (Fig. 41).

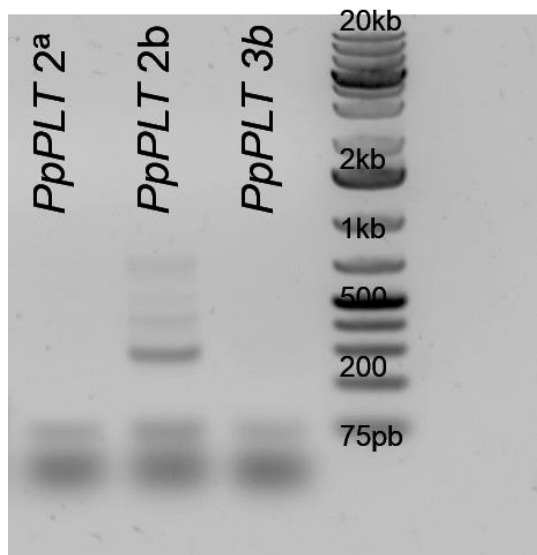


Fig. 41. Productos de la reacción secundaria 3'-RACE-PCR con la enzima *Advantage2*.

Para el gen *PpPLT2b* se obtuvo un patrón de bandas muy similar al de la reacción primaria para el mismo gen con la enzima de Thermo Scientific, con una banda mayoritaria de tamaño menor a 300 bp. De tratarse del producto específico de la reacción nos adicionaría 100 pb más del UTR de la secuencia de este mRNA. Pero la cantidad de bandas sugiere que son productos inespecíficos. Para los otros dos genes (*PpPLT2a* y *PpPLT3b*) no hubo amplificación. Al igual que con los productos de las reacciones 5' podríamos estar teniendo productos inespecíficos o falta de amplificación debido a la baja integridad de la muestra de RNA. Por ello, para obtener las secuencias completas de los transcritos PpPLT por RACE-PCR, sería necesario aislar nuevamente RNA de *P. pringlei* y repetir estas reacciones. Sin embargo, en el trabajo reciente del laboratorio, cuando ya estuvo terminada la parte experimental de esta tesis, se obtuvo mucho mayor número de lecturas más largas y apareadas de mRNA-seq de muestras de puntas de la raíz de *P. pringlei* (appr. 280 mln), lo que permitió un segundo ensamblaje *de novo* del transcriptoma de puntas de la raíz de *P. pringlei* muy superior al que utilizamos en este trabajo. Para la mayoría de los genes que nos interesan, los *contigs* contienen secuencias de UTRs 5' y 3', lo cual eliminó la necesidad de amplificar los extremos en reacciones de RACE-PCR.

## 8. Nuevo mRNA-seq

El análisis de las secuencias de *contigs* y datos del segundo ensamblaje provenientes del mRNA-seq que se usaron en este trabajo de tesis indicó que varios de los transcritos no se ensamblaron completamente. Entre esos transcritos incompletos se encontraron que codificaban para los genes *PLT*, de los cuales el único que contó con la secuencia completa de la secuencia codificante era un posible *contig* quimérico. Por todo lo anterior, para obtener el transcriptoma más completo, fue necesario incluir nuevas muestras biológicas y obtener una mayor cantidad de lecturas de mRNA-seq. Recientemente, y de manera posterior a la realización del trabajo experimental de esta tesis, por otros integrantes del laboratorio se llevó a cabo el ensamblaje *de novo* de lecturas del nuevo mRNA-seq. Nos dimos la tarea de analizar los genes *PpPLT* obtenidos en este nuevo mRNA-seq, lo que se describirá en este capítulo.

Análisis preliminares mediante RT-qPCR de los niveles de expresión de algunos genes en etapas intermedias del crecimiento de la raíz, sugerían la existencia de cambios en la abundancia de algunos transcritos durante dicha etapa en comparación con las etapas inicial y terminal. Por esta razón, se tomó la decisión de secuenciar muestras de las tres etapas. En el nuevo análisis por mRNA-seq se incluyeron además dos réplicas biológicas de cada etapa. La inclusión de las réplicas biológicas representó diversas ventajas, entre ellas la posibilidad de analizar estadísticamente la expresión diferencial virtual de los *contigs* ensamblados.

En promedio, para cada una de las seis muestras se obtuvieron 46 millones de lecturas que se generaron mediante el método de *paired-ends* o extremos apareados, es decir, las bibliotecas de cDNA se secuenciaron en las direcciones 3'→5' y 5'→3'. En contraste con la secuenciación anterior, en esta ocasión se contó con un mayor número de lecturas, de mayor longitud y de mejor calidad. Por ejemplo, las lecturas se secuenciaron durante 100 ciclos y por el método de *paired-ends*, en contraste con la secuenciación *single-end* y de 85 ciclos que se utilizó para la secuenciación anterior.

Durante la secuenciación previa se generaron ca. 21 millones de lecturas, las cuales provienen de las etapas inicial y terminal del crecimiento de la raíz primaria, sin réplicas biológicas. Para evitar el ensamblaje de los *contigs* quiméricos, se eliminaron las lecturas repetidas, lo que redujo todavía el número de lecturas utilizadas para ensamblar los *contigs*. En la secuenciación de las nuevas muestras biológicas se obtuvieron aproximadamente 280 millones de lecturas, las cuales provienen de las etapas inicial, intermedia y terminal del crecimiento de la raíz primaria, cada una con dos réplicas biológicas. La secuenciación de ambos extremos de fragmentos permitió utilizar para el ensamblaje todas las lecturas, sin la necesidad de eliminar las lecturas repetidas.

El tener un mayor número de lecturas, resultó en *contigs* mejor ensamblados, con una longitud promedio mayor y con una mejor cobertura. Esto permitió obtener un número considerablemente mayor de *contigs* que incluyen el marco de lectura completo y al

menos parte de las secuencias 5' y 3' no traducidas, reducir el número de *contigs* quiméricos y mejorar la anotación funcional de los transcritos de *P. pringlei*.

Se realizaron diferentes ensamblajes al utilizar distintos valores para los parámetros "Tamaño de palabra" y "Tamaño de Burbuja". De los ensamblajes generados se seleccionó el mejor, uno de los criterios principales para esta selección fue el porcentaje de lecturas de ambos extremos incluido en el ensamblaje teniendo en cuenta también que se maximizara el tamaño promedio de los *contigs* ensamblados.

### 8.1 *PLT* en el nuevo ensamblaje del transcriptoma

Enfocándonos en el análisis de los transcritos de los genes *PLT*, buscamos dichos genes entre los nuevos *contigs* ensamblados. Para identificar los *contigs* con similitud a los genes *AtPLT* se realizó el análisis BLASTx de dichas secuencias contra la base de datos de las secuencias del ensamblaje del último mRNA-seq. Como resultado, en esta ocasión se encontraron 7 *contigs* con similitud a *PLT*, uno más que en la secuenciación previa. Anteriormente contábamos con 6 *contigs* que no contenían las secuencias codificantes completas y faltaban los extremos UTR 5'. En esta ocasión, en los resultados del ensamblaje las secuencias de los *contigs* para los mismos 6 genes contenían las secuencias codificantes completas, además de los UTRs de extremos 5' y 3'. Sin embargo el nuevo *contig* encontrado con similitud a *PLT* no contiene la secuencia codificante completa. En la Tabla 6 se muestra la longitud de los *contigs* encontrados.

Tabla 6. Longitud de *contigs* de los genes *PpPLT* obtenidos en la segunda secuenciación

Contig	Longitud (pb)	UTR 5' (pb)	Secuencia Codificante(pb)	UTR 3' (pb)	Longitud en secuenciación previa (pb)
PpPLT2a	2475	360	1758	357	944
PpPLT2b	2249	86	1703	460	1195
PpBBM	2737	288	2064	385	1315
PpPLT3b	2288	498	1533	266	3762*
PpPLT3c	2484	542	1536	406	1315
<i>Contig</i> 0026499	679	-	679 (incompleto)	-	-
PpAIL	2036	410	1374	252	1506

La comparación entre las nuevas secuencias de *PpPLT* y las secuencias que se utilizaron durante este trabajo de tesis mostró que las secuencias corresponden a los mismos 6 genes (*PpPLT2a*, *PLT2b*, *PLT3b*, *PLT3c*, *PpBBM* y *PpAIL*) con los que se habían estado trabajando por lo que se conservaron los nombres dados a los *contigs*. La comparación entre estas secuencias se muestra en los alineamientos de la figura 42.

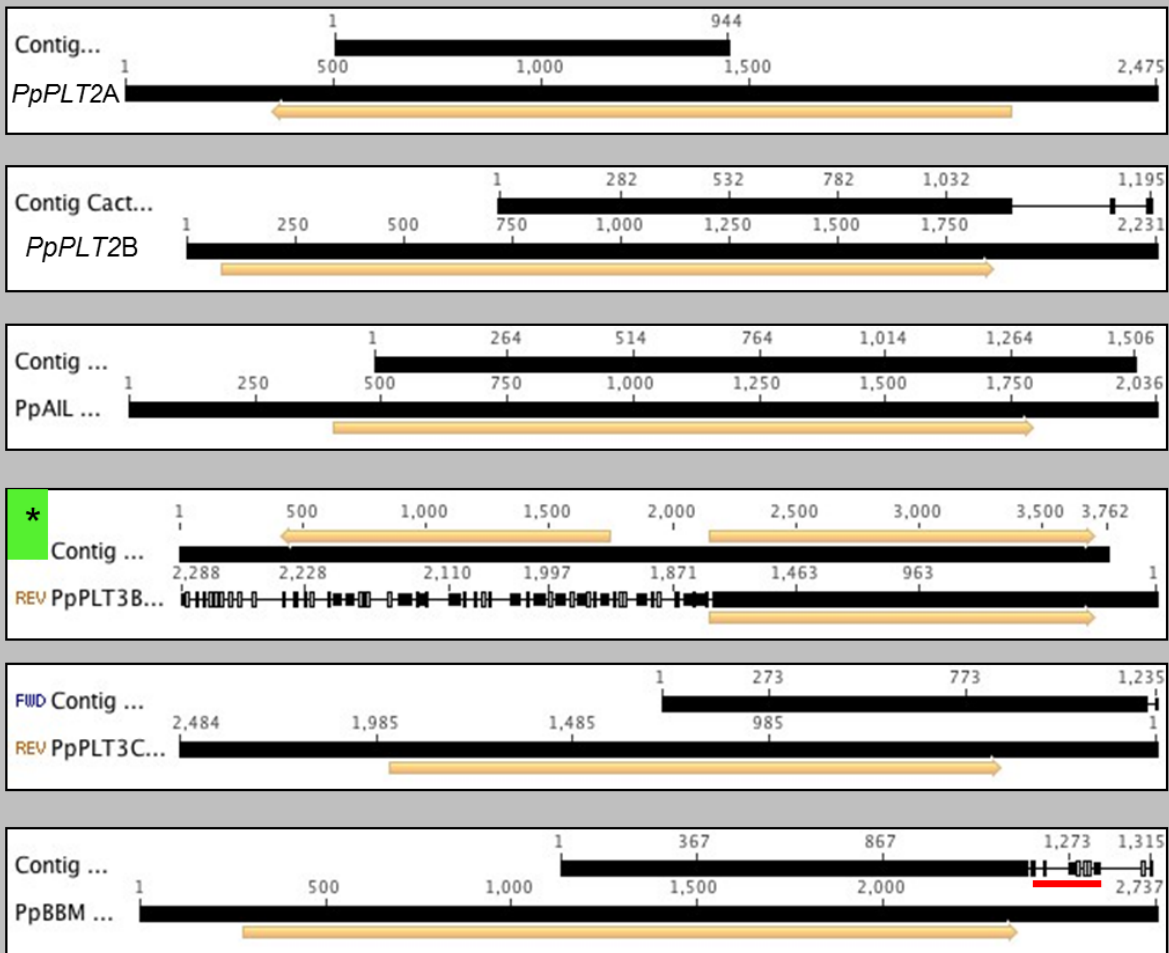


Fig. 42. Alineamiento de secuencias de los *PpPLT* del mRNA-seq anterior *Vs.* secuencias del nuevo mRNA-seq. En cada panel se representan los diferentes *PpPLT*. Para cada uno de los *contigs* se muestran dos barras negras, la barra superior representa al *contig* de la secuenciación previa utilizada en este trabajo de tesis, y la inferior al de la nueva secuenciación. Con flechas amarillas se representa la secuencia codificante.

En cada panel se representan los diferentes *PpPLT*. Si observamos como ejemplo el alineamiento de *PpPLT2a*, en este ambas secuencias se alinean perfectamente por lo que las barras negras son continuas, lo que significa que no existe ningún *gap* en el alineamiento. Son muy pocos los errores que se encuentran entre las secuencias utilizadas en esta tesis y las nuevas secuencias, los cuales pueden reflejar variaciones entre diferentes individuos de la población de plantas utilizadas para obtener muestras biológicas. Esto confirma que la calidad del ensamblaje que se utilizó en este trabajo de tesis fue muy buena. Sin embargo esta vez, gracias a todas las mejoras en la secuenciación y en la calidad y cantidad de las muestras biológicas utilizadas, pudo reconstruirse la secuencia completa de los genes de interés.

Una de las correcciones que apareció con la nueva secuenciación muestra un cambio de unas 15 pb en la región UTR 3' del gen *PpBBM*, aunque se trata de pocas bases, es justo en esta región en donde se había diseñado uno de los oligonucleótidos para

cuantificar la abundancia del transcrito por RT-qPCR, el cual presentaba más de un producto de amplificación.

En la secuenciación previa el *contig PpPLT3b* parecía ser quimérico, es decir que durante su reconstrucción parecían haberse unido dos *contigs* no relacionados. Esto fue algo que pudo confirmarse con la nueva secuencia obtenida. La secuencia del último mRNA-seq para este *contig* coincide perfectamente con la secuencia codificante del *contig* de la secuenciación previa, y esa es la única sección en la que ambos se alinean (Fig. 43). En la figura que corresponde a dicho alineamiento se observan muchos *gaps* justamente en el extremo UTR 5', esto debido a que la nueva secuencia no coincide con la parte quimérica del *contig* anterior.

El nuevo *contig* tiene un UTR 5' de 498 pb. La región UTR 5' del *contig* actual y del anterior muestra similitud en 20 pb, dicha similitud debió originar el ensamblaje erróneo en el *contig* quimérico que se tenía anteriormente.

En la nueva secuenciación apareció un nuevo *contig* que tiene una alta similitud con *AtPLT5*, como lo muestran los resultados del BLASTx en la base de datos de RefSeq. Sin embargo la secuencia de este *contig* es muy corta, 679 pb, no contiene la secuencia codificante completa, y no podríamos afirmar que este es efectivamente el homólogo de dicho gen en *P. pringlei*.

Usando las secuencias de las proteínas PLT y AIL de *A. thaliana* y *P. pringlei* se construyó la siguiente filogenia (Fig. 43).

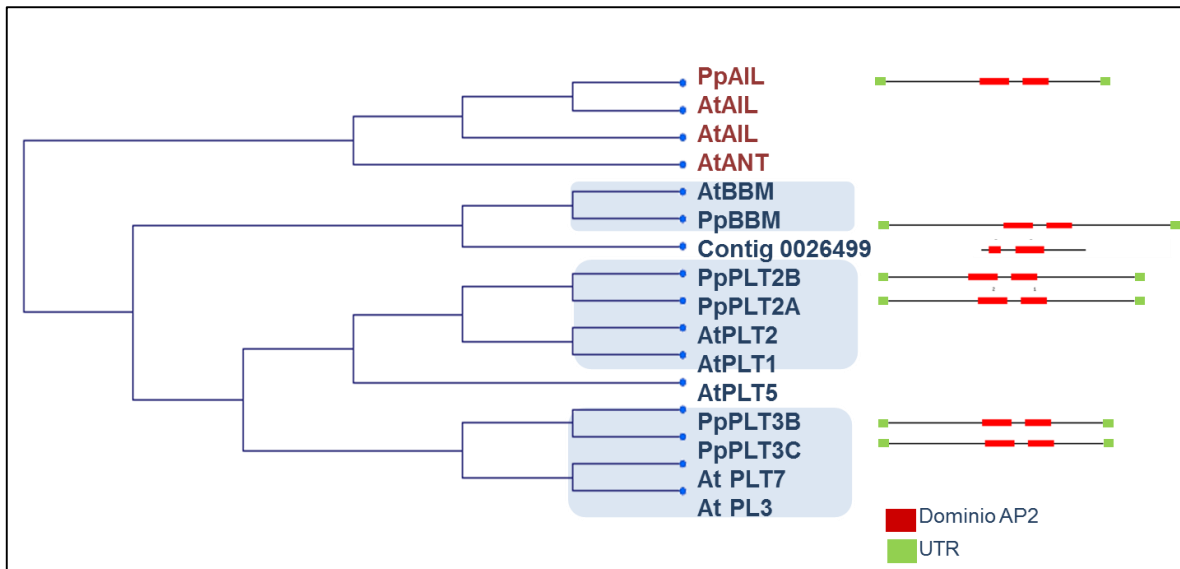


Fig. 43. Árbol filogenético de proteínas PLT y AIL de *A. thaliana* y *P. pringlei*. Del lado derecho se muestra una representación gráfica los *contigs* ensamblados.

Las proteínas PLT de *P. pringlei* en este análisis filogenético se agrupan con sus posibles homólogos más cercanas en *A. thaliana*, este resultado es consistente con los resultados de las filogenias que fueron construidas con los *contigs* de secuencias incompletas (Fig. 22).

El nuevo *contig* encontrado se ubica dentro de la rama de las proteínas BBM, aunque, una vez más, este no puede tomarse como un resultado concluyente porque el tamaño de dicho *contig* que no tiene la región codificante completa.

El hecho de que AtPLT5 fuese la única proteína de *A. thaliana* que en este análisis no se agrupó con uno de los *contigs* de *P. pringlei* sigue apuntando a la posibilidad de que el nuevo *contig* encontrado pudiese ser su homólogo. Además como se mencionó anteriormente, este *contig* mostró alta similitud con proteínas PLT5 de diferentes especies.

## 8.2 Expresión virtual de los *contigs* de *PpPLT* del último mRNA-seq

Los resultados de expresión virtual de los *contigs* de *PpPLT* obtenidos del último mRNA-seq se muestran a continuación (Fig. 44). Los datos corresponden a los promedios normalizados de las dos muestras biológicas y la abundancia está expresada en RPKMs (lecturas por kilobase por millón) para cada una de las etapas: inicial, intermedia y terminal.

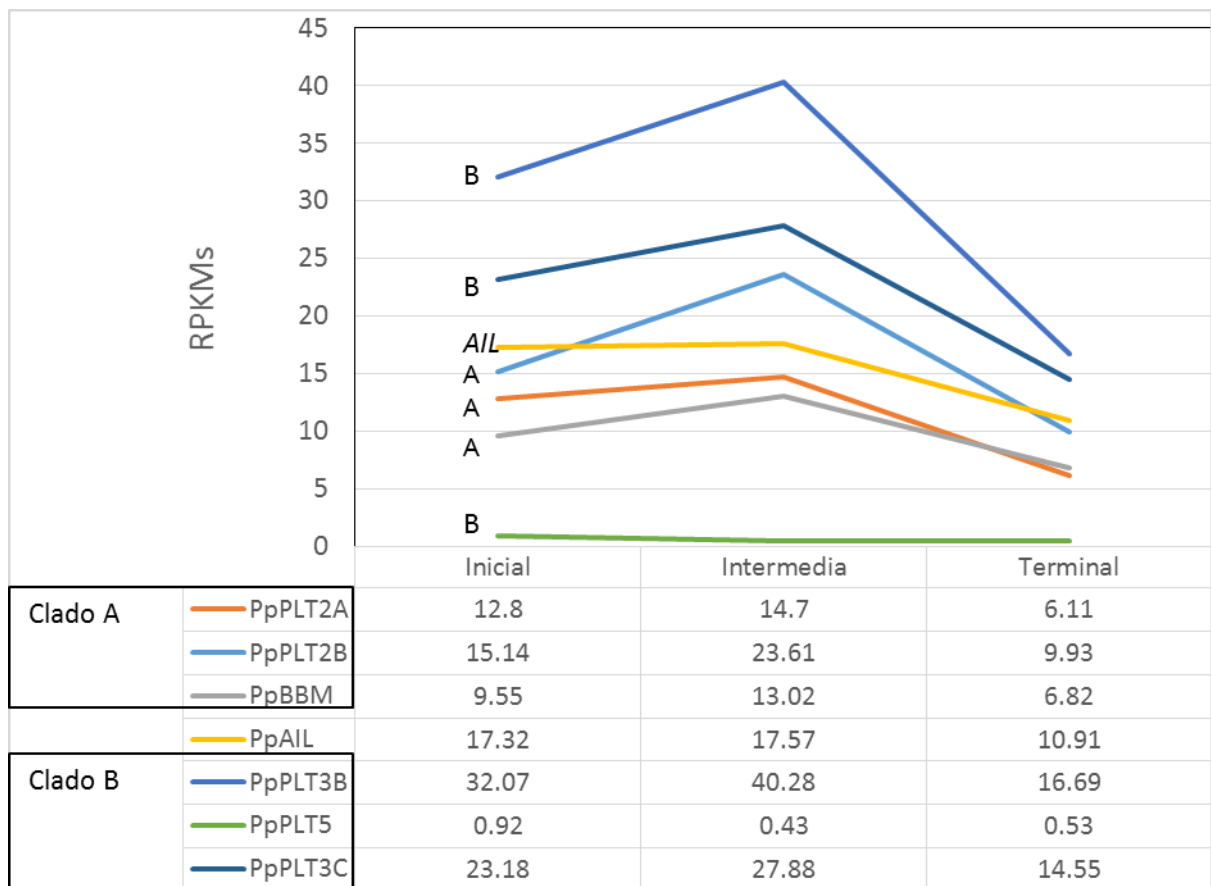


Fig. 44. Expresión virtual de los *contigs* de *PpPLT* y *PpAIL* del último mRNA-seq.

Los resultados revelan que existe un pequeño aumento en la expresión de la mayoría de los genes *PLT* en la etapa intermedia del desarrollo de la raíz. Aun así, los valores

RPKM para los *contigs* con similitud a los genes del clado A de *PLT* (*PLT2a*, *PLT2b* y *BBM*) varían entre 6 y 23, lo que indica una abundancia relativamente baja del transcrito de estos genes. Los valores más altos de RPKMs para estos *contigs* se observan en la etapa intermedia, en la que van de 13 a 23 RPKMs. Si comparamos el nivel de expresión de estos genes con los del resto de los transcritos ensamblados, estos se encuentran entre los más bajos.

Recordemos que en *A. thaliana* los genes pertenecientes al clado A de *PLT* se expresan más fuertemente en el RAM, en donde forman gradientes. En cambio la expresión de los genes del clado B de *PLT* no está únicamente localizada en el RAM, sino que se extiende a lo largo de la raíz (Fig. 25). De acuerdo a nuestros datos los genes del Clado A de *PLT* tienen una expresión menor a la de genes del Clado B en la punta de la raíz de *P. pringlei* en todas la etapas del desarrollo analizadas. Este es el resultado que se esperaba obtener en la etapa en la que el meristemo ya se ha agotado, o en tejido diferenciado, pero encontramos dicho comportamiento desde la etapa inicial, en la que el meristemo está presente aún.

El hecho de que el nivel de expresión de los genes *PLT2a*, *PLT2b* y *BBM* no alcance el nivel de expresión de genes del Clado B sugiere que la expresión de dichos genes puede no ser suficiente para alcanzar los umbrales necesarios en el mantenimiento del RAM.

Normalizando los valores RPKM sobre la etapa terminal, encontramos que los genes *PLT* se expresan entre 1.6 y 2.4 veces más en la etapa intermedia respecto a la etapa terminal (Fig. 45) y entre 1.4 y 2.15 veces más en la etapa inicial respecto a la etapa terminal.

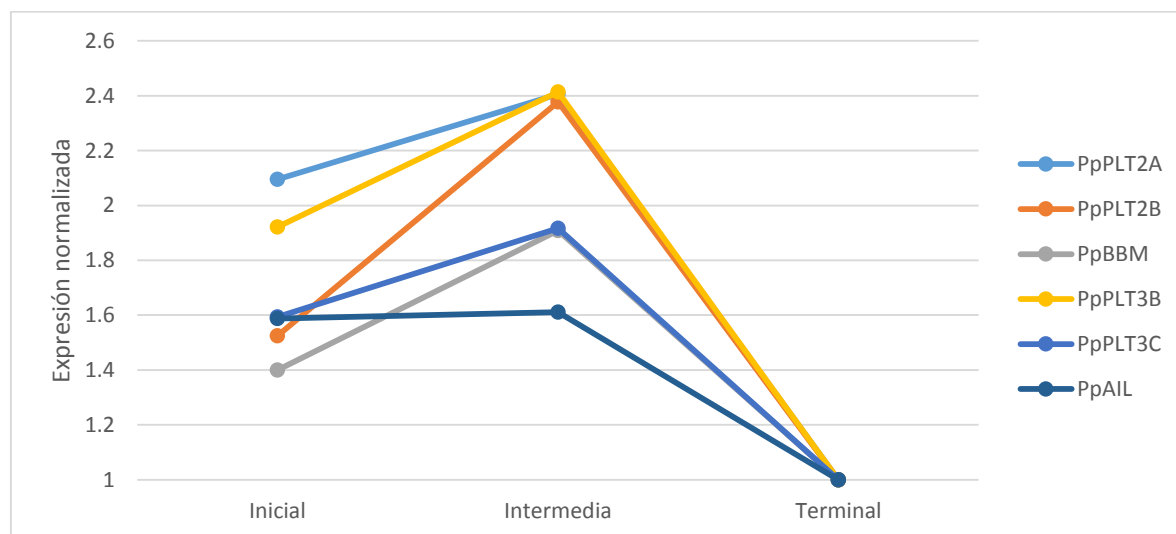


Fig. 45. Expresión virtual de los *contigs* de *PpPLT* del último mRNA-seq, normalizada a 1 en la etapa terminal.

La expresión virtual de estos genes en la secuenciación previa, mostró un cambio mayor, de entre 3 y 10 veces más en la etapa inicial con respecto a la etapa terminal (Fig. 30). La misma tendencia de disminución se encontró para la mayoría de los

contigs, tanto más abundantes como menos abundantes, en la etapa inicial de desarrollo (datos no mostrados).

Al comparar los resultados de la secuenciación actual contra la anterior (Fig. 46), se puede observar el mismo comportamiento en la expresión de los seis genes analizados, es decir, los valores de RPKM en la etapa inicial son mayores a los de la etapa terminal.

Debemos tomar en cuenta que resultados de la última ronda de secuenciación están respaldados por dos replicas biológicas, y que las correcciones técnicas realizadas para el presente ensamblaje mejoraron la calidad de las secuencias, por lo que los valores RPKM indicadores de la abundancia de transcritos son más confiables. Además, los resultados obtenidos respaldan el patrón de expresión documentado que se obtuvo en la primera ronda de secuenciación, es decir, una mayor abundancia de transcritos de todos los genes *PpPLT* en la etapa inicial en comparación con la etapa terminal.

Las diferencias cuantitativas entre los resultados de ambos eventos de secuenciación pueden explicarse como el resultado de las diferencias en la calidad de las lecturas y del método de ensamblaje que se usó anteriormente.

En la etapa terminal la expresión de los genes es superior a lo que se había estimado en la secuenciación previa, de más del doble para la mayoría de los genes *PLT*. Al parecer se subestimó la cantidad del transcrito presente en la etapa terminal, que podría ser consecuencia de la baja calidad del RNA de la muestra de esta etapa, por lo que, partiendo de la misma cantidad de RNA de las etapas inicial y terminal, de esta última se obtuvieron dos veces menos de lecturas, las cuales que tuvieron una peor calidad que las de etapa inicial. Cabe mencionar que en el segundo mRNA-seq, probablemente por el avance técnico, aunque la calidad de RNA de la etapa terminal fue peor, el número y la calidad de lecturas de todas las muestras fue comparable. Como consecuencia, los valores de expresión virtual también son más confiables.



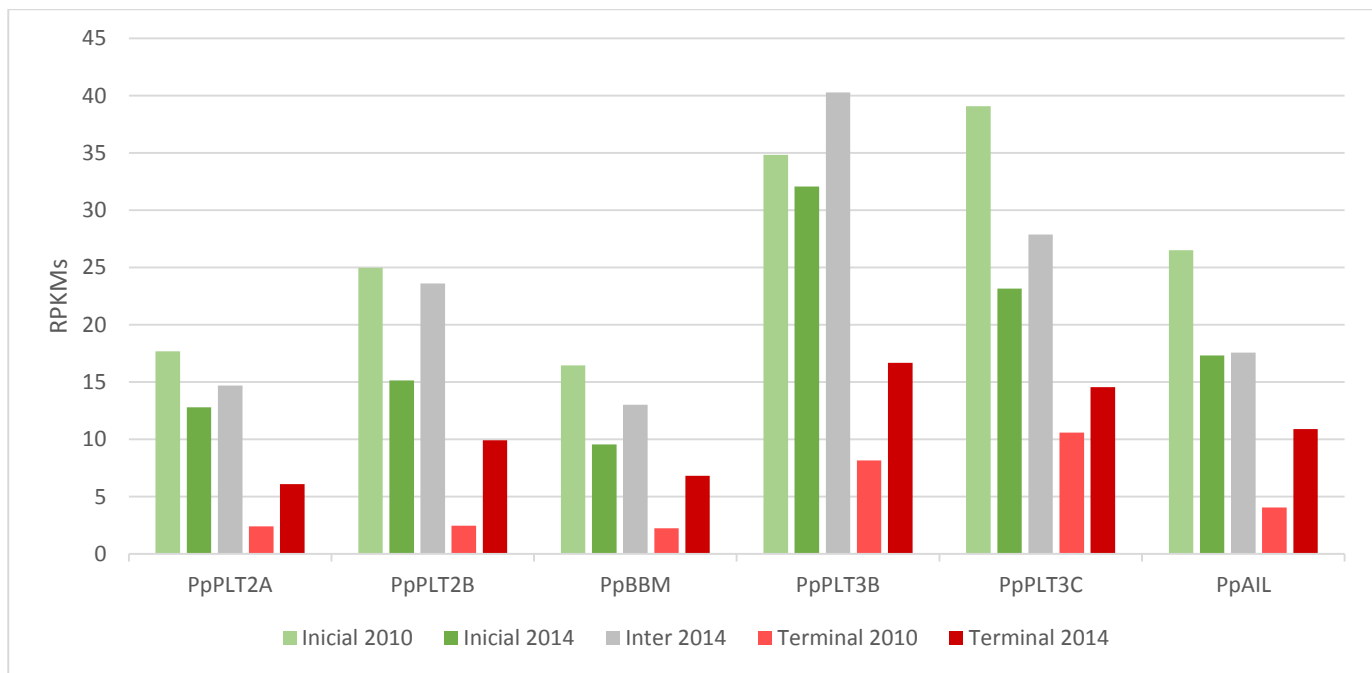


Fig. 46. Valores RPKM de los *contigs* con similitud a PLT obtenidos por mRNA-seq 2010 Vs. 2014. (Para mRNA-seq 2014 se muestran valores promedios de dos réplicas biológicas)

### 8.3 Genes de referencia para RT-qPCR

En la presente tesis se analizó la abundancia de los transcritos *PpPLT* mediante RT-qPCR. Para normalizar los datos del RT-qPCR fue necesario usar genes de referencia cuya expresión en la raíz es estable en las diferentes etapas del desarrollo de *P. pringlei*. Dichos genes se seleccionaron anteriormente con base en una lista de genes que se expresan de manera relativamente estable durante el desarrollo de *A. thaliana* publicada por Czechowski *et al.*, 2005. De esa lista se seleccionaron los genes *At4G24550* (subunidad del complejo adaptador de clatrina), *At5G65050* (proteína MADS-box AGL31) y *AT1G69960* (fosfatasa alcalina 2A). Las secuencias con similitud a estos genes de *A. thaliana* se buscaron entre los *contigs* resultantes de nuestro primer mRNA-seq.

Al analizar la expresión virtual de los *contigs* para los genes de referencia en los datos de la primera secuenciación, se encontró que los 3 presentan pequeñas variaciones. El valor de RPKM de *PpAGL31* aumenta 2.6% entre la etapa inicial y terminal, mientras que el de *PpCLA* disminuye 18% de la etapa inicial a la terminal y el de *PpPP2A* disminuye 22% de la etapa inicial a la terminal (Fig. 47). Utilizar más de un gen de referencia, de los cuales uno(s) se expresan un poco más en la etapa inicial, y otro(s) en la etapa terminal, permite una correcta normalización de la abundancia de transcritos en experimentos de RT-qPCR.

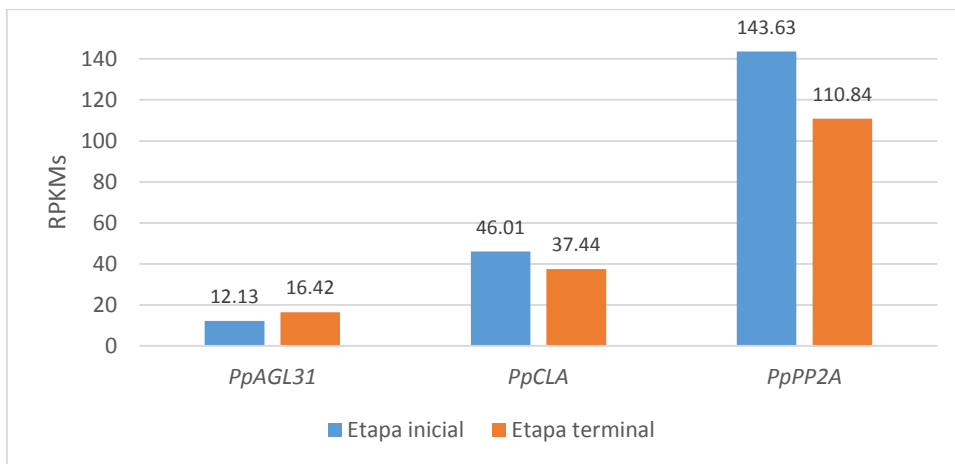


Fig. 47. Valores de RPKM para genes de referencia de RT-qPCR. Datos de la primera secuenciación.

Anteriormente estos tres genes se utilizaron exitosamente como genes de referencia en el análisis de abundancia de transcritos de otra especie, la cactácea *Stenocereus gummosus*, porque entre las pocas secuencias de esta especie que estaban disponibles en el laboratorio se encontraban partes de los transcritos de los genes mencionados. Usando las secuencias de los *contigs* de *PpAGL31*, *PpCLA* y *PpPP2A* con un alto nivel de similitud a los fragmentos de secuencias de *S. gummosus*, se diseñaron oligonucleótidos para amplificar en reacciones de RT-qPCR fragmentos de los tres genes de ambas especies. En el artículo (Shishkova *et al.*, 2013) finalmente se utilizaron *SgAGL31* y *SgCLA* como los genes de referencia suponiendo que en *S. gummosus*, presentarían el mismo comportamiento, uno disminuiría su expresión y el otro aumentaría en la etapa terminal permitiendo balancear estos cambios para la normalización.

En la fig. 48 se comparan las secuencias de los *contigs* de los genes de referencia obtenidas en el primer mRNA-seq contra el segundo.

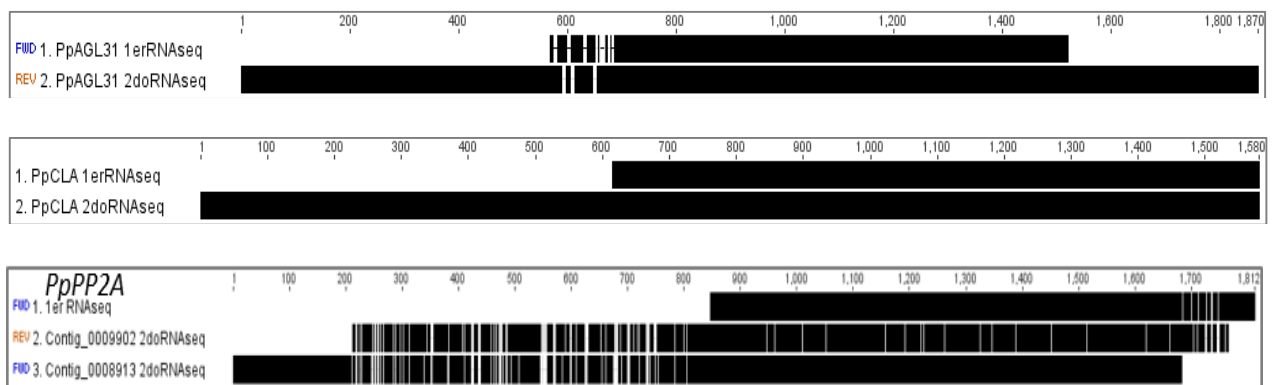


Fig. 48. Alineamiento de secuencias de los genes de referencia *PpAGL31*, *PpCLA* y *PpPP2A* del mRNA-seq anterior *vs.* secuencias del nuevo mRNA-seq. Para cada uno de los *contigs* las barras superiores del panel representa a los *contigs* de la primera secuenciación (utilizados en este trabajo de tesis) y las inferiores las de la nueva secuenciación. La imagen es una representación gráfica del alineamiento en donde las barras que se sobrelapan son secuencias que se alinean, los *gaps* se representan como espacios en blanco dentro de las barras negras.

Encontramos que para *PpAGL31* el *contig* de 926 nt de la primera secuenciación difiere en 91 nt del extremo 5' de la nueva secuencia, lo cual indica que casi todo este *contig*, estaba correctamente ensamblado excepto el extremo 5'. El nuevo *contig* para *PpAGL31* es del doble de longitud del anterior y contiene la secuencia codificante completa además de los UTRs 5' y 3'. En el caso de las secuencias para *PpCLA*, el *contig* de la secuenciación anterior se alinea perfectamente con el de la nueva secuenciación. El nuevo *contig* de *PpCLA* es al menos 600 nt más grande y contiene la secuencia codificante completa además de los UTRs 5' y 3'.

El *contig* *PpPPP2A* del primer mRNA-seq se alinea con dos diferentes *contigs* del nuevo mRNA-seq: el *contig* 0008913 y el *contig* 0009902, que guardan una muy alta similitud entre sí (Fig. 48, panel inferior). Entre las regiones altamente conservadas que comparten ambos *contigs* está precisamente la región en la que se diseñaron los oligonucleótidos para RT-qPCR. Por esta razón, aun cuando en las reacciones de PCR con estos oligonucleótidos pareciera haber un único amplicón, en realidad se generan dos amplicones de la misma longitud, con apenas unas cuantas bases diferentes. Por lo descrito, se recomienda no utilizar este gen en los experimentos posteriores.

Los datos obtenidos en el nuevo mRNA-seq de expresión virtual de estos genes de referencia se muestran en la figura 49.

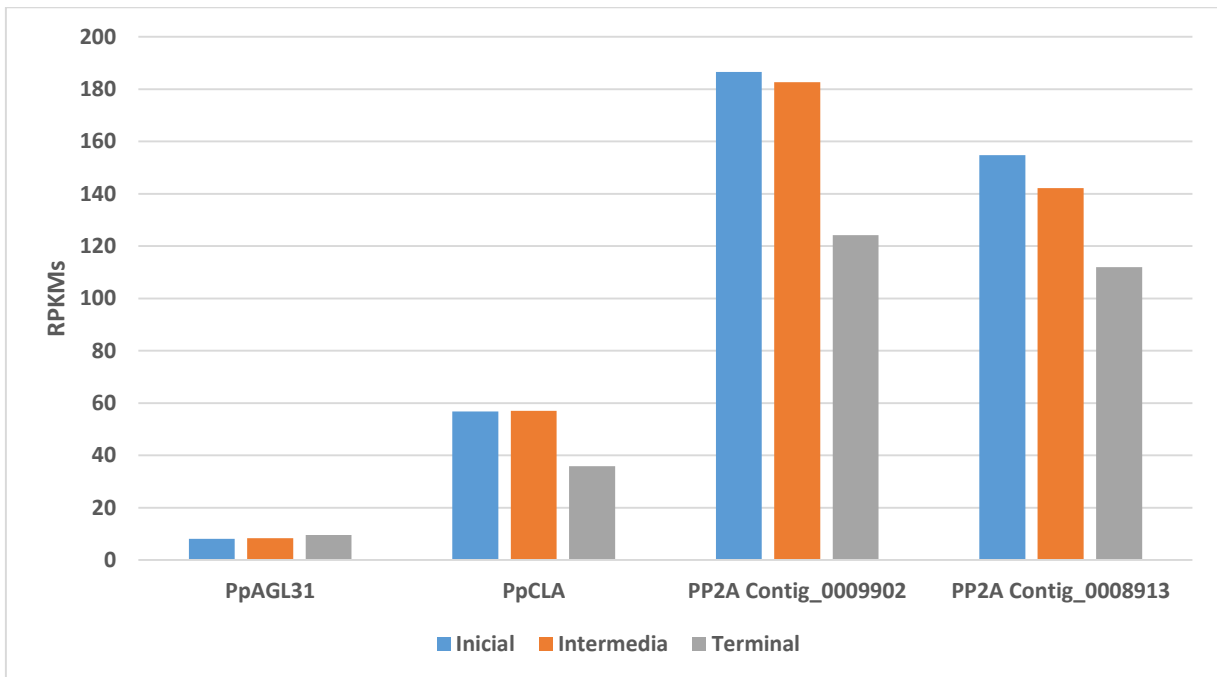


Fig. 49. Valores de RPKM para genes de referencia de RT-qPCR. Datos de la segunda ronda de secuenciación.

Los datos de expresión virtual indican que los genes de referencia elegidos previamente no son idóneos y se sugiere en este trabajo buscar otros genes normalizadores. Particularmente, los valores de RPKM de *PpAGL31* son de alrededor de 10 en todas las etapas, lo que indica que la abundancia del transcrito de este gen, aunque es estable a lo largo de las etapas de desarrollo, es muy baja. Este bajo nivel

de expresión queda confirmado en las reacciones de RT-qPCR (Fig. 36) en las que se reportan Cqs elevados, de más de 25 ciclos para todas las etapas del desarrollo de *P. pringlei*.

Los valores de RPKM de la subunidad de clatrina *PpCLA* indican un decremento de expresión de aproximadamente 35% en la etapa terminal del desarrollo. Por lo tanto, usar a *PpCLA* como gen de referencia podría afectar el resultado final si no se usa a la par de un gen de referencia que tenga un incremento parecido en la etapa terminal para contrarrestar este efecto.

Los dos *contigs* con similitud a *PP2A* muestran el mismo patrón de expresión. Los valores de RPKM indican un decremento de expresión de 27% para uno de los *contigs* y 33% para el otro en la etapa terminal del desarrollo. Este gen (o al menos este par de oligonucleótidos) no debe usarse como un gen de referencia porque podrían alterar el resultado final.

Todo lo anterior indica que los resultados de RT-qPCR obtenidos en este trabajo deben tomarse como resultados preliminares hasta contar con nuevos genes de referencia, más robustos y confiables.

Para normalizar la expresión de los genes *PpPLT* se usaron los tres genes de referencia antes descritos y se utilizó el método publicado por Pfaffl, 2001. Además, se probaron las diferentes combinaciones con dos de los tres genes de referencia o con un solo gen de referencia (datos no mostrados). El patrón de expresión, es decir, mayor abundancia de los transcritos en la etapa intermedia y menor abundancia en la etapa terminal, se mantuvo con las diferentes combinaciones de los genes de referencia. Aún cuando al cambiar los genes de referencia que se utilizaron para normalizar los niveles de expresión génica en este ensayo se modifica cuantitativamente la expresión medida, el patrón observado es el mismo.

#### **8.4 Comparación entre la cuantificación por RT-qPCR y el análisis de expresión virtual del nuevo mRNA-seq**

A continuación se presentan los resultados de las reacciones de RT-qPCR normalizados con los tres genes descritos anteriormente y comparados con los resultados de expresión virtual del último mRNA-seq (Fig. 50-52). Las barras anaranjadas corresponden a los datos normalizados de expresión virtual para las etapas inicial, intermedia y terminal. Como lo mencionamos antes, para este mRNA-seq se utilizaron muestras de puntas de la raíz en tres etapas de desarrollo: las etapas intermedia y terminal son las mismas, que las utilizadas en el análisis por RT-qPCR. Para el mRNA-seq se utilizó una única etapa inicial, que contenía en su mayoría puntas de la raíz de la etapa clasificada como inicial 2 para RT-qPCR. Las barras en azul corresponden a los datos de RT-qPCR normalizados con tres genes de referencia de las etapas inicial 1, inicial 2, intermedia y terminal. Dichas etapas se acomodaron en estas graficas de izquierda a derecha siguiendo el orden cronológico del desarrollo de la raíz de *P. pringlei*, con el fin de poder comparar ambos métodos de cuantificación.

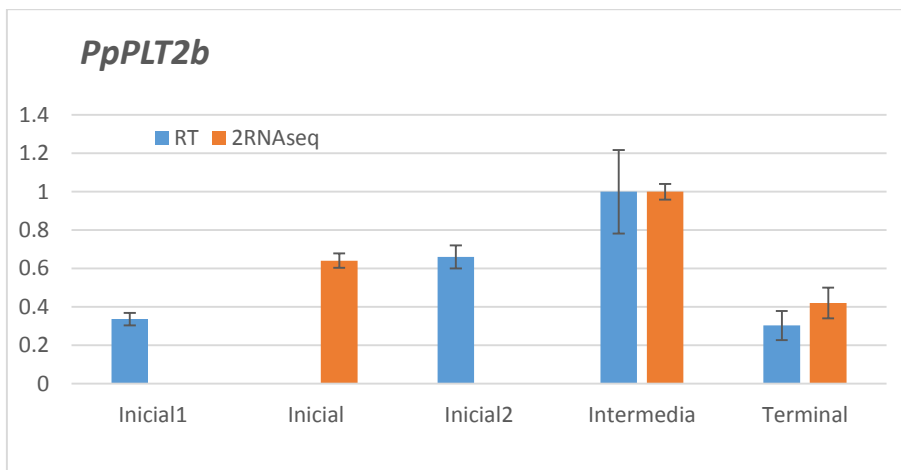


Fig. 50 Expresión relativa normalizada del transcrito *PpPLT2b* analizada por RT-qPCR y expresión virtual proveniente de los datos de mRNA-seq. Para los datos de RT-qPCR se muestran promedios de dos réplicas biológicas.

Para el gen *PpPLT2b* los datos de expresión obtenidos *in silico* y en RT-qPCR son muy similares. Después de las etapas iniciales del desarrollo hay un aumento en la expresión que llega a su máximo para la etapa intermedia, y finalmente un decaimiento en la expresión en la etapa terminal.

El gen *PpPLT2b* que tiene una alta similitud con *AtPLT2*, que tiene un importante papel en el mantenimiento del RAM, por lo que se esperaba justamente que su expresión bajara para la etapa terminal como ya se había discutido anteriormente.

En cambio, los datos obtenidos mediante RT-qPCR para los genes *PpAIL* y *PpPLT3b* difieren de los datos de expresión virtual (Fig. 51 y 52).

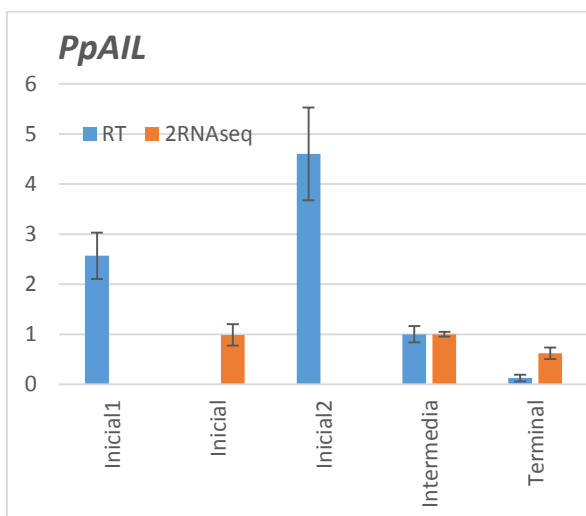


Fig. 51. Expresión relativa normalizada del transcrito *PpAIL* analizada por RT-qPCR y expresión virtual proveniente de los datos de mRNA-seq. Para los datos de RT-q PCR se muestran promedios de dos réplicas biológicas.

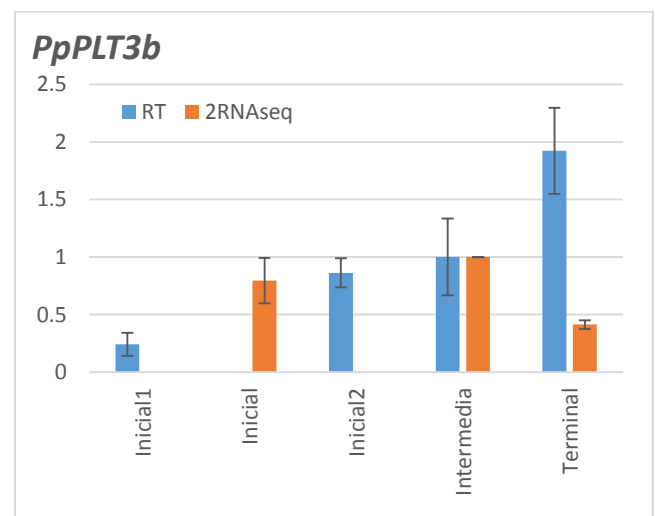


Fig. 52. Expresión relativa normalizada del transcrito *PpPLT3b* analizada por RT-qPCR y expresión virtual proveniente de los datos de mRNA-seq. Para los datos de RT-q PCR se muestran promedios de dos réplicas biológicas.

Cabe recordar que en las reacciones de RT-qPCR de estos dos genes (*PpAIL* y *PpPLT3b*) hubo diferencias considerables entre las réplicas biológicas de algunas de las etapas del desarrollo analizadas (Fig. 37).

Para *PpAIL* los datos provenientes del RT-qPCR indican una mayor expresión en las etapas iniciales y un decaimiento hacia las etapas intermedia y terminal, en esta última la expresión es muy baja y apenas cuantificable, por lo que su cuantificación está sujeta a errores probabilísticos. En general, si los datos de expresión virtual sugieren que la abundancia del transcrito *PpAIL* es similar entre las etapas inicial e intermedia y que existe un decaimiento en la etapa terminal, los datos del RT-qPCR apuntan a que existe un aumento en la abundancia de transcrito de la etapa inicial 1 a la inicial 2, seguido de un decremento que continua hasta la etapa terminal. La diferencia entre los datos provenientes de ambos métodos está en las etapas iniciales, mientras que en la etapa intermedia y terminal son más equiparables. Probablemente, utilizando nuevos genes de referencia, tal y como se propuso en párrafos previos a esta sección, e integrando más réplicas biológicas, será posible lograr resultados reproducibles.

La abundancia del transcrito *PpPLT3b*, de acuerdo a los datos de expresión virtual muestra un ligero incremento entre la etapa inicial y la intermedia; para estas etapas encontramos una tendencia similar en RT-qPCR (recordemos que la muestra colectada para la etapa inicial del mRNA-seq se parece más a la etapa inicial 2 del RT-qPCR). Sin embargo, los valores RPKM sugieren que la abundancia del transcrito decae en la etapa terminal. Este dato no concuerda con lo que se encontró por RT-qPCR, es decir, un incremento desde las etapas iniciales e intermedia hacia la etapa terminal. Sin embargo, si observamos los datos crudos sin normalizar (Fig. 36) este incremento en la etapa final podría ser explicado como un artefacto de la normalización.

En general, los tres genes analizados en este estudio presentan un patrón similar de expresión virtual: los valores RPKM aumentan ligeramente o no cambian de la etapa inicial a la intermedia y decaen en la etapa terminal. Este es el comportamiento que esperábamos encontrar para estos factores de transcripción; sin embargo, en los experimentos de RT-qPCR este patrón se confirmó únicamente para *PpPLT2b* por RT-qPCR. Aunque, como ya mencionamos antes, esto podría deberse a artefactos generados en la normalización debido a que nuestros genes de referencia no son los ideales.

## 9. Conclusiones

El análisis filogenético sugiere la existencia de al menos un posible ortólogo en *P. pringlei* para cada uno de los genes *PLT* de *A. thaliana*. Por lo tanto, estos genes podrían estar cumpliendo los mismos roles que tienen sus posibles ortólogos en *A. thaliana*.

El grado de conservación de los dominios AP2 de unión a DNA en las secuencias de proteínas PLT en las especies analizadas sugiere que las funciones de estas proteínas pueden estar conservados.

La mayoría de las especies analizadas cuentan con genes pertenecientes al Clado A y al Clado B de *PLT*, esta posible duplicación podría estar reflejando la neofuncionalización de dichos genes.

El análisis filogenético sugiere que en el Clado A una duplicación dio origen a la rama PLT1 y PLT2, mientras que en el Clado B otra duplicación originó la rama de PLT5. Ambos eventos de duplicación fueron posteriores a la separación que dio origen a las plantas monocotiledóneas y dicotiledóneas.

La abundancia del transcrito *PpPLT2a* medida por RT-qPCR durante el desarrollo de la raíz primaria de *P. pringlei* muestra un incremento en la etapa intermedia respecto a la inicial, y un decremento en la etapa terminal. Estos resultados son consistentes con los del mRNA-seq.

## 10. Perspectivas

### a) Selección y uso de nuevos genes de referencia para RT-qPCR en *P. pringlei*

En este trabajo se realizó el análisis de la expresión de los genes *PpPLT* mediante RT-qPCR. Dichas mediciones pueden contener un sesgo debido a cambios en la abundancia de los transcritos de los genes normalizadores entre las etapas del desarrollo de la raíz primaria de *P. pringlei* (favor de referirse a la pág. 80). Con el fin de aumentar la precisión de las mediciones se requiere del uso de nuevos genes de referencia, que se elegirán usando los datos de expresión virtual del segundo mRNA-seq, para verificar que sus niveles de expresión virtual sean constantes entre las diferentes etapas del desarrollo analizadas.

Una vez seleccionados, deberán diseñarse oligonucleótidos apropiados para RT-qPCR que serán usados para establecer los valores de referencia en cada una de las muestras a analizar.

### b) Analizar los dominios y cajas conservadas de las proteínas PLT

Con las secuencias completas de los transcritos de *PpPLT* es posible realizar un análisis comparativo entre las secuencias de posibles genes ortólogos de *P. pringlei*, *A. thaliana* y las especies vegetales que se incluyeron en los análisis filogenéticos, enfocándonos en los dominios AP2 de unión a DNA y cajas conservadas entre dichas secuencias.

### c) Analizar los patrones de expresión de los genes *PpPLT* mediante hibridación de RNA *in situ*

Otra de las perspectivas de este trabajo consiste en identificar los tejidos y tipos celulares en los que se expresan los transcritos de los genes *PpPLT* usando hibridación *in situ*. Dichos genes pertenecen a una gran familia génica, dentro de la cual se observa un alto grado de similitud. Actualmente contamos con las secuencias completas o casi completas de los transcritos *PpPLT*, por lo que será posible diseñar las sondas de hibridación utilizando las regiones 5' y 3' UTR que no están conservadas y que nos permitirían realizar hibridación *in situ* para cada uno de los genes de interés. Una vez diseñadas las sondas de hibridación será necesario montar y estandarizar en el laboratorio el método de hibridación *in situ*, amplificar y marcar las sondas, preparar el tejido usando raíces completas o realizando cortes longitudinales de las mismas para fijarlas y realizar la detección de los transcritos.



## 11. Bibliografía

- Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., et al. (2004). The *PLETHORA* Genes Mediate Patterning of the *Arabidopsis* Root Stem Cell Niche. *Cell*, 119, 109-120.
- Bertrand, C. et al. (2003) Arabidopsis histone acetyltransferase AtGCN5 regulates the floral meristem activity through the WUSCHEL/AGAMOUS pathway. *J. Biol. Chem.* 278, 28246–28251
- Boutillier, K. et al. (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14, 1737–1749.
- Carrier, M.F. et al. (1997) Actin depolymerizing factor (ADF/cofilin) enhances the rate of filament turnover: implication in actin-based motility. *J. Cell Biol.* 136, 1307–1322.
- Czechowski, T., Stitt M., Altmann T., Udvardi, M. K. & Wolf-Ru (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. *Plant Physiology*, 139, 5–17.
- Clontech. (2009). SMARTer™ PCR cDNA synthesis kit User Manual. Clontech Laboratories Inc.
- Dolan, L., Janmaat, K., Willemsen, V., Linstead, P., Poethig, S., Roberts, K., et al. (1993). Cellular organisation of the *Arabidopsis thaliana* root. *Development*, 119, 71-84.
- 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.
- Dubrovsky, J. G. & Shishkova, S. (2012). Plant roots: the hidden half, 4<sup>a</sup> ed., Capítulo 28, Developmental adaptations in roots of desert plants with special emphasis on Cacti. CRC Press.
- 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, 33-42.
- Esau, K. (1997). Anatomy of Seed Plants, 2<sup>a</sup> edición. *John Wiley and Sons, New York*.
- Fischer, U., Ikeda Y., Ljung, K., Serralbo O., Singh, M., Heidstra, R., Palme, K., Scheres, B. & Grebe, M. (2006). Vectorial information for Arabidopsis planar polarity is mediated by combined AUX1, EIN2, and GNOM activity. *Curr. Biol.* 16, 2143–2149.
- Floyd, S.K., and Bowman, J.L. (2007). The ancestral developmental tool kit of land plants. *Int. J. Plant Sci.* 168, 1–35.

Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., et al. (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature*, 449, 1053-1057.

Gilbert, Scott F. (2005). *Biología del Desarrollo 7ª Edición*. Ed. Médica Panamericana.

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. *Curr. Biol.* 23, 956–962

Horstman, A., Willemsen, V., Boutilier, K. & Heidstra, R. (2013). AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends in Plant Science.* 19(3):146-57.

Hunt, D., Taylor, N. & Charles, G. (2006). *The new cactus lexicon*. Milborne Port: dh books.

Kim, S., Soltis, P.S., Wall, K., & Soltis, D.E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Mol. Biol. Evol.* 23, 107–120.

Kornet, N. & Scheres, B. (2009) Members of the GCN5 histone acetyltransferase complex regulate PLETHORA-mediated root stem cell niche maintenance and transit amplifying cell proliferation in Arabidopsis. *Plant Cell* 21, 1070–1079

Krizek, B. (2009). AINTEGUMENTA and AINTEGUMENTA-LIKE6 act redundantly to regulate Arabidopsis floral growth and patterning. *Plant Physiol.* 150, 1916–1929.

López-Valle, M. L. (2011). Tesis de Licenciatura: Identificación de genes involucrados en el crecimiento determinado de la raíz de las cactáceas desérticas. Directora de Tesis: Shishkova, S. UAEM (Universidad Autónoma del Estado de Morelos. 25-27.

Mähönen, A.P., Ten Tusscher K., 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.

Malamy, J. E. & Benfey, P. N. (1997). Down and out in Arabidopsis: the formation of lateral roots. *Trends in Plant Science*, 2, 390-396.

Matsuzaki, Y., Ogawa-Ohnishi M., Mori A., Matsubayashi Y. (2010). Secreted peptide signals required for maintenance of root stem cell niche in Arabidopsis. *Science*, 1065, vol. 329.

Mauseth J.D. (2009). *Botany: An Introduction to Plant Biology*, 4ª Edición. Sudbury, MA: Jones and Bartlett Publishers.

Miwa, H., Kinoshita, A., Fukuda, H., & Sawa, S. (2009). Plant meristems: CLAVATA3/ESR-related signaling in the shoot apical meristem and the root apical meristem. *Journal of Plant Research*, 122, 31-39.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 5(7), 621-628.

Mudunkothge, J.S. & Krizek, B.A. (2012). Three *Arabidopsis* AIL/PLT genes act in combination to regulate shoot apical meristem function. *Plant J*. 71, 108–121

Niklas, K.J., Molina-Freaner, F., Tinoco-Ojanguren, C., Paolillo, Jr. D.J. (2002). The biomechanics of *Pachycereus pringlei* root systems. *Am J Bot* 89:12–21.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*. Vol 29, No. 9.

Pfaffl, M.W. (2004). Quantification Strategies in Real-Time PCR. In: A-Z of Quantitative PCR S.A.Bustin, editor. International University Line. La Jolla, p. 87-120.

Passarinho, P. et al. (2008) BABY BOOM target genes provide diverse entry points into cell proliferation and cell growth pathways. *Plant Mol. Biol*. 68, 225–237.

Pinon V., Prasad K., Grigga S. P., Sanchez-Perez G. F. & Scheres B. (2012). Local auxin biosynthesis regulation by PLETHORA transcription factors controls phyllotaxis in *Arabidopsis*. *PNAS*, vol. 110, 1107–1112.

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. *Arabidopsis* PLETHORA transcription factors control phyllotaxis. *Current Biology* 21, 1123–1128.

Rodríguez-Rodríguez, J. F., Shishkova, S., Napsucialy-Mendivil, S., & Dubrovsky, J. G. (2003). Apical meristem organization and lack of establishment of the quiescent center in Cactaceae roots with determinate growth. *Planta*, 217, 849-857.

Sabatini S., Beis D., Wolkenfelt H., Murfett J., Guilfoyle T., Malamy J., Benfey P., Leyser O., Bechtold N., Weisbeek P., Scheres B. (1999). An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell*, 99: 463-72.

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, 354-358.

Sablowski, R. (2007). Flowering and determinacy in *Arabidopsis*. *Journal of Experimental Botany*, 58, 899-907.

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*, 290:998–1009.

- Scheres, B., Benfey, P., & Dolan, L. (2002). Root Development, The *Arabidopsis* Book American Society of Plant Biologists (1-18).
- Scheres, B., Laurenzio, L. D., Willemsen, V., Janmaat, K., Weisbeek, P., Benfey, P. N., et al. (1995). Mutations affecting the radial organisation of the *Arabidopsis* root display specific defects throughout the embryonic axis. *Development*, 121, 53-62.
- Shane, M. W., & Lambers, H. (2005). Cluster roots: A curiosity in context. *Plant and Soil*, 274, 101-125.
- Shane, M. W., Dixon, K. W., & Lambers, H. (2005). The occurrence of dauciform roots amongst Western Australian reeds, rushes and sedges, and the impact of phosphorus supply on dauciform-root development in *Schoenus unispiculatus* (Cyperaceae). *New Phytologist*, 165, 887-898.
- Shishkova, S. & Dubrovsky, J. G. (2005). Developmental programmed cell death in primary roots of Sonoran Desert Cactaceae. *American Journal of Botany* 92, 1590 -1594.
- Shishkova, S., Rost, L. T., & 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., 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*, 111 (6).
- Sievers, A. & Braun, M. *The Root Cap: Structure and Function*. (1996). *Plant roots: the hidden half*, 2a ed., Marcel Dekker Inc.
- Skene, K. R., Raven, J. A., & Sprent, J. I. (1998). Cluster root development in *Grevillea robusta* (Proteaceae). *New Phytologist*, 138, 725-732.
- Sánchez-Calderón, L., López-Bucio, J., Chacón-López, A., Cruz-Ramírez, A., Nieto-Jacobo, F., Dubrovsky, J. G., et al. (2005). Phosphate Starvation Induces a Determinate Developmental Program in the Roots of *Arabidopsis thaliana*. *Plant & Cell Physiology*, 46, 174-184.
- Svistoonoff, S., Creff A., Reymond, M., Sigoillot-Claude C., Ricaud L., Blanchet A., Nussaume, L. & Desnos, T. (2007). Root tip contact with low-phosphate media reprograms plant root architecture. *Nature Genetics*, 39, 792-796.
- Stahl, Y. & Simon, R. (2009). Is the *Arabidopsis* root niche protected by sequestration of the CLE40 signal by its putative receptor ACR4? *Plant Signaling & Behavior*, 4, 634-635.
- Steeves, Taylor A. & Ian M. Sussex. (1989). "Patterns in Plant Development" 2<sup>nd</sup> Edition. Cambridge University Press.
- Taiz, Lincoln & Eduardo Zeiger. (2006). "Plant Physiology", 4<sup>a</sup> Edición, Sinauer Associates.

Watt, M., Silk, W. K., & Passioura, J. B. (2006). Rates of Root and Organism Growth, Soil Conditions, and Temporal and Spatial Development of the Rhizosphere. *Annals of Botany*, 97, 839-855.

Weigel, D., & Jürgens, G. (2002). Stem cells that make stems. *Nature*, 415, 751-754.

Webster, P. L., and R. D. MacLeod. (1996). *Plant roots: the hidden half*, 2<sup>a</sup> ed., Marcel Dekker Inc. 51–109

Wilcox, H. (1962). Growth Studies of the Root of Incense Cedar, *Libocedrus decurrens*. II. Morphological Features of the Root System and Growth Behavior. *American Journal of Botany*, 49, 237-245.

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

**Secuencias de nucleótidos de contigs de *P. pringlei* con similitud a *PLT***

>ContigA|PpPLT2a|Cactassy\_W22B50\_14680\_Cov\_15\_selection|PREDICTED:AP2-like ethylene-responsive transcription factor PLT2-like [*Vitisvinifera*]~2.162E-105~XP\_002285539

CACCACCACCGCCACCATAGCTCCCACTAGGGCTCCCATCATTAGGGTTAGCATCAAGGGCACCGGAGGTA  
GTAGATCCCCAATGCTTATAAGCCCATGAAGCAAATTGGGGTTAGAATGGAGATTACTACAAGTGTGTA  
GATGCTAGGATTTGATGTGTTTCTTGATTGTTGGTATGGCTGAAGGAGTTGAGCCTTTGGAAGTAATTGTG  
ATGGGAGAGAGAGAGAGGTTGCTTCTGGCTTTGTAGGGTTAGCAAGGCTTGTGGTTGGTGATCTTGGTATT  
GTGGTTGACTATATGTCAGAAGAGGGTAAGCAATCGGAAGGTAGTGGGCCCATATCTACCACTCGAGTTT  
GAGCCAGACCATACTGAAAGCTGTTATTTCTAGAGGCTTCTAGCATTGAGCTTCTTTAGGCGCTTAGCT  
GCCCCTGACCCTATGGGGAGCGTACTACTTTCCAAGATGGCCTTGACATCATAACGGCTCATGTCAAAGTTG  
GTGACCGCATTGAGACCTCTGAACCTGATGGCTGCTATGTCATAAGCTTCTGCTGCCTCCTCCTCAGTGCTGA  
ATGTTCCCAAGTAGAGATCTTTGTTTCTGCCACTCTTCCAATTCTGCCTGCCAACGCCATGTTGGTGATG  
CCTTGTTACCCACGATAAATGGAAGCACCTCTTAAAATCCGCTACTCTTCTTATAGCAGCAATGAAC  
TCTTGTCTAGTCATGTGTTTCATCTCTTCTATTTCTTTTCGTAATTGCTTATTGGAAAATTAGTGGTAGTAGA  
AGTTCCCAGTACTTGATAGCAGCTAAATCATAAGCCCTGCAGCCTTCTTCTTGTACATACCCTCCAGA  
TAGACTTGGCGACCCTCCTGGACTGACCTCCCTTCTACAGCTATTATCCCAAAGATGAGCTTCATATCTTC  
CCGTC

>ContigB|PLT2b|Cactassy\_W22B50\_4089\_Cov\_19|PREDICTED: AP2-like ethylene-responsive transcription factor PLT2-like [*Glycine max*]~1.046E-105~XP\_003539421

TAATAGCTGTAGGAGGGAAGGACAGTCAAGGAAGGGTCGACAAGTTTATCTTGGAGGGTATGATAAGGAA  
GAGAAAGCTGCCAGGGCTTATGATCTTGCTGCCATCAAATACTGGGGAAGTCTACTACTACTAATTTTCCA  
ATAAGCAATTACGAGAAAGAAATAGAAGAGATGAAGCACATGACTAGACAAGAGTTTGTGCGGGCTATTA  
GGAGGAAGAGTAGTGGATTCTCAAGAGGTGCATCAATTTATCGTGGTGTAAACAAGGCACCACCAACATGG  
TAGATGGCAAGCAAGGATTGGAAGGGTGGCGGAAACAAAGATCTGTACTTGGGAACATTTAGCACCGAG  
GAAGAAGCAGCAGAAGCCTATGATATAGCAGCCATCAAGTTCAGAGGCATGAATGCCGTGACCAACTTCG  
ACATGAGTCGTTATGACGTCAAGGCCATCCTAGAGAGCAGCACACTCCCTATAGGAGGTGGTGGCGCCAA  
GCGGCTAAAGGAAGCCCAAGCCTTAGAATCTTCTAGAAGAAGGGAGGAAATGTTTGCCTAGCCACGGCTC  
CCACCACCTTCCAGTATGGCAGTGCCACCTCCTCAAGCTCGAGTAGGTATGGGTGCATAACCAGTATCCTC  
TTCTTTCTACCACCAAGATCACCCCAACCCTTGCTAACCTGCAAAATGGAGGAGACAAAACATCAGTCT  
CTCTCATTACTCCAAGAGGCTCAATTCCTTCAACTATATCAACAACAATCATCAAGCTACTCGTCAAACCT  
AGTATTTACAACGCTGCTAGTTATAACGGCCTTCACTCAAACCTAACTTGTTCAGGGCTCATGAGCATG  
CATGGAGTGTCTAGCTCTTCTGGGGTCTTGATGCTGGTACCAACAACAATGGGAGCTCAAGTGGGAGTTA  
TAGTGGTGGGTACATAGGCGGTGGGATGGGGTCCACATCGAGCAATGCAGTGGGGCCATGGACCCTGA  
GCCCCTTGCAATGGTGAAGGTAGATTATGACATGCCCTCGACTGGCGGTGGTGTGCTGGTACGGCTATGGCG  
GGTGGCCGACTGAGTCAGTTCAGGGTCCCATAATGGTGGCGTTTTTGCATGTGGAATGACTGATTCGTG  
ATTGGCGATGGTCCACGTTGAGTATTTAATTAATTTGTTTTTTTTTTTTGGGGGGGGGG

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

>ContigC|BBM|Cactassy\_W22B50\_11798\_Cov\_13|AP2 domain-containing transcription factor [Populus trichocarpa]~1.888E-119~XP\_002311259

ATGGACGGGAAGATATGAAGCGCATTGTGGGATAATAGTTGCAGGAGAGAAGGCCAGACTCGCAAAGG  
AAGACAAGTGTATTTGGGGGGTTATGATAAAGAAGAAAAAGCAGCTCGAGCTTATGATTTGGCTGCTATTA  
AATACTGGGGCACTACTACTACTAATTTTCCATTAGTGACTACGAGAAGGAGATTGAGGAAATGAAG  
CATATGACTCGTCAAGAATATGTAGCATCTCTAGAAGGAAAAGCAGTGGATTTTCTCGTGGTGCATCGATA  
TATCGAGGAGTAACAAGGCATCATCAACATGGGCGTTGGCAAGCAAGAATTGGCAGAGTTGCTGGCAACA  
AAGACCTCTACCTTGGCACTTTAGTACACAAGAGGAAGCAGCAGAGGCCTATGACATAGCAGCAATCAAG  
TTCAGAGGCCATAATGCAGTGACAACTTCGAGCTCAGCAGATATGATGTAAGCCATCCTTGAGAGCAC  
GACTCTCCCATCGGAGGAGCTGCGAAGAGGCTCAAGGATGTCGATGACTTGACCTCTGCAGCGGATACCA  
GGCATCAGATTCTGAGGGTGGCTGCTGTTACTGATCACGACAACCAACTCCCAACTTAGCGGTGATCAC  
GCAGGAGTCATTCCCAACTCCATTCTGTGGCCTGGCATTGCGTACCCCAACAAGCTCACCCGCTTTCCATGC  
ATTACCCCTATGCCAGAGGCTCTGGTGCAAGCAGGAGAATGAACCTGATGTGGGTTTCAACAATGGTAAC  
TACCAGGATCATCTTACCACCAGCTTCAAGTTGAACAGTGGCGGTACCCACAACCTTCAATCAATGCATAATT  
TGATGGGGTTGGAGAGTACCTCTGCTTCTTGAACATAGCTCGGGGTCCAATCCATTGTGTATGGGAAT  
GGGAACGGCAATGGCAATGATGCAGGTGCGTACAATGGGATCCACCCATTTGGTGTGATCGCCCCAGAGG  
GGAATGGGAATGGGTATATGGAGAGCAATGATAATCAGGGGAATTACTACTATTTCTCTCAACAAGGAAAT  
GGAGGGGGTGGTGGGTTAAGGGCGCCGTGGGAACCTTATGATCAAGGCTCGGCCTGCAACAACCTGGGTT  
CCAACCGCTATCCCAGCGCTCACCCCAAGGACAACCACCATGGCTGCCGGAGCTCCTCTTTTCACTGTGTGG  
AATGATACATAATCTCTGATTAATGGAGGAATTAAGGAAAGGGAAAAGATCGATGCATGGTGGGAAGA  
TCATCCGGTTGTACATGTGCGACTAGCCTTTTGCCTT

>ContigD|PLT3a|Cactassy\_W22B50\_2402\_Cov\_28|DNA binding protein, putative [Ricinus communis]~1.245E-85~XP\_002525305

CATTGTCATCGTCATCATCATCATGCGTTGTTGATTATCATGATCATGAAGAGGATAAGCCCTCCAAGCCCA  
AAAGGGCTAGAAGAACCAGAGCAATTAACCCTAAGTCTGATGAGAAATGCCCCATTTCCAATAACACCCCC  
ACCACCAATCGGAGAAGCTCCATTTACAGAGGCGTCACCAGGCATAGATGGACAGGAAGATTTGAAGCGC  
ATTTGTGGGATAAAAGTTGTTGGAATAGCATTCAAACAACAAAAAGGGCGACAAGTTTATTTGGGAGCATAT  
GACAGTGAAGTTGATGCAGCACGGACTTATGATCTGGCTGCACTTAAGTATTGGGGACCAGAGACCACATT  
AAATTTCCCTATAGAAAATTATACGAAAGAGCTCGAGGAAATGCAAAATGCGTCCAAGGAGGAGTACTTAG  
CATCATTGAGACGGAGAAGCAGTGGATTTTCAAGGGGTGTCTCTAAATATCGTGGCGTTGCTAGGCATCAC  
CATAATGGGCGATGGGAGGCACGAATTGGGAGGGTCTTCGGGAACAATACTCTACTTGGGCACTTTCA  
GTTACAGGAAGAGGCGGCAGCAGCTTATGATCTGGCGCAATTGAGTACAGGGGTCCGAATGCTGTTAC  
CAATTTGACATTAGCAACTACATTGGTCTGTGAAGAATCTAATGCCAACGGGTGCCAGCCGCAACAACC  
AAATACCAACAAAACAAACCTGTGCAACGACGACAGCAGCAGCAGCAACAACAATAAAACAGCAG  
GAAGAGGAAGATGAGCAACAAGATCAACCACGAGTACAAGAGCAGCAAGATTTGCAAGAAGTTCGAAAAC  
AGGAGGAAATGGCGACGCCATCTCAGTTCGAGGAGTCTAAGCTTGATATTAGCTCAGATGACCACTCTCT  
ATGCTCGTGATGGGCACACATGTTGATGACGAGGACAGAGAAAACAACCTTGGAGCTTTTGGCTGGACAC  
CGGGGCTTATGACCCCTCCTTGTCCCTGATATCCCCCTCGAAGAAGATGGGAGGGTGTCTGATTTATTTGA  
TCAAACAGGGTTCGAAGACAACATCGATTTGATCTTCAGGGGGAGGGGAATTCAATAACTTTGGATGAGA  
CCTACTTTGGTGGCACCAAGGATGTAGGTTGTCAAACAGGGATGAAACCTCAGAGGACTCAAACAAAAAA  
ATGCAGATCTCGTTGAAGATAGAGAACTCATTTTCTTATCCTCCCCTTCTCCGTCTTCAACTACTAGCTT

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

AGTCACTTGTCAACAGTGA TGCTCAAATGGCGCCTTTCCTTTTTCTTTTTTAAAAAAAACCTCCCCATGTGT  
TCTCTTCTCTTTTTCTTAGAACTTACCCATTGACTTGTTGTTGTTCTTGAGCTTTGGTTGGGATTTTTAG  
TTAGGAGGAACTTAAGGATGTGTGTGTTCTCTCTGGGAAAAAACAAAAATTTTGTGCTTGAATCT  
CAAAGT

>ContigE|PLT3b|Cactassy\_W22B50\_750\_Cov\_31|PREDICTED: AP2-like ethylene-responsive  
transcription factor AIL6-like [Vitis vinifera]~3.411E-136~XP\_003635541

CAGGATAACAATCGCTATATTATTGACATATCGTTGCACAACAGCAACGACCAAACGAGATACAATCCCTGG  
TACATAGAGCACAAATATGGGTTATCTCCATATGCTCTGAGAATCTGAGATCCAATTCAGAATCTCAAACAG  
TTTGAAAACCAATAAAGGCTGCATCAATTTCAACGAGACAAAACCTCAGGCTATCCAGCCCTTCCCTGCTAAT  
CAGGCTAGAGAAAGATACAGTCATTCGATTCTTTACAATCAGCTACAAAGAACAACTAACTAACTAATC  
ACTCGACAGAAGCCCTGGGCTAAACCTAACTCCACATGGAACCTTGCAATGGAGGGAGCAAAGAGGAT  
GAACGGCACATAGAATTTGTTAATTTCTCAAATGAACTATGTGAGTTGATCATCTAAACGGCATCCATCAC  
CGTGGACCTGTTCTATTTGTGCTGCTGTCTGCATTTGATGCTGATACAGGTATGATGAGCTGTCAGCCACA  
AACTTCAAAAACCTTGTCTTCTCAGTGAAAAGTTGATTTTTCTTGATCTAATAAAGCTTGTCTCTTCGTAT  
GTCATCTAGAATCCTCTCCCTTCCCATTTCTTCTCCTTCAACCGTTTCTCCAGCTCATAACAGTTCTCCGCAAG  
AACTGATCTTTTGGAGCATTTAGGGAACCGCACGTCCACATCTTCCCTGCTACCCAAGGAGAACACACCAGG  
AGCCCTAATCATTCCACCATCTAGGGAGACAGCTCCATCACTGATGCATGGAAGAGCAAGAAGCATCTCAG  
CTCTAGTTCTGTATACTTGAAGACGCGAAAAGAGATTATAGAATAACGTCTCCCTGAGGCCACACCCACTTG  
ATGTCAAACAAAATAAATGAGGCCTTCTACATGGATCATATTCAGTCAAAAACCTAGAAATCCTGCCGGAA  
CGTACCACCTTGGCAATCTTGGCTTGGGAATATCAAGTCTCCTCTGAGGATCATCACTATAACATCACCAG  
CATATGGTCTTAAAGTATCAAGACATATCACAAGAAAGCGACCCTGTAAAACCCGTCCAATAGAAGCTCCA  
GCCCTTGAATGCCAGAGCCTTTATTGACAAGTCTTACCATCATGATTTCCAGAGCTTTAATGCCTTCATA  
AGTCTTACAGACAATAGTAAGCATAGTCCCAGGCCAGGTATTGGCAAAAAGCCTGCTAAGGTCATCAT  
CATCTACTTTTCCGAGGGTAGCAACCACACCAAGAACATCTGGGACAAAAGTAAGGTTTGAAGCTGTGAA  
CCGTGATCCCTTTGAAGCTTAAATATATACCAGCAGCAGAATTTTCATGCTTCAAGATCTGCTCAACAGTT  
CCTCTTACCTCGACTATGAGAGGTATCATCATCTTCACTTGGGTACACTTGAGGAAATGTGCTTACCAA  
GGCTAACTTGCAAATCCAGAATGGCGTCATCCAACCTGGTTTTTTGAGCCTCAGAAATTAAGATTATCCTC  
ATGCCATTTTATTTACTCCCACTTTTCTCCACTTCATCTTGTAAATTTCTGGGAACTAGACATTAGAGAATTTG  
CATAGGCAGGTGCACCATTTGGTACTTCGTTTGCAGTGTTACAAAAGTGGTATTCGGATCCACATGATGTA  
TAGCTGTAATATCTTGAACAATAATGCCCTGCTATTAATTGGAACCGGAGAATTCGTCGGGTACATGGTTG  
ATGAGGTTGATCGATGACGCCCCCTCCCCCTTTGTGTGGCTTTGTGATGATGTTGGTTTCTCTCTCTCT  
CCTCTCTCTCCATTCATCAATGGCGTTCCTTGTGTCATGCAGTCATTAATGCGCGCTCTTCTTCTCCATTT  
ATTTTATCCCATCTTCGTTTCGTATCGGAACTTTATAACTCCATATATATTCATCTTTCTGTTAAACAAGAGA  
ATGTATGAATAGTAGTAGCAGTAGCAGCAGTTGATGAGTGTAAACGGCAAGCACTTAATTTTCTGCTTCTG  
GGTTGGTTTCTTCTCATCAAACCAATGAGACTTTCTCTCTAGCTTCTACTTGCTTACCCTTAGACTAGACTAG  
AGAGAACCATCATTAAGTCTGGGAAGAAATTAATTAAGCTTTTGATTTGAGTTGAAAAAATGGCACC  
AGCAGCAGCGAGCAACTGGCTCTCATTCTCTCTCTCCAATGGAAATGTTGGCTACTTCTCATCATCATCAC  
CATCATCATCAAAACCCCCAGATTAATCTTCCACTGCTCTTCTCCATCTCACCACCATTATTCTGCCTTC  
TTCGACACCTTCTATTCTAATGATTGGGCGAATCCAAAACCAAGTATGTACGCGGACTACGCGGAGGA  
GCACACGACGGCTGTGCCGAAGTTGGAAGACTTCTGGGCGGCGACAACACGACGTCGTATAACAGCCAG  
ACTGAGGCCACAACCTCAGGACGAGTCGTCCAGCTGACTCACCTCTACGACCCACATCACGCCGCCGCAAT



Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

TACTTCGCCGACCACCACCATCACCACCAAGATCTGAATTCGTTGGCGGAAAATGGGTTTCAGTCGGCGGC  
GGCGAACTCGGGCTCGGAGGTCGATGACTCGTGCTCGGTGGGTAGAGTGACTCAGCTGAGTTGCGCCGGT  
TCTGAGTTCGTGGGTGAGTCCATCGAGACGGAGCTGGTGGCGGGTTTCGCCGTCGCCGCCAGTGCTTCCGC  
TTCTAAGAGCGCCGAGAAGGCGTTGGTTTCGGTGGCGGTGGATTGAGATCGTGCCAAGAAGATTTCTGACA  
CTTTTGGCCAGCGAACCTCCATTTACAGAGGAGTTACCAGGCATAGATGGACAGGCAGATATGAAGCACAT  
CTATGGGATAACAGCTGCAGGAGGGAAGGGCAAGCTAGAAAAGGCCGTCAAGTTTACTTGGGTGGTTATG  
ATAAAGAAGAGAGGGCAGCTAGAGCTTATGATTTGGCCGCTCTGAAGTACTGGGGTCCAAGTCCACCACA  
AATTTCCCTATTTCCGATTACGCCAAGAGTTGGAGGAGATGAAAAATGTGACCAAGCAGGAATTCATTGC  
ATCACTACGAAGGAAGAGTAGTGGCTTCTCAAGAGGAGCATCCATATATAGGGGTGTCACAAGGCATCATC  
AGCAAGGGCGCTGGCAAGCAAGAATAGGCCGTGTTGCTGGCAACAAAGATTTGTATCTCGGAACATTTGCT  
ACTGAAGAGGAAGCAGCCGAGGCATATGACATAGCAGCCATCAAGTTCAGGGGCGTGAATGCAGTCACCA  
ACTTTGAGATGAGTCGGTATAATGTGGAGGCCATCATCAACAGCAACCTTCCAGTAGGAGGCGCTGCCAAG  
CGACTGAACTAGCCACCGAGTCAAGCGATCAAAGGTCAGCACTAAGCCTGGAGCAGCAGTCCCAATCAA  
GCAACGGAAGTAGCGGTACAAGCAGCACCATCAACTTTGCCGCTGTTCAAGCAGCTTCTCCATTACATGCA  
ACGTGCCCTTTGACACAGCCACCGCCATATTCATCAAAAACCTTTCACAACTTTGGCTGCAATGGTGGCG  
GCAGTGGTGGGGCCCAAGTGACCGCTTCAGAGTCTTTCCGATGGCCCTCATGCCACAGACGTTATCAGAG  
TTCTACATATGGCCCACCAGTCTACTAACTGCGCAAGAACCAAGAAGCTAGAAGCTGTAGTTTAGCCAC  
CTAGTTTAGTTTAAGCTTAGTTCA

>ContigF|PLT3c|Cactassy\_W22B50\_12156\_Cov\_21|PREDICTED: AP2-like ethylene-responsive transcription factor AIL6-like [Vitis vinifera]~9.568E-110~XP\_003635541

GAACCTCCATTTACAGAGGAGTCACAAGGCATAGGTGGACTGGAAGATATGAAGCACATCTCTGGGATAAT  
AGCTGTAGGAGGGAAGGTCAGGCTAGGAAAGGGCGTCAAGTGTATTTAGGTGGGTATGACAAAGAAGAA  
AAGGCAGCCAGGGCTTATGACTTGGCAGCTCTAAAGTACTGGGGTCCCCTGCCACCACCAACTTCCCTATT  
TCGAATTACACCAAGGAGCTCGAGGAGATGAAACACGTGACCAAGCAAGAATTTATTGCATCACTCCGAAG  
GAAGAGTAGTGGCTTCTCAAGGGGCGCGTCCATATACAGGGGTGTCACAAGGCATCATCAGCAAGGCCGA  
TGCAAGCAAGGATTGGCCGTGTTGCAGGCAACAAAGATTTGTATCTCGGCACTTTTGGTACTGAAGAGGA  
GGCGGCCGAGGCATATGACATAGCAGCGATCAAGTTCAGGGGGGTGAATGCTGTACCAACTTTGAGATG  
AGCCGGTATGATGTTGAGGCCATCTCAAGAGCAACCTTCTATAGGAAGTGCAGGCAAGAGGCTGAAGCT  
CGCCACTGAGTCAAGTATCAGAAACCGGCACTCAGCAGTACTGCAAGTAGTGGCAGCAGCACCATCAACT  
TTGCTGCCATCCAGGGGGTCTCCCAATCTCATCCGCTGTGCCATATGACTGCCACAGCCATCTTCCATGA  
GAACCTCTGCATCACTTTGGCTCGACTGGTGGGGCCCAATGATTGGGTCTACCGAGCCCTTTGGGTGAGC  
TATTACCCAGTGGCTCTGATGCCACAGACGCCAGCTGAGTTTTACATATGGCCTCACCATCTTACTAACT  
GGCCGGAGTTAGGAAGGGCGTAGTTTAAAGTCATTAGTTCAGTTGATTAGTTAGTTAGTTAGTTGTAAGT  
GCTATTGATCTAAGAAAGCTAATTTTGGTTCCTAGAATGTCTCTGTTTTGACAGGTTTCTAATGGTCTTTT  
ACTGACCAACATCTTGATCTGACCTGACCCATTTTCGCTTGTAGCTTAGTGGTTGACCTTATGATAGAAT  
GAAAGAAGAGGGGGATTGGACTATGCTTTGGGGGTTGTATTAGAGGCAAGGAAGTACTGAAAATTTTGGT  
GTATGTATAGCTTTATTTCAATAGCTGGATGGTTATGACATTCGGTGAATGGGTCTTTGTACCAACTTATGC  
TCTGAAAAAGAAAAGAAGAG

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

**Secuencias de aminoácidos de los contigs traducidos de *P. pringlei* con similitud a PLT**

>PLT2A|15

TGRYEAHLWDNSCRREGQSRKGRQVYLGGYDKEEKAARAYDLAAIKYWGTSTTTNFPISNYEKEIEEMKHMTR  
QEFIAAIRRKSSGFSRGASIRGVTRHHQHGRWQARIGRVAGNKDLYLGTSTEEEAAYDIAAIKFRGLNAV  
NFDMSRYDVKAILESSTLPIGSGAAKRLKEAQMLEASRNNSFQYGSNSSSGRYGAHYLPIAYPLLYSQPQYQD  
HQPQALLTLQSQESNLSLHHNYFQEAQLLQPYQQSRNTSNPSIYNCSNLHSPNLLHGLISIGGSTTSGALDA  
NPNDGSPSGSYGGGG\*

>PLT2b|19|extraction

NSCRREGQSRKGRQVYLGGYDKEEKAARAYDLAAIKYWGTSTTTNFPISNYEKEIEEMKHMTRQEFVAAIRRKSS  
GFSRGASIRGVTRHHQHGRWQARIGRVAGNKDLYLGTSTEEEAAYDIAAIKFRGMNAVTFNFDMSRYDVK  
AILESSTLPIGGGAARKLKEAQALESSRRREEMFALATPTTFQYGSATSSSSRYGSHNQYPLLSYHQDHPQPLT  
LQNGGDQNISLSHYSQEAQFLQYQQSSSYSSNPSIYNAASYNGLHSPNLLQGLMSMHGVSSSGVVDAGT  
NNGSSSGSYGGYIGGGMGSTSSNAVGPMDPEPLAMVKVDYDMPSTGGGAGDGYGGWPTESVQGPNG  
GVFAMWND\*

>BBM|13|extraction

WTGRYEAHLWDNSCRREGQTRKGRQVYLGGYDKEEKAARAYDLAAIKYWGTSTTTNFPISDYEKEIEEMKHM  
RQEYVASLRRKSSGFSRGASIRGVTRHHQHGRWQARIGRVAGNKDLYLGTSTQEEAAAYDIAAIKFRGLNA  
VTNFELSRVDKAILESTLPIGGAAKRLKDVDDLTSAADTRHQILRVAAVTDHDNTNSQLSGDHAGVIPNFH  
PGIAYPQQAHPMSMHPYAQLRWCKQENEPDVGFNNGNYQDHLHHQLQLNSGGTHNFIQMHNLMGLESTS  
ASLEHSSGSNSIVYGNNGNDAGAYNGIHPFVIAPEGNGNGYMESNDNQGNYYYFSQQGNGGGGGLR  
AAVGTYDQGSACNNWVPTAIPALTPRTTMAAGAPLFTVWVNDT\*

>PLT3A|28|extraction

MSSSSSSCVVDYHDHEEDKPSKPKRARRTRAINPKSDEKCPISNNTPTTNRSSSIYRGVTRHRWTGRFEAHLWD  
KSCWNSIQNKKGRQVYLGAYDSEVDAARTYDLAALKYWGPETTLNFPNIENYKLEEMQNASKEEYLASLRRSS  
GFSRGVSKYRGVARHHHNGRWEARIGRVFGNKYLYLGTSSQEEAAAAYDLAAIEYRGPNAVTFDISNYIGRV  
KNLMPTGAQPQQPNTNKTQVQRRQQQQQQQIQKQEEDEEQDQPRVQEQDQLEVRKQEEEMATPSQ  
FEESKLDISSDDHSPMLVMGTHVDEDEDRENNPWSFCLDTGAYDPLLVPDIPEEDGRVLDLFDQTFEDNIDLIF  
EGEGNSITLDETYFGGTDVGCQTGMKPSSEDSNKKMQISLKIENSFSLSSPSSSTTSLVTCQQ\*

>PLT3b|31

MAPAAASNWLSFSLSPMEMLATSHHHHHHHHQTPQINLPTALPPSHHHYSAFFDFTFYSNDWANPKTQVMYA  
DYAEHHTTAVPKLEDFLGGDNNTSYNSQTEATTQDESSLTHLYDPHAAAANYFADHHHHHQDLNSLAENGFQ

## Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

SAAANSQSEVDDSCSVGRVTQLSCAGSEFVGQSIETELVAGFAVAASASASKSAEKALVSVAVDSDRAKKISDTF  
GQRTSIYRGVTRHRWTGRYEHLWDNSCRREGQARKGRQVYLGGYDKEEKAARAYDLAALKYWGPTATTNFP  
ISDYAKELEEMKNVTKQEFIASLRRKSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLYLGTFFATEEEAAE  
AYDIAAIKFRGVNAVTFEMSRYNVEAIINSNLPVGGAAKRLKLATESSDQRSALSLEQQSQSSNGSSGTSSTINF  
AAVQAASPITCNVPFDATAIFHQNLLHNLGCNGGGSGGAQVTASESFPMALMPQTLSEFYIWP HQSY\*

>PLT3C|21|extraction

TSIYRGVTRHRWTGRYEHLWDNSCRREGQARKGRQVYLGGYDKEEKAARAYDLAALKYWGPTATTNFPISNY  
TKELEEMKHVTKQEFIASLRRKSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLYLGTFFATEEEAAEAYDI  
AAIKFRGVNAVTFEMSRVDVEAILKSNLPIGSAKRLKLATESSDQKPAALSSTASSGSSSTINFAAIQGVLPISAVP  
YDTATAIFHENLLHHFGSTGGAQMIGSTEPFGSAITPVALMPQTPAEFYIWP HQSY\*

### Secuencias de proteínas PLT y AIL anotadas en la base de datos RefSeq usadas para en los análisis filogenéticos de las figuras 19 y 20

>Zm\_BBM\_extraction

QSLALSMSTGSHLPMVPSGAASGAASESTSENKRASGAMDSPGSAVEAVPRKSIDTFGQRTSIYRGVTRHR  
WTGRYEHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARAYDLAALKYWGTTTTTTFPISNYEKELEEMKHM  
TRQEYIAYLRRNSSGFSRGASKYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTEEAAEAYDIAAIKFRGLNA  
VTNFDMSRYDVKSILESTLPGGAARRLKDAVDHVEAGATIWRADMDGAVISQLAEAGMGGYASYGHHGW  
PTIAFQQPSPLSVHYYPYQPSRGWCKPEQDAAAAAAHSLQDLQLHLGSAAHNFFQASSSTVYNGGAGASG  
GYQGLGGSSFLMPSSTVVAADQGHSSSTANQGSTCSYGDHHEGKLGIDYDAAMVATAAGGDPYAAARNGY  
QFSQGSSTVSIARANGYANNWSSPFNNGM

>Vv\_PLT2\_like

NLQSLTSLMSGKSTCETSGNSSTSIVEAAPRRLDTFGQRTSIYRGVTRHRWTGRYEHLWDNSCRREGQSR  
KGRQVYLGGYDKEEKAARAYDLAALKYWGTTTTTTFPISNYERELEEMKHMTRQEFVASIRRKSSGFSRGASMY  
RGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTEEAAEAYDIAAIKFRGLNAVTFNFMNRYDVKSILENTLPI  
GGGAARKLKEAQAISSSRKREEMIALGSSSQYGSSSSRLQTYPLMQQFEQPQLLTLQNQEPLLTQNPEISQ  
YPQDSQFHQNYIQTQLQLHQSGSYLNHSSQSPQFYNSYLHNNPALLHGLMSMSGSSSSVMENNGSSSGSYNG  
GYFNNGLGVASNSTVASAVGSAEELPLIKVDYDMPAAGYGSWSGDSVQGGQAGVFTMWND

>Vv\_AIL6-like

GALS LGVTTTTTTQRSDQQAIVSADSDCSKKIADTFGQRTSIYRGVTRHRWTGRYEHLWDNSRREGQARKG  
RQVYLGGYDKEEKAARAYDLAALKYWGASATTNFPVSNYTKLEEMKHVTKQEFIASLRRKSSGFSRGASIRGV  
TRHHQQGRWQARIGRVAGNKDLYLGTFFATEEEAAEAYDIAAIKFRGVNAVTFEMNRYDVEAIANSALPIGGA  
AKRLKLSLESDQKPCNLNDQQTQSSASSSGITFGAIQPIAIPCGIPLDTATAFYHHNLFNHLHSSNMGASNSPGT  
TSSMPTPMALLPPPAEFFIWP HQSY

>Vv\_AIL5

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

ASFLRGFSSEQSEAQKQLALTPEPSPKKSVDTFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVY  
LGGYDKEEKAARAYDLAALKYWGPTTTTTNFPVSNYEKELENMKNMTRQEFVASLRRKSSGFSRGASIYRGVTRH  
HGHGRWQARIGRVAGNKDLYLGTFTSTQEEAAEAYDIAAIKFRGLNAVTFDMSRYDVKSIANSNLPIGGAIITG  
KPKTSPSDSASDSGSRREEQVQLSGPSNTLIFSKPLKQDHHQSVDYWSVLGYQASTLNSVVKNPILESSFPFQT  
SSNLTMDFSTASAGFLCNGGGYIQQQSNSSSSTASSNSIPYATPVNGLNSSTSIEVSGYNSWAAATAAAAAA  
AAAPSLHSFQAAPSLSVFQTPIFGME

>SI\_PLT2-like

NIQSLTSMGSGKSTSETSASPSANAATASATAENSNTSIVEAAPRRTLDTFGQRTSIYRGVTRHRWTGRYEA  
HLWDNSCRREGQSRKGRQVYVGGYDKEEKAARAYDLAALKYWGSTTTNFPISNYEKELEDMKHMTRQEFVA  
AIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTTEEEAAEAYDIAAIKFRGLNAVTFD  
MNRYDVKAILESNTLPIGGGAARKLKEAQALESSRKRQEMMALNSSFQYGNNSNPLQAYPLMQQHQPFDNS  
QPLLTQNHDISQYNIQDSSQFHQSYLQTLQLQNSHQVSLYNNYLQNNQVFLHGLMNNNEGSSSGSYSTG  
GYFGNSPGLGGMSSNSTSGNNGGGGAHEEVALVKVDYDNMPSYNGWSGESSVQGSNPGVFSMWNE

>SI\_BBM2-like

RSLSLMSMGDGGGDNCSSENNKQEI VSGNNDGQISGAIVPKKSIDTFGQRTSIYRGVTRHRWTGRYEAH  
LWDNSCRREGQTRKGRQVYVGGYDKEEKAARAYDLAALKYWGPTTTTTNFPISNYEKELEEMKHMTRQEFV  
RRKSSGFSRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSTQEEAAEAYDIAAIKFRGLNAVTFEINRY  
DVKSILESSLPIGGAAKRLKDVEQAEIALDYQRASQENIGTTHLMNGSMSAYGAGHCWPNITLQQAQPLSAM  
HYSNPYNSQQRLWCKQEVQDHDLT SQNFQDFQLGNAHNYLMGLDNSSSSSMEQNNVYTHQGVGYASNNN  
NHHGGGGFVLPHLNGSSSYGGENDNEVKQLQLGTHENIFGNYYHSQPATSNDSIKVSNLYDHDQESVCNSNN  
WMSTAVPMALASRTSTAACHGVAPTFTMWND

>SI\_AIL6-like

GALSLAVNAAAATTTNQSENEKAMVAVADSQSCCKISDFTFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRRE  
GQARKGRQVYVGGYDKEEKAARAYDLAALKYWGPTATTNFPVTNYTKELEEMKHMTKQEFIASLRRKSSGFSR  
GASIYRGVTRHHQQGRWQARIGRVAGNKDLYLGTFTSTQEEAAEAYDIAAIKFRGVNAVTFEMNRYDVEAIMK  
SSLVPGGTAKRLKLSLESEQSSSSNNNNQQQMTQCNSNNINFGAMPVSAIPCGVPFDNTAPAFYHHNFF  
QYLHPGNASVPDASNATTTMAAAMP LLPSAADLFIWPHQSY

>SI\_AIL5-like

AATFLPTTFSDDHKEKLPLIVTQPKKPHESEFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRRQGQSRKGRQGGY  
DKEEKAARAYDLAALKYWGPTTTTTNFPVSNYERELEEMKKNMTRQEFVASLRRKSSGFSRGASIYRGVTRHHQH  
GRWQARIGRVAGNKDLYLGTFTSTQEEAAEAYDIAAIKFRGLNAVTFDISRYDVKSIASSNLPVGGMSNKS  
DETTHVENRPNLRDQDQDRDISSAPNYTSKNLAMVNLGFLPIKQDAFDWSSVGYNNNRSKNGLLFQGT  
MNAQGTTFSSNEVNSSVMLNEQGYDNHDQQQSGSSCEINMALNSNIATTSTTIDSTNFGNWMTPSLHS  
FQSSTKNDLGPYHTPIFGME

>Mt\_BMM

QTLSLSMSTGSQSSSVPLLNANVMSGEISSSENKQPPTTAVVLDNSQTSVVESAVPRKSVDTFGQRTSIYRGV  
TRHRWTGRYEAHLWDNSCRREGQTRKGRQVYVGGYDKEEKAARAYDLAALKYWGTTTTNFPISHYEKEVEEM  
KHMTRQEFVVASLRRKSSGFSRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSTQEEAAEAYDVA  
RGLSAVTFDMSRYDVKTILESSLPIGGAAKRLKDMEQVELNHVNVDISHRTEQDHSIINNTSHLTEQAIYAATN

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

ASNWHALSFQHQQPHHHYNANNMQLQNYPYGTQTQKLWCKQEQSDDHSTYTTATDIHQLQLGNNNNNT  
HNFFGLQNMISMDSASMDNSSGSNSVYGGGDHGGYGGNGGYMIPMAIANDGNQNPNSNNNFGESEIKGF  
GYENVFGTTTTDPYHAQAARNLYYQPQQLSVDQGSNWVPTAIPRTLAPRTTNVSLCPPFTLLHE

>Gm\_PLT2

SLQSLTSMGSGKDSTCETSGDNSTNTTTTTVEAAPRRLDTFGQRTSIYRGVTRHRWTGRYEAHLWDNSCR  
EGQSRKGRQVYLGGYDKEEKAARSYDLAALKYWGTTTTNFPISNYEKELDEMKNMTRQEFVAAIRKSSGFSR  
GASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTEEAAEAYDIAAIKFRGLNAVTFDMSRYDVKAILE  
SNTLPIGGGAARKLKEAQALESRRKREEMIALGSSTFYQGTSSNSRLHAYPLMQHHHQFEQPQLLTLQNHDIS  
SHFSHQDPLHQGYIQTQLQLHQQQSGSSSYFQNNNINNAQFYNGYNLQNHHPALLQGMINMGSSSSSVL  
ENNSNNNNVGGFVGSFGMASNATSGNTVGTAEELGLVKVDYDMPTGGYGGWSAAAAESMQTSNSGVF  
TMWND

>Gm\_BBM1

QTLSLMSTGSSSTSLPLLTASVDNGESPNDKQPNTSAALDSTQTGAJETAPRKSIDTFGQRTSIYRGVTRHRW  
TGRYEAHLWDNSCRREGQTRKGRQVYLGGYDKEEKAARAYDLAALKYWGTTTTNFPISHYEKELEEMKNMTR  
QEYVASLRRKSSGFSRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTQEEAAEAYDVAAIKFRGLSAV  
TFDMSRYDVKSILESTTLPIGGGAARKLKDMEQVELSVDNGHRADQVDHSIIMSSHLTQGINNNYAGGGTATH  
HNWHNAHAFHQPPCTTMHYPYQQRINWCKQEQQDNDAPHLSYSYDIHQLQLGNNGTHNFFHTNSGLHP  
MLSMDSASIDNSSSSNSVYDGYGGGGGYNVMPMGTAVVASDGDQNPNSNHGFGDNEIKALGYESVYGS  
ATDSYHAHARNLYLTQQQSSVDTVKASAYDQGSACNTWVPTAIPHTAPRSTTSMALCHGATTPFSLLE

>Gm\_AIL6-like

GTLSLAVALSSEKAVVAAESNSSKKIVDTFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRREGQARKGRQVYLG  
YDKEEKAARAYDLAALKYWGPTATTNFPVSNYSKEVEEMKHVTKQEFIASLRRKSSGFSRGASIYRGVTRHHQ  
GRWQARIGRVAGNKDLYLGTFFATEEEAAEAYDIAAIKFRGANAVTFEMNRYDVEAIMKSSLPVGGAAKRLRLS  
LESEKAPPVNSSSQQNPQCGNVSGSINFSAIHQPIASIPCGIPFDSTTAYYPHNLQHFHPTNAGAAASAVTS  
ANATALTALPASAATEFFIWP HQSY

>Gm\_AIL5-like

ACFPRAFAAEPTTEPQKPSPKKTVDTFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVYLG  
KEDKAARAYDLAALKYWGPTTTTTNFPISNYEKELEEMKNMTRQEFVASLRRKSSGFSRGASIYRGVTRHHQ  
HGRWQARIGRVAGNKDLYLGTFFSTQEEAAEAYDIAAIKFRGLNAVTFDMSRYDVKSIANSTLPIGGLSGKNK  
NSTDSASEKSHEASRSDERDPSAASSVTFASQQQPSSTLSFAIPIKQDPSDYWSILGYHNSPLDNTGIRNTTS  
VTATSFPSSNNGTTSSLTPFHMEFSNAPTSTGSDNDAAFFSGGGIFVQQQSGHNGHSGSSSSSSLSCSIPFAT  
PIFSLNSNTSYENSAGYGNWIGPTLHTFQSHAKPSLFQTPIFGME

>Fv\_PLT2\_like

SLQSLTSMGSGKGSSTCETSADNTSTATVEAPRRALDTFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRREG  
QSRKGRQVYLGGYDKEEKAARAYDLAALKYWGTTTTNFPISNYEKEVEEMKNCTRQEFVASIRKSSGFSRGAS  
MYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTEEAAEAYDIAAIKFRGLNAVTFDMSRYDVKSILENTL  
PIGGGAARKLKEAQALETSRKREEMIALGSTFQYAAGASSASSAARNLQAYSMLQQQPPQSAYDQTLLSLQN  
HDMSQYTHHDPSSYQNYIQTQLQLQQSSNSYNIHQSSQSPHPYYNTYLQSTHPALLQGLMDMGTASTTVM  
DSSAGGYLGHNGLGMASNSTTTNGVGSAAELALVKVDYDLPSSGGYGSWSGGSVQGSNPGVFTMWNE

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

>Fv\_AIL6-like

NALSLGVANTTANTQSSNREKAIVAVENDSNKKISDTFGQRTSIYRGVTRHRWTGRYEHLWDNSCRREGQAR  
KGRQGGYDMEDKAARAYDLAALKYWGPTATTNFPYSTYSKELEDMLVTKQEFIASLRRKSSGFSRGASMYRGV  
TRHHQQGRWQARIGRVAGNKDLYLGTFFATEEEAAEAYDIAAIKFRGINAVTNFEMNRYDVEAISKSSLPVGSSA  
KRLKISLESEEQKSSVNNRDLHQQQNSQCSSGSNSNTINFANSMAQPIQTIPCGIPYDHAAAFYHHNLLQQFQSY  
NYYGTTSDSSAVSQPNIATSQMTMAMMPQQQEQQQPAEFFLWPQYQTHQS

>Fv\_AIL5-like

ASFLRGFSTTPTATATAPTTLQQTHTHPVGTETSTAKKTVDTFGQRTSIYRGVTRHRWTGRYEHLWDNSCR  
REGQSRKGRQVYLGGYDKEDKAARAYDLAALKYWGPTTTTTFPVTNYEKELTEMKNMSRQEFVSLRRKSSGF  
SRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTQEEAAEAYDIAAIKFRGLNAVTFDMSRYDVKSIA  
NSNLPIGGMSGKSKNSSDQSLSDSKSLDDRDLSSASSVTFASSSQQPASNSSTLSFAIPIKQDPSTEYWSNIFGY  
QVPNNNPSANNVRNPNVSGAAPTYPFHSNMDFSATSTATTTTTETNGLFYSGIYNVHGQQQQQQEQQQQSTS  
SGTNNPIPYSTPLGLSSNMGYDQSSSPGYGTWIGPGLNLFQTFIFGME

>Cs\_PLT2\_like

CLQSLTSMGSGKPKSTCETTSTPDNNTTSNNSNNNSNTTLDVTPRRLDFTFGQRTSIYRGVTRHRWTGRYEHL  
WDNSCRREGQSRKGRQVYLGGYDKKEKAARAYDLAALKYWGTTTTNFPISNYEKEVEEMKHMTRQEFVAAIR  
RKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTEEAAEAYDIAAIKFRGLNAVTFDMSR  
YDVKSILESNTLPIGGGAARKLKEAQAVESSRKRDEMIALGSSSSSSCFQYGTSSSSTNSSHYPNLLQQPNLND  
HHHLQTQPLLSLQNHHDISHYSTHHPSSFHNPSSSYIHHSSDHSSYPNNNNNHPFYGAGYLHNHPALLHGMIN  
MSGGGGGYLGNAFGIGSASGSTAEEYALVKVDYDMPNSGGYGGWGTGDSVQGSNAGVFSMWND

>Cs\_BBM2-like

HTLSLSMNTGPPPPSQSSSGSAALPLLTASGGESSSDNKQKSSGASIDAENGAVEAAPRKSVDFTFGQRTSIYR  
GVTRHRWTGRYEHLWDNSCRREGQTRKGRQVYLGGYDKKEKAARAYDLAALKYWGTTTTNFPISDYEKELE  
DMKHMTRQEFVSLRRKSSGFSRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFGTQEEAAEAYDIAAI  
KFRGLNAVTFDMSRYDVKAILESNTLPIGGGAARKLDIHHSDIPLDPQRAEDNITSHLSDGIINITSTTTPYSGGG  
WPTIAFHHPYAFHYPYATPQQRMWCKQEQDATAIAADNFFPHGGAMDSDMHSSGSYSSANGDGYNGNF  
LIPMGAEGSSSNNGGLGDIGEVNLFGGCSNVDDPFHATRTSNLYNYHNSHQLOPPPPPSGLQGSNCNNWLP  
PARSVCQGGAPPFTIWNNDT

>Cs\_AIL6-like

GALSLAVAQSSDTAPAPVVAVDSDSSKKIADFTFGQRTSIYRGVTRHRWTGRYEHLWDNSCRREGQARKGRQ  
VYLGGYDKKEKAARAYDLAALKYWGPTATTNFPVSNYAKELEEMKQVTRQEFIASLRRKSSGFSRGASIYRGVTR  
HHQQGRWQARIGRVAGNKDLYLGTFFATEEEAAEAYDIAAIKFRGLNAVTFEMSRDYVEIAIKSALPIGGAAKR  
LKLCELDQKPIPNHDQATQCSSGSNNINFGTAMQAVPPPCGIPYDTAAVLYHHNYFHHLQPNIAIGSSESTSPG  
IAVPGTVGPHQAAEFFVWPHQSY

>Cs\_AIL5

ASFPTAAAAYSTRYPSAPSPKAVDFTFGQRTSIYRGVTRHRWTGRYEHLWDNSCRREGQSRKGRQVYLGGYD  
KEEKAARAYDLAALKYWGPTTTTTFPVSNYEKELEDKMNMTQEFVSLRRRSGFSRGASIYRGVTRHHQHGR  
RWQARIGRVAGNKDLYLGTFFSTQEEAAEAYDIAAIKFRGLNAVTFDISRYDVKSIASSNLPIGGMSGAKSKTAS

Anexo I

Secuencias de contigs de *P. pringlei* y proteínas anotadas en RefSeq con similitud a PLT

DSAASDGGSRSTDERDVHHSPPSSSTSTFISSSSQPNNSSTLSFTMPIKQDPSDQYWSILSYNPDAFNANLPKR  
DDNNVPVPLFQQPESTMFAITELGSSGSGMTEGGMYVQQQQQYGTMPMAFAKASFFQTPIFGME

>At\_PLT3\_AIL6

GALSLGVNVNNTNHRNDNDNHYRGNNGERINNNNNNDNEKTDSEKEKAVVAVETSDCSNKKIADTFGQR  
TSIYRGVTRHRWTGRYEAHLWDNSCRREGQARKGRQVYLGGYDKEDKAARAYDLAALKYWNATATTNFPITN  
YSKEVEEMKHMTKQEFIASLRRKSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLYLGTFFATEEEAAEAY  
DIAAIKFRGINAVTNFEMNRYDVEAIMKSALPIGGAARKRLKLSLEAAASSEQKPILGHHQLHHFQQQQQQQLQ  
LQSSPNHSSINFALCPNSAVQSQQIIPCGIPFEAAALYHHHQQQQHQQQQQQQNFFQHFPANAASDSTGSN  
NNSNVQGTMGMLMAPNPAEFFLWPNQSY

>At\_PLT2

NLQSLTSMGSTGAAAAEVATVKASPAETSADNSSSTNTSGGAIVEATPRRTLETFGQRTSIYRGVTRHRWTGR  
YEAHLWDNSCRREGQSRKGRQVYLGGYDKEEKAARAYDLAALKYWGPSTTTNFPITNYEKEVEEMKNMTRQE  
FVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTEEAAEAYDIAAIKFRGLNAV  
TNFEINRYDVKAILESNTLPIGGGAARKLKEAQALESSRKREEMIALGSNFHQYGAASGSSSVASSRLQLQPYPLSIQ  
QPFEHLHHHQLLTLQNNNDISQYHDSFSYIQTQLHLHQQQTNNYLQSSSHTSQLYNAYLQSNPGLLHGFVSDN  
NNTSGFLGNGIGIGSSSTVGSSAEEFPAVKVDYDMPPSGGATGYGGWNSGESAQGSNPGGVFTMWNE

>At\_BBM

GARGLSLSMNSSTSDSNYNNDVQEKTIQVVDVETPKKTIESFGQRTSIYRGVTRHRWTGRYEAHLWDNS  
CKREGQTRKGRQVYLGGYDKEEKAARAYDLAALKYWGPSTTTNFPVSEYEKEVEEMKHMTRQEYVVASLRRKSS  
GFSRGASIRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTQEEAAEAYDIAAIKFRGLSAVTNFDMNRYNVK  
AILESPSLPIGSSAKRLKDVNNPVPAMMISNNVSEANNVSGWQNTAFQHHQGMDLSLLQQQERYVGYNG  
GNLSTESTRVCFKQEEEQHFLRNSPSHMTNVDHHSSTSDSVTCGNVVSYGGYQGFAIPVGTSVNYDPFTA  
AEIAYNARNHYHYAQQHQQQQIQQSPGGDFPVAISNNHSSNMYFHGEGGGEGAPTFSVWNT

>At\_AIL7

GVLSLGVNNTSDQPLSCNNGERGGNSNKKKTVSKKETSDDSKKIVETLGQRTSIYRGVTRHRWTGRYEAHLW  
DNSCRREGQARKGRQVYLGGYDKEDRAARAYDLAALKYWGSTATTNFPVSSYSKEEEMNHMTKQEFIASLRR  
KSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLYLGTFFATEEEAAEAYDIAAIKFRGINAVTNFEMNRYD  
IEAVMNSSLPVGGAAAKRHKLKLALLESPPSSSDHNLQQQLLPSSPSDQNPNSIPCGIPFEPVLYYHQNFQ  
YPLVSDSTIQAPMNQAEFFLWPNQSY

>At\_AIL5

ASFLGNYSGGHSSEVSVHKQQPNPLAVSEASPTPKKNVESFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRRE  
GQSRKGRQVYLGGYDKEDKAARAYDLAALKYWGPSTTTNFPISNYESELEEMKHMTRQEFVASLRRKSSGFSRG  
ASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFFSTQEEAAEAYDIAAIKFRGLNAVTFDISRYDVKSIASCN  
LPVGGMLPKPSPATAAADKTVDLSPSDPSLTPSLTFNVATPVNDHGGTFYHTGIPIKDPADHYWSNIFGFQA  
NPKAEMRPLANFGSDLHNPSPGYAIMPVMQEGENNFSGSFGSDGYNNHSAASNVPVSAIPLSSTTTMSNGNE  
GYGGNINWINNNISSYQTAKSNSVLHTPVFGL