



# **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

## **Maestría y Doctorado en Ciencias Bioquímicas**

EXPLORACIÓN MEDIANTE RNAseq DE LA REGULACIÓN GENÉTICA DEL  
CRECIMIENTO DETERMINADO DE LA RAÍZ PRIMARIA DE LA CACTÁCEA  
*Pachycereus pringlei*

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. GUSTAVO RODRÍGUEZ ALONSO

TUTOR PRINCIPAL

DRA. SVETLANA SHISHKOVA

INSTITUTO DE BIOTECNOLOGÍA - UNAM

MIEMBROS DEL COMITÉ TUTOR

DRA. ELIZABETH CORDOBA MARTÍNEZ – IBt-UNAM

DRA. VERÓNICA LIRA RUAN – CIDC - UAEM

CUERNAVACA, MOR.; SEPTIEMBRE, 2018



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.



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

Este proyecto fue parcialmente financiado por los donativos CONACyT (CB240055, FOINS219522) y UNAM-DGAPA- PAPIIT (IN207115, IN201318, IN200818).

A Gustavo Rodríguez Alonso se le otorgó una beca doctoral del CONACyT (registro 290654) y recibió una beca para realizar una estancia en la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba-CONICET, en el marco del proyecto bilateral CONACyT-FOINS 219522; así como una beca por parte del proyecto UNAM-DGAPA-PAPIIT IN207115. Además, recibió financiamiento mediante el Programa de Apoyo a los Estudiantes de Posgrado (PAEP-UNAM) para asistir a un congreso nacional y uno internacional.



# **Agradecimientos**

## **A mis mentores:**

La **Dra. Svetlana Shishkova**, la **Dra. Marta Matvienko** y el **Dr. Joseph G. Dubrovsky**.

Por dar inicio a este proyecto, el cual me ha permitido aprender nuevas aproximaciones para hacer ciencia, y por proporcionar el ambiente adecuado para el desarrollo del pensamiento crítico y analítico. Especialmente gracias a la **Dra. Shishkova** por permitirme crecer académica y profesionalmente, y por propiciar y fortalecer mi autonomía intelectual y en la toma de decisiones.



## **Por el apoyo técnico:**

Al **Dr. Francisco Molina-Freaner**, por la donación de semillas de *Pachycereus pringlei* y por recibirnos y proporcionar las instalaciones de su laboratorio para la limpieza de semillas que colectamos en el Desierto de Sonora.

Al **Q.B. José Fulgencio Martínez Rodríguez**; por su apoyo en la colecta de frutos de *P. pringlei* en el Desierto de Sonora y por sus enseñanzas respecto a la colecta de frutos, las técnicas de trabajo en campo y la limpieza de semillas.



## **A mis colaboradores:**

Quienes compartieron su tiempo y dedicación para el desarrollo de este proyecto:

A la **I.B.I Marcela Ramírez Yarza**, por su valioso apoyo técnico en el laboratorio, incluyendo su participación en la preparación de medios de cultivo, esterilización y germinación de semillas, así como en la colecta del material biológico que se utilizó en este trabajo.

A la **Dra. Selene Napsucialy Mendivil**, por su apoyo con la purificación de RNA y los ensayos de RT-qPCR.

Al **Dr. Pedro Eduardo Lázaro Mixteco**, por su apoyo con los ensayos de RT-qPCR.

A la **M.C. Mayra Liliana López Valle**, por su apoyo con los ensayos de RT-qPCR.

A los jóvenes que realizaron servicio social: A la Tec. Lab. Daniela Fernanda Pecina Ornelas y a la Tec. Lab. Angélica Contreras Valladares; así como a Sinhué Raphael Albarrán Hernández y Ramsés Uriel Albarrán Hernández. Gracias por su confianza, tiempo y entusiasmo.



# **Agradecimientos**

## **A los miembros de mi comité tutor:**

La Dra. Verónica Lira Ruan y la Dra. Elizabeth Cordoba Martínez.  
Por sus evaluaciones, críticas y comentarios en nuestras reuniones semestrales. Su pericia y experiencia abonaron a la robustez y calidad de este proyecto.



## **A los miembros de mi jurado de tesis:**

Al Dr. Enrique Merino Pérez, Dr. Arnaud Jacques Marcel Ronceret, Dra. Viviana del Carmen Valadez Graham, Dr. Ulises Yunuén Rosas López y el Dr. Víctor Manuel Chávez Ávila.  
Por su tiempo, críticas, correcciones y cuestionamientos en la revisión de este escrito de Tesis.



## **A las Unidades y Servicios del IBt.**

A la **Unidad de Cómputo**, particularmente al M. en T.I. Juan Manuel Hurtado Ramírez, por su apoyo con la gestión y manejo de datos masivos.

A la **Unidad de Biblioteca**, particularmente a la B.A. Dip.Lib. A.L.A. Shirley Elizabeth Ainsworth Gore, por su apoyo para conseguir material bibliográfico de difícil acceso.

A la **Unidad de síntesis y secuenciación de ADN**, particularmente al Dr. Rubén Paul Gaytán Colín y el M.C. Eugenio López Bustos, por su accesibilidad, profesionalismo y apoyo técnico con la síntesis de oligonucleótidos.

A la **Unidad de Docencia y Formación de recursos humanos**, particularmente al Ing. Jalil Saab Hassanille, al Lic. J. Antonio Bolaños Guillén y a la Oficinista de Docencia Gloria Villa Herrera. Por su apoyo y gestión administrativa durante mi estancia en el posgrado.



## **A mis compañeros del laboratorio:**

Cuyos comentarios, observaciones y discusiones durante los seminarios de grupo fortalecieron este proyecto e hicieron disfrutable mi estancia en el posgrado: Leslie Acevedo Sánchez, Aylin Ávila Linares, M.C. Ramsés de Jesús García, Felipe Hernández Bermúdez, Dra. Blanca Jazmín Reyes Hernández, M.C. Héctor Hugo Torres Martínez y Biol. Aranza Xhaly Quintana Armas.



## ***Dedicatorias***

A los mexicanos, cuyos impuestos permitieron mi acceso a la educación pública y financiaron todos los recursos necesarios para el desarrollo de esta Tesis.

A quienes adoptan y promueven el acceso abierto al conocimiento, tanto en los esquemas de publicación Open Access como en el desarrollo de Open Source y la gestión del Open Data.

A mi familia. Por ser mi primera escuela.



# CONTENIDO

LISTA DE ABREVIACIONES .....	i
RESUMEN.....	iii
SUMMARY .....	v
I. INTRODUCCIÓN .....	1
I.1 GENERALIDADES SOBRE LA RAÍZ.....	1
I.1.1 El Meristemo Apical de la Raíz .....	2
I.1.2 El RAM mantiene el crecimiento de la raíz .....	2
I.2 CRECIMIENTO DETERMINADO DE LA RAÍZ.....	4
I.3 REGULACIÓN GENÉTICA DE LA ACTIVIDAD DEL RAM .....	5
I.3.1 Vía PLT .....	6
I.3.1 Vía SHR y SCR .....	7
I.3.3 Vía WOX5-CLE40 .....	8
II. JUSTIFICACIÓN.....	9
III. OBJETIVOS.....	11
IV. MATERIALES Y METODOS .....	12
IV.1 OBTENCIÓN DE MUESTRAS BIOLÓGICAS .....	12
IV.2 RNAseq Y ENSAMBLAJE DEL TRANSCRIPTOMA DE <i>Pachycereus pringlei</i> .....	13
IV.3 ANOTACIÓN DEL TRANSCRIPTOMA.....	14
IV.4 ANÁLISIS DE EXPRESIÓN DIFERENCIAL.....	14
IV.5 VALIDACIÓN DEL RNAseq MEDIANTE RT-qPCR .....	15
IV.6 TRANSCRIPTÓMICA COMPARATIVA Y REDES DE REGULACIÓN TRANSCRIPCIONAL.....	15
V. RESULTADOS Y DISCUSIÓN.....	17
V.1 OBTENCIÓN DE MATERIAL BIOLÓGICO .....	17
V.2 SECUENCIACIÓN Y ENSAMBLAJE DEL TRANSCRIPTOMA .....	21
V.2.1 Secuenciación del transcriptoma .....	21
V.2.2 Ensamblaje del transcriptoma.....	24
V.3 ANOTACIÓN FUNCIONAL DE LOS TRANSCRITOS.....	25
V.4 CARACTERIZACIÓN DEL TRANSCRIPTOMA DEL ÁPICE DE LA RAÍZ PRIMARIA DE <i>Pachycereus pringlei</i> .....	29
V.5. EVALUACIÓN DE LA EXPRESIÓN DIFERENCIAL .....	33

V.5.1 Análisis de conglomerados .....	35
V.6 VALIDACIÓN DEL RNAseq MEDIANTE RT-QPCR .....	38
V.6.1 Selección de genes normalizadores para ensayos de RT-qPCR.....	38
V.6.2 Ensayo de eficiencia para oligonucleótidos de RT-qPCR.....	39
V.6.3. Cuantificación de la expresión relativa de genes de interés.....	42
V.7 CONSERVACIÓN DE GENES EN LA FAMILIA CACTACEAE.....	43
V.8 RED DE REGULACIÓN TRANSCRIPCIONAL PARA EL ÁPICE DE LA RAÍZ DE <i>Pachycereus pringlei</i> .....	46
VI. CONCLUSIONES.....	51
VII. PERSPECTIVAS.....	53
VIII. BIBLIOGRAFÍA.....	54
VIII. ANEXOS.....	62
1. ACCESO PÚBLICO DE LOS DATOS GENERADOS EN ESTA TESIS.....	62
A) REPORTE GENERAL DE LA SECUENCIACIÓN POR RNAseq .....	62
B) ACCESIÓN EN GEO-OMNIBUS .....	62
2. ANÁLISIS DE CALIDAD DE LAS LECTURAS QUE SE GENERARON A PARTIR DE LA SECUENCIACIÓN POR RNAseq .....	63
A) PHRED Score por posición. ....	63
B) Calidad por secuencia .....	64
C) Contenido de GC por secuencia .....	65
D) Distribución de nucleótidos por base.....	66
E) Secuencias cortas (k-mers) sobre-representadas.....	67
3. DEFINICIÓN DE PARÁMETROS DURANTE EL ENSAMBLAJE <i>DE-NOVO DEL</i> <i>TRANSCRIPTOMA</i> .....	68
4. ALGUNAS RUTAS Y PROCESOS BIOLÓGICOS CONTENIDOS EN EL TRANSCRIPTOMA DEL ÁPICE DE LA RAÍZ PRIMARIA DE <i>Pachycereus pringlei</i> .....	70
5. PERFILES DE TRANSCRIPCIÓN IDENTIFICADOS MEDIANTE <i>K-MEAN CLUSTERING</i> . .....	80
6. RED DE REGULACIÓN TRASCRIPTACIONAL .....	82
A) Tabla de interacciones .....	82
B) Tabla de propiedades.....	87
7. PUBLICACIONES DERIVADAS DE ESTA TESIS.....	92



## ÍNDICE DE FIGURAS Y TABLAS

Figura 1. Estructura y organización de la raíz primaria.....	3
Figura 2. Etapas del crecimiento de la raíz primaria de <i>Pachycereus pringlei</i> .....	18
Figura 3. Integridad del RNA de las muestras que se secuenciaron.....	20
Figura 4. Calidad de las lecturas crudas de secuenciación de RNAseq.....	22
Figura 5. Calidad de las lecturas de RNAseq después de su procesamiento.....	23
Figura 6. Especies de referencia a partir de las cuales se anotaron a los transcritos de <i>Pachycereus pringlei</i> .....	26
Figura 7. Anotación funcional de los transcritos de <i>Pachycereus pringlei</i> .....	28
Figura 8. Variabilidad de los niveles de expresión (RPKM) de transcrito entre réplicas biológicas.....	29
Figura 9. Gráfica de Scree de los componentes principales del transcriptoma de <i>Pachycereus pringlei</i> .....	31
Figura 10. Análisis de componentes principales del transcriptoma de <i>Pachycereus pringlei</i> .....	31
Figura 11. Expresión diferencial entre las etapas del desarrollo de la raíz primaria de <i>Pachycereus pringlei</i> .....	34
Figura 12. Análisis hipergeométrico y de conglomerados de los transcritos con expresión diferencial entre las etapas inicial y terminal del desarrollo de la raíz primaria de <i>Pachycereus pringlei</i> .....	36
Figura 13. Niveles de expresión (RPKM) de los posibles genes de referencia de <i>Pachycereus pringlei</i> para ensayos de RT-qPCR.....	41
Figura 14. Determinación de la eficiencia de amplificación de los oligonucleótidos para genes de referencia de <i>Pachycereus pringlei</i> .....	41
Figura 15. Abundancia relativa de algunos transcritos de interés.....	43
Figura 16. Conservación de secuencias nucleotídicas entre <i>Pachycereus pringlei</i> y <i>Lophophora williamsii</i> .....	45

Figura 17. Tamaño de los ORF putativos en transcritos de <i>Pachycereus pringlei</i> no anotados y con expresión diferencial.....	46
Figura 18. Red de regulación transcripcional inferida para el ápice de la raíz primaria de <i>Pachycereus pringlei</i> .....	50
Figura 19. Módulos de regulación transcripcional del ápice de la raíz primaria de <i>Pachycereus pringlei</i> .....	50
Tabla I. Semillas de <i>Pachycereus pringlei</i> disponibles para la obtención de material biológico.....	17
Tabla II. Caracterización espectrofotométrica del RNA que se extrajo de los ápices de la raíz primaria de <i>Pachycereus pringlei</i> .....	19
Tabla III. Comparación entre los transcriptomas ensamblados <i>de novo</i> de <i>Pachycereus pringlei</i> y <i>Lophophora williamsii</i> .....	25
Tabla IV. Expresión diferencial de genes entre las zonas de desarrollo de la raíz primaria de algunas angiospermas.....	35
Tabla V. Expresión (RPKMs) de los posibles genes de referencia de <i>Pachycereus pringlei</i> para ensayos de RT-qPCR.....	40
Tabla VI. Oligonucleótidos diseñados para la amplificación de los genes de referencia que se utilizaron en este trabajo.....	40
Tabla VII. Conservación de las secuencias nucleotídicas ensambladas para los transcriptomas de <i>Pachycereus pringlei</i> y <i>Lophophora williamsii</i> .....	44

## LISTA DE ABREVIACIONES

- ACR4 – CRINKLY 4.
- AP2 – Dominio Apetala 2.
- BBH – Mejor hit bidireccional (Best Bidirectional Hit).
- BBM – BABY-BOOM, también PLETHORA 4.
- BGI – Beijing Genomics Institute.
- BLAST – Basic Local Alignment Search Tool .
- cDNA – Ácido desoxirribonucleico complementario a una cadena de mRNA (complementary DNA).
- CLE – CLAVATA/ESR-RELATED.
- CLV – CLAVATA.
- CQ – Centro quiescente.
- d.d.g – Días después de la germinación.
- dF/dT – Diferencial de la fluorescencia con respecto al tiempo.
- DNA – Ácido desoxirribonucleico (Deoxyribonucleic acid).
- EDGE – Detección empírica de la expresión diferencial (Empirical detection of differential gene expression).
- EF1 $\alpha$  – TRANSLATION ELONGATION FACTOR 1 $\alpha$
- EREBP - Ethylene-responsive element binding protein.
- E-value – Valor esperado (Expected-value).
- FDR – Tasa de descubrimientos falsos (False discovery rate).
- GAI - GIBERELLIC ACID INSENSITIVE.
- GASA – GA-STIMULATED ARABIDOPSIS.
- GH3 – GRETCHEN HAGEN 3.
- H3K4me3 – Trimetilación de la lisina 4 de la histona 3.
- HD2A – HISTONE DEACETYLASE 2A.
- HD2B - HISTONE DEACETYLASE 2B.
- HSP – High-scoring segment pair.
- IAA14 – INDOLE-3-ACETIC ACID INDUCIBLE 14.
- JKD – JACKDAW.
- KAAS – Servidor de anotación automática del KEGG (KEGG automatic annotation server).
- Kb – Kilobases.
- KEGG – Enciclopedia de genes y genomas de Kyoto (Kyoto encyclopedia of genes and genomes).
- LFY3 – LEAFY 3.
- mRNA – Ácido ribonucleico mensajero (messenger RNA).

NAC – No apical meristem domain containing protein.

nt – Nucleótidos

OneKP – One thousand plants initiative.

ORF – Marco abierto de lectura (Open reading frame).

PCA – Análisis de componentes principales (Principal component analysis).

PCR – Reacción en cadena de la polimerasa (Polymerase chain reaction).

PLT – PLETHORA.

RAM – Meristemo apical de la raíz (Root apical meristem)

RBR – Retinoblastoma related.

RGA – REPRESSOR OF GA1

RIN – Número de integridad del RNA (RNA integrity number).

RNA – Ácido ribonucleico (Ribonucleic acid).

ROW – Repressor of WUSCHEL.

RPKMs – Lecturas por kilobase por millón de lecturas mapeadas (Reads per kilobase per million mapped reads).

rRNA – RNA ribosomal (ribosomal RNA).

RT-qPCR – PCR cuantitativo a partir de cDNA (Retrotranscribed quantitative PCR).

SAM – Meristemo apical del brote (Shoot apical meristem).

SCR – SCARECROW.

SHR – SHORT-ROOT.

SUR1 – SUPERROOT.

TDEs – Transcritos diferencialmente expresados.

UBC – UBIQUITIN CONJUGATING ENZYME.

WOX - WUSCHEL RELATED HOMEODOMAIN.

WRKY - WRKY DNA-BINDING PROTEIN.

WUS – WUSCHEL.

YLS – YELLOW LEAF SPECIFIC.

## RESUMEN

En la mayoría de las especies de plantas las raíces pueden crecer de manera indefinida, o indeterminada, mientras el Meristemo Apical de la Raíz (RAM, por sus siglas en inglés) está presente y activo. Sin embargo, la subfamilia Cactaceae representa un caso único entre las plantas angiospermas, pues las células que componen al RAM de la raíz primaria dejan de dividirse después de pocos ciclos celulares, lo cual conduce al agotamiento del RAM pocos días después de la germinación. Como consecuencia del agotamiento del meristemo todas las células en el ápice de la raíz se diferencian y el crecimiento de la raíz cesa. A este patrón de crecimiento lo denominamos crecimiento determinado. El crecimiento determinado de la raíz primaria es un programa genético estable y se ha propuesto que representa una ventaja adaptativa en los ambientes áridos, tales como los desiertos.

En esta tesis se utilizó a la cactácea desértica *Pachycereus pringlei*, como modelo de estudio, para entender cuáles son los mecanismos que subyacen al crecimiento determinado de la raíz primaria. *P. pringlei* es una cactácea columnar endémica del Desierto de Sonora. El meristemo apical de la raíz primaria de *P. pringlei* se agota a los 7-10 días post-germinación. En este trabajo se reportan los resultados de la secuenciación del transcriptoma del ápice de la raíz primaria de *P. pringlei* y la evaluación de la expresión diferencial de genes en tres etapas de desarrollo de la raíz: (i) la etapa inicial, cuando el RAM está presente y activo; (ii) la etapa intermedia, cuando el meristemo está presente pero es más pequeño debido a que la zona de diferenciación se acerca al ápice de la raíz; y (iii) la etapa terminal, cuando todas las células de la raíz primaria, incluyendo a las que ocupan el lugar del RAM, están diferenciadas.

El transcriptoma se secuenció a partir de dos muestras biológicas por etapa. Se utilizó la plataforma Illumina HiSeq 2000 y para cada una de las 6 muestras se obtuvieron ~46 millones de lecturas pareadas de 100 nucleótidos (nt). Las lecturas se ensamblaron en 49,045 contigs mayores a 400 nt, de 1,081 nt de longitud en promedio. 63% de los contigs ensamblados se anotaron funcionalmente. Los contigs

ensamblados se utilizaron como referencia para estimar la abundancia de cada transcrito en las diferentes etapas de crecimiento de la raíz. Se identificó y extrajo a los contigs que se expresan diferencialmente entre distintas etapas, y se integraron a algunos de estos genes en una red de regulación transcripcional inferida a partir de lo descrito para la raíz primaria de la planta modelo *Arabidopsis thaliana*. Los valores de expresión virtual que se obtuvieron para el transcriptoma, así como algunos de los nodos de la red de regulación transcripcional, fueron validados experimentalmente mediante ensayos de RT-qPCR. El análisis del transcriptoma de la punta de la raíz de *P. pringlei* que se presenta en esta tesis, así como los resultados que obtuvimos al descargar y analizar las lecturas de RNAseq generadas para *Arabidopsis thaliana* y publicadas por otros grupos, corroboró que existen mecanismos moleculares conservados en el desarrollo de la raíz entre ambas especies. Además, a partir del análisis de expresión diferencial en *P. pringlei* y otras especies fue posible identificar algunos transcritos, aparentemente específicas de la familia Cactaceae, que pudieran estar involucradas en el desarrollo de la raíz de las cactáceas y en la regulación del agotamiento del RAM.

## SUMMARY

The root of most angiosperm species can grow for indefinite time as long as the Root Apical Meristem (RAM) is present and active, and therefore it is generally assumed that roots exhibit indeterminate growth. The species from the Cactaceae family, however, exhibit determinate growth of the primary root as the cells that constitute the RAM differentiate after a few cell division cycles, leading to the RAM exhaustion shortly after germination. As a consequence of RAM exhaustion and differentiation of all root apex cells, root growth ceases. Determinate growth of the primary root is a genetically stable program in Cactaceae and it has been suggested to represent an adaptive trait to arid environments, such as deserts. Although this growth pattern was reported for lateral and adventitious roots of some species, the determinate growth of the primary root has only been reported for Cactaceae species.

In this work *Pachycereus pringlei* was used to explore the molecular mechanisms underlying the determinate root growth. *P. pringlei* is a columnar cactus endemic to the Sonoran Desert. The RAM of the *P. pringlei* primary root becomes exhausted 7 to 10 days after germination. We report here the results of sequencing and the *de novo* assembly of the transcriptome of the *P. pringlei* root apex at three developmental stages: i) initial, when the RAM is present and active; ii) intermediate, when the RAM is still present but is smaller as the differentiation zone approaches the root tip; and iii) terminal, when the RAM is exhausted and all cells at the root tip, including those that were part of the RAM, are differentiated.

The *P. pringlei* root tip transcriptome was sequenced from two biological replicates per developmental stage by using the Illumina HiSeq 2000 sequencing platform. An average of 46 million paired-end reads (2x100 nt) were obtained per sample. After read processing the transcriptome was assembled into 49,045 contigs longer than 400 nt, with an average contig length of 1,081 nt. 63% of the assembled contigs were annotated. The assembled contigs were used as reference sequences to assess the transcript abundance by mapping the reads to the assembled contigs in each growth stage. The differentially expressed transcripts among growth stages were

identified and extracted. Some transcripts were integrated into an inferred transcriptional regulatory network based on published research articles for *Arabidopsis thaliana*. The expression level of some assembled transcripts were corroborated by RT-qPCR. In addition, the raw RNAseq reads generated by independent groups for *A. thaliana* were downloaded, processed, and mapped to the respective reference genome. The differential expression analysis of these transcriptomes confirmed the existence of conserved developmental programs of the primary root. We also identified some putative lineage specific transcripts that might be involved in the regulation of the RAM exhaustion, providing new opportunities for evo-devo research.



# I. INTRODUCCIÓN

## I.1 GENERALIDADES SOBRE LA RAÍZ

La raíz es el órgano de las plantas que absorbe agua y los nutrientes a partir del suelo y que fija a las plantas a su sustrato (Taiz and Zeiger, 2010). En la raíz se pueden distinguir longitudinalmente tres zonas: (i) la zona meristemática, (ii) la zona de elongación y (iii) la zona de diferenciación o de maduración (Fig. 1a, b). Una región adicional, la cofia, protege al ápice de la raíz al evitar el daño físico de este órgano mediante la secreción de mucopolisacáridos, los cuales facilitan la progresión de la raíz a través del suelo. Además, la cofia está involucrada en la percepción de señales de posición u orientación, tales como el gravitropismo y el hidrotropismo, entre otras (Kumpf and Nowack, 2015). Inmediatamente después de la cofia se encuentra la zona meristemática. Las células del meristemo apical de la raíz (RAM, por sus siglas en inglés; Fig. 1c) se dividen frecuentemente para mantener el desarrollo de la raíz. En general, el RAM se encuentra presente y activo en la raíz por largos períodos durante la vida de la planta, lo cual permite el crecimiento indeterminado de la raíz; es decir, el crecimiento de la raíz de manera indefinida. Las divisiones mitóticas de las células del RAM desplazan funcionalmente a las células generadas hasta que éstas salen de la zona meristemática y entran, eventualmente, en la zona de elongación. En la zona de elongación ocurre el alargamiento rápido de las células de la raíz. La división celular constante en la zona meristemática, en conjunto con la expansión longitudinal de las células en la zona de elongación son los dos mecanismos que permiten el crecimiento de la raíz, por lo cual, la zona meristemática y la zona de elongación constituyen a la zona de crecimiento. Finalmente, en la zona de diferenciación, la cual es la zona radical más alejada del ápice, las células totalmente elongadas adquieren características que las distinguen en tipos celulares específicos. La zona de diferenciación se distingue fácilmente debido a que algunas células epidérmicas forman extensiones llamadas pelos radicales, los cuales aumentan la superficie de contacto entre la raíz y el sustrato y maximizan el área disponible para la absorción de nutrientes minerales y agua. En la

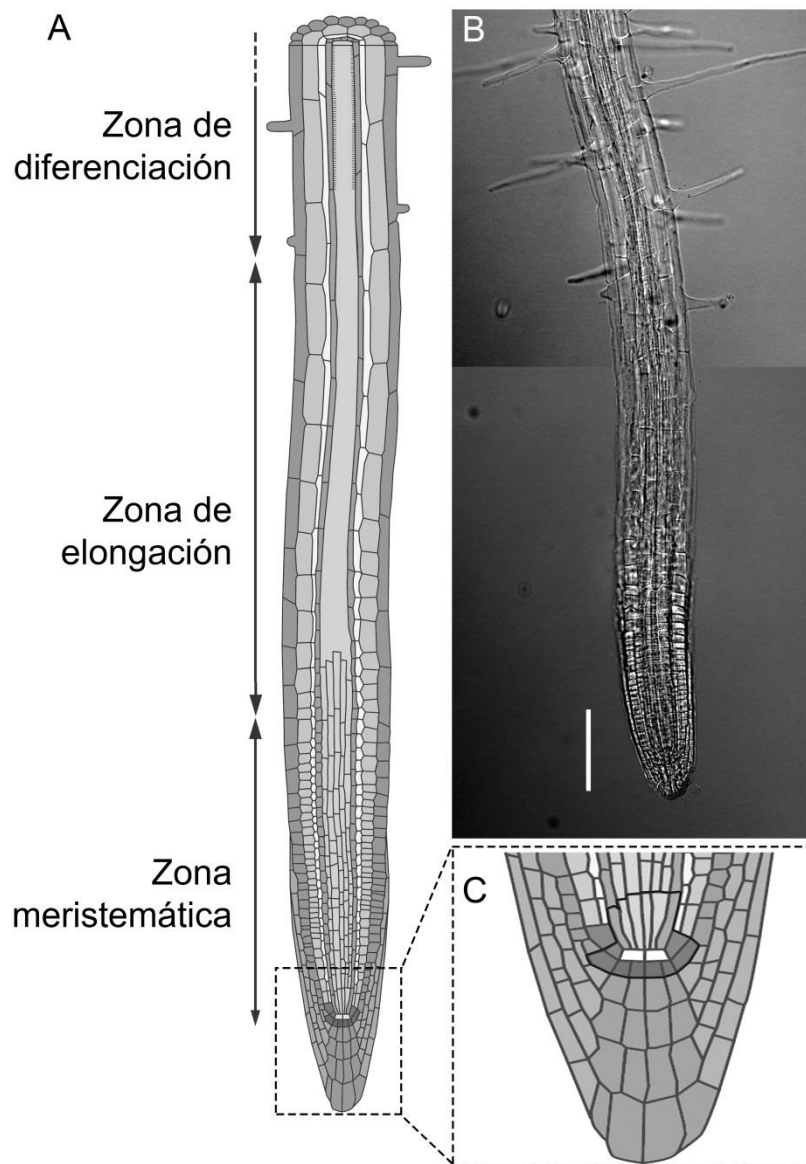
raíz se pueden distinguir tipos celulares ordenados concéntricamente. Estos son, del exterior hacia el interior, la epidermis, el córtex, la endodermis y el periciclo, el cual es la capa más externa del estele o cilindro vascular (Scheres *et al.*, 2002). En la raíz joven de *Arabidopsis thaliana*, la epidermis, el córtex, la endodermis y el periciclo están compuestos cada uno por una capa celular, lo cual ha facilitado los estudios anatómicos y ontogénicos de este órgano (Dolan *et al.*, 1993).

### **I.1.1 El Meristemo Apical de la Raíz**

Las células que componen al RAM son mitóticamente competentes y se mantienen en un estado relativamente indiferenciado (Evert y Eichhorn, 2006). En la década de 1950, Clowes (1958) identificó a un grupo de células en la parte distal del RAM que a diferencia de las demás células del meristemo presenta una actividad mitótica muy baja. A esta región la denominó centro quiescente (CQ). El término CQ acuñado por Clowes incluía a células con extremadamente baja o nula actividad mitótica, a las cuales se denomina actualmente CQ de forma estricta; y a células adyacentes al CQ, las cuales se denominan actualmente células iniciales y tienen actividad mitótica más baja que el resto de las células en el RAM pero se dividen más frecuentemente que el CQ (Fig. 1c). Se ha propuesto que el CQ actúa como centro organizador del RAM al mantener indiferenciadas a las células iniciales. Las células iniciales dan lugar a las diferentes filas de tejidos radicales mediante divisiones asimétricas, es decir, una célula hija mantiene su estado como célula inicial, con divisiones mitóticas poco frecuentes, mientras que la otra adquiere una tasa mitótica alta (Doerner, 1998; Dolan *et al.*, 1993).

### **I.1.2 El RAM mantiene el crecimiento de la raíz**

Los experimentos de remoción física del ápice de la raíz de *Z. mays*, incluyendo a las células del CQ, mostraron que las células del estele pueden desdiferenciarse y formar un nuevo CQ (Feldman, 1976). Esta observación se corroboró posteriormente mediante la ablación por láser de las células del CQ de la raíz de *A. thaliana* (Berg *et al.*, 1995), en donde se demostró que el contacto entre las células del CQ y las células iniciales del RAM mantiene a estas últimas en un estado indiferenciado (Berg *et al.*, 1995).



**Figura 1.** Organización de la raíz primaria (A) y de la raíz lateral (B) de *Arabidopsis thaliana*. (A) Diagrama de un corte longitudinal de la raíz primaria de *A. thaliana* en donde se señala a las zonas meristemática, de elongación y de diferenciación. En las raíces laterales (B) también se distinguen a estas zonas. (C) El meristemo apical de la raíz (RAM) está contenido en la zona meristemática. El RAM contiene a las células del centro quiescente, las cuales tienen muy baja o nula actividad mitótica y se muestran con color blanco en el diagrama. Las células adyacentes al centro quiescente, delimitadas con un borde grueso, se conocen como células iniciales y son las progenitoras de las células con alta tasa mitótica que forman al RAM. Barra en B = 100  $\mu$ m. A y C, tomadas y modificadas de Peret, B. (2017): doi:10.6084/m9.figshare.5143987.v4, disponibles en FigShare como contenido open access (licencia CC BY 4.0).

En las plantas mutantes de *A. thaliana* que tienen CQ con identidad comprometida durante el crecimiento de la raíz, las células iniciales se dividen anormalmente y las células que previamente formaban al CQ se activan mitóticamente (Aida *et al.*, 2004; Benfey *et al.*, 1993; Di Laurenzio *et al.*, 1996). El agotamiento del meristemo mediante la diferenciación de todas las células del ápice abate el crecimiento de la raíz; es decir, cuando la identidad del CQ se compromete, las raíces exhiben un crecimiento del tipo determinado.

## **I.2 CRECIMIENTO DETERMINADO DE LA RAÍZ**

Mientras el RAM se encuentra presente en la raíz y es potencialmente funcional la raíz puede continuar su crecimiento de manera indefinida, por lo cual se considera que el patrón de crecimiento de la raíz es indeterminado. Sin embargo, existen ciertas situaciones en las cuales el RAM se agota, es decir, las células meristemáticas se diferencian; por ejemplo ante la falta de algunos nutrimentos tales como el fósforo (Sánchez-Calderón *et al.*, 2005), o en algunas especies cuando la raíz encuentra un obstáculo físico (Watt *et al.*, 2006). Cuando el RAM se consume, es decir, cuando se agota, se dice que la raíz presenta crecimiento determinado. En las plantas cactáceas el agotamiento del RAM se observó por primera vez en las raíces laterales de las especies *Opuntia arenaria* y *O. tunicata* (Boke, 1989). Posteriormente, Dubrovsky reportó que las cactáceas *Stenocereus gummosus* (Dubrovsky, 1997) y *Pachycereus pringlei* (Dubrovsky and Gómez-Lomelí, 2003), entre otras, presentan crecimiento determinado de la raíz primaria. El crecimiento determinado de distintos tipos de raíces se ha observado en varias especies, por ejemplo, en las raíces proteoides de especies de la familia Proteaceae (Dinkelaker *et al.*, 1995), en las raíces adventicias de *Ficus pumila* (Groot *et al.*, 2003) y en las raíces laterales de *O. arenaria*, *O. tunicata* y *Zea mays* (Boke, 1989; Varney y McCully, 1991); sin embargo, el crecimiento determinado como programa de desarrollo de la raíz primaria sólo se ha observado en especies de la familia Cactaceae (Shishkova *et al.*, 2013).

En algunas cactáceas tales como *S. gummosus*, *S. thurberi* y *Ferocactus peninsulæ* se demostró que el agotamiento del meristemo precede a la emergencia de las raíces laterales (Dubrovsky, 1997), por lo cual se ha propuesto que el agotamiento

del RAM, así como el desarrollo temprano de las raíces laterales, representa una ventaja adaptativa para el establecimiento de las plántulas de estas especies pues favorece la absorción del agua y los minerales en los ambientes desérticos, en donde la precipitación pluvial es escasa (Dubrovsky, 1997; Shishkova *et al.*, 2008). Además, se ha demostrado que el agotamiento del RAM de la raíz primaria de *S. gummosus* y *P. pringlei* no ocurre mediante muerte celular programada (Shishkova and Dubrovsky, 2005), aunque la evidencia sugiere que este proceso sí ocurre en las células de la cofia y en los pelos radicales de las cactáceas. Las observaciones realizadas en las raíces que se regeneraron a partir de callos, indican que el crecimiento determinado corresponde a un programa de desarrollo conservado en *S. gummosus* y *F. peninsulae* y en diferentes tipos de raíces (Shishkova *et al.*, 2007). Adicionalmente, se demostró que en las cactáceas *S. gummosus* y *P. pringlei* el CQ no se establece o se establece de manera transitoria durante el desarrollo postembrionario, respectivamente, aunque en ambos casos el crecimiento de la raíz primaria es del tipo determinado; es decir, el programa genético que conduce al agotamiento del RAM es independiente del establecimiento del CQ, y cuando éste se establece pierde su identidad rápidamente, de modo que en la etapa terminal del desarrollo no se distinguen células de CQ (Rodríguez-Rodríguez *et al.*, 2003).

### **I.3 REGULACIÓN GENÉTICA DE LA ACTIVIDAD DEL RAM**

La actividad de las células meristemáticas está estrictamente regulada, siendo la regulación genética un componente importante en el mantenimiento y actividad del meristemo. Gran parte del conocimiento que se tiene acerca de los mecanismos moleculares que regulan el mantenimiento del RAM proviene de los experimentos que se han realizado en la planta modelo *A. thaliana*. Los ensayos clásicos de mutagénesis y escrutinio de mutantes, y la posterior caracterización de los genes responsables del fenotipo de estas mutantes, han permitido identificar a algunos genes importantes en el desarrollo de la raíz. En *A. thaliana* se han identificado tres vías de regulación preponderantes en el mantenimiento del meristemo: la vía mediada por los factores de transcripción PLETHORA; la vía mediada por cofactores de transcripción de la familia GRAS; y la vía mediada por WOX5-CLE40.

### I.3.1 Vía PLT

La vía más importante de control y especificación del RAM que se ha descrito hasta ahora es aquella mediada por los factores de transcripción PLETHORA (PLT), los cuales son parcialmente redundantes. Las proteínas PLT pertenecen a la superfamilia AP2/EREBP (Dietz *et al.*, 2010), la cual es específica de plantas y es también una de las familias proteicas más grandes descritas para estos organismos (Okamoto *et al.*, 1997). Las proteínas codificadas por los genes de la superfamilia AP2/EREBP se distinguen por la presencia de uno o dos dominios AP2, el cual consiste en una secuencia conservada de aproximadamente 60 aminoácidos (Jofuku *et al.*, 1994). En el brote, la proteína AP2 está involucrada en el establecimiento de la identidad de los meristemas florales (Krizek, 2011), mientras que algunas proteínas PLT participan en el control de la filotaxis (Prasad *et al.*, 2011). Aunque las funciones de los factores de transcripción AP2/EREBP son diversas, una gran parte de ellas están relacionadas con la regulación de procesos del desarrollo (Riechmann and Meyerowitz, 1998). En la raíz, los genes *PLT* son importantes para el establecimiento de la identidad de las células meristemáticas, incluyendo la especificación de la identidad de las células del CQ e iniciales (Aida *et al.*, 2004). Los genes *PLT1* y *PLT2* se expresan en el ápice de la raíz y son altamente redundantes, por lo cual las mutantes simples de pérdida de función de estos genes no presentan un fenotipo marcadamente distinto al de las plantas del tipo silvestre; sin embargo, las mutantes dobles *plt1 plt2* conducen a un sistema radical muy corto y ramificado compuesto por raíces con crecimiento determinado (Aida *et al.*, 2004). Las mutantes triples *plt1 plt2 plt3* no desarrollan raíz postembrionaria, y la expresión ectópica de los genes *PLT1* o *PLT2* conduce a la organogénesis de raíces en el brote, e incluso a la inhibición de la formación del SAM mediante el establecimiento de un RAM, lo cual resulta en el desarrollo de una plántula con raíces en los dos ápices y sin meristemo del brote (Aida *et al.*, 2004; Galinha *et al.*, 2007). Los genes *PLT* se expresan en forma de gradiente como respuesta al establecimiento de máximos locales de auxina. La expresión de los genes *PLT1*, *PLT2*, *PLT3* y *PLT4* (*BBM*), tanto a nivel de transcrito como a nivel de proteína, es máxima en el RAM, y disminuye gradualmente hasta que en la zona de diferenciación la expresión de estos genes se abate casi completamente (Galinha *et al.*, 2007; Mähönen *et al.*, 2014). En

nuestro grupo se han estudiado los niveles de expresión transcripcional de los genes *PLT* en el ápice de la raíz primaria de la cactácea *P. pringlei* en diferentes etapas del crecimiento (López-Valle, 2015). Los análisis preliminares de transcriptómica y de RT-qPCR a partir de muestras del ápice de la raíz primaria de *P. pringlei* sugieren que los genes *PpPLT* disminuyen su expresión, aunque sin abatirla completamente, cuando las células del ápice de la raíz están diferenciadas.

### I.3.1 Vía SHR y SCR

Las proteínas SHORTROOT (SHR) y SCARECROW (SCR) pertenecen a la familia GRAS, la cual es específica de plantas y se denomina así debido al nombre de los primeros genes de esta familia que se descubrieron: *GIBBERELLIC ACID INSENSITIVE* (**GAI**), *REPRESOR OF GA1* (**RGA**), y *SCARECROW* (**SCR**) (Bolle, 2004). Diversas proteínas que pertenecen a esta familia participan en diferentes procesos del desarrollo de las plantas. SCR fue el primer miembro de la familia GRAS en ser aislado (Di Laurenzio *et al.*, 1996). Las mutantes de pérdida de función *scr* y *shr* de *A. thaliana* tienen la raíz primaria más corta (Benfey *et al.*, 1993; Sabatini *et al.*, 2003). En estas mutantes la identidad del CQ está comprometida, por lo cual la raíz crece durante un período breve y posteriormente el meristemo se agota (Benfey *et al.*, 1993; Sabatini *et al.*, 2003). En la raíz de ambas mutantes se desarrolla una sola capa en lugar de las capas que corresponderían a la epidermis y al córtex. Esta observación indica que las proteínas SCR y SHR se requieren para la división asimétrica que da lugar a la endodermis y al córtex (Di Laurenzio *et al.*, 1996; Helariutta *et al.*, 2000). Las células de la capa celular que se desarrolla en la mutante *scr* exhibe características tanto de la endodermis como del córtex (Di Laurenzio *et al.*, 1996) mientras que las células de ésta capa en la mutante *shr* presentan sólo marcadores del córtex (Helariutta *et al.*, 2000). Estudios posteriores mostraron que aunque la proteína SHR se sintetiza en las células del estele y se mueve hacia la endodermis, en donde induce la transcripción de *SCR* (Nakajima *et al.*, 2001). El movimiento intercelular de SHR requiere de su acumulación en el citoplasma, sin embargo, no se trata de difusión pasiva, sino que ocurre mediante la interacción con otras proteínas (Gallagher and

Benfey, 2009; Long *et al.*, 2015; Welch *et al.*, 2007a). En el núcleo, SHR actúa como cofactor transcripcional e induce la expresión de SCR, lo cual a su vez refuerza la traslocación de SHR hacia el núcleo, estableciendo un ciclo de retroalimentación positiva que secuestra a SHR e impide su difusión más allá de la endodermis (Cui *et al.*, 2007). Aunque inicialmente se consideró a SHR y SCR como factores transcripcionales, no existía evidencia sólida de la unión directa de estas proteínas al DNA y tampoco se conocían las secuencias cortas del DNA, o cajas, que funcionaban como sitios de unión para SHR-SCR. Recientemente, los estudios realizados a partir de la cristalografía de complejos multiproteicos y de interacción proteína-proteína *in vivo*, demostraron que el dímero SCR-SHR interacciona con los dedos de zinc de JACKDAW (JKD) y que es esta proteína la que interacciona de manera directa con el DNA (Hirano *et al.*, 2017). JKD puede además interactuar con los monómeros de SCR o SHR, de modo que la actividad de JKD está regulada de manera dependiente del contexto celular (Long *et al.*, 2017). Así, el dímero SHR-JKD es importante en el mantenimiento del CQ mientras que el dímero SCR-JKD es importante en la regulación de la identidad de las células de la endodermis (Long *et al.*, 2017). Un interactor adicional en la vía de regulación SCR-SHR es la proteína RETINOBLASTOMA-RELATED (RBR), la cual regula de manera negativa a SCR (Cruz-Ramírez *et al.*, 2012). Se ha demostrado que RBR participa en la especificación del tamaño del meristemo, y en consecuencia de la longitud de la raíz, al integrar las vías de regulación de las fitohormonas auxina y giberelina (Cruz-Ramírez *et al.*, 2013; Perilli *et al.*, 2013).

### **I.3.3 Vía WOX5-CLE40**

La proteína WUSCHEL (WUS) se sintetiza en el centro organizador del SAM, equivalente al CQ del RAM, y mantiene a las células iniciales en un estado indiferenciado (Mayer *et al.*, 1998). Debido a su papel preponderante en el mantenimiento del SAM, se estudió la posible participación de proteínas homólogas de WUS en el mantenimiento del RAM de *Oryza sativa* y *A. thaliana* (Kamiya *et al.*, 2003; Sarkar *et al.*, 2007). De este modo, se identificó a un homólogo de WUS que se expresa específicamente en las células del CQ del arroz (Kamiya *et al.*, 2003), con lo cual se sugirió que un mecanismo análogo al del mantenimiento del SAM podría operar



en la raíz. Por su parte, se demostró que en *Arabidopsis* el patrón de expresión de *WUSCHEL RELATED HOMEBOX 5 (WOX5)* en el RAM es comparable al de *WUS* en el SAM y que su actividad es necesaria para mantener la identidad del CQ y el estado indiferenciado de las células iniciales (Sarkar *et al.*, 2007). En el brote, el dominio de expresión de *WUS* es delimitado mediante una vía de retroalimentación negativa que involucra a *CLAVATA1 (CLV1)*, una proteína-cinasa de tipo receptor; al coreceptor *CLAVATA2 (CLV2)*, y al ligando peptídico *CLAVATA3 (CLV3)* (Jeong *et al.*, 1999; Stone *et al.*, 1998; Trotochaud *et al.*, 2000). Dadas las similitudes en secuencia, perfiles de expresión y aparente actividad entre *WUS* y *WOX5*, se buscaron a los homólogos de las proteínas *CLV* en la raíz. El péptido *CLE40 (CLAVATA3/ESR-RELATED 40)*, homólogo de *CLV3*, se expresa de manera específica en la raíz, en donde regula el dominio de expresión de *WOX5* (Stahl *et al.*, 2009). En esta vía de señalización también interviene el receptor *CRINKLY4 (ACR4)* una proteína cinasa de tipo receptor que se expresa en la columella. *ACR4* se ha propuesto como un interactor de *CLE40* (Stahl and Simon, 2009). *CLE40* tiene un papel dual al regular positivamente la expresión de *ACR4* y funcionar como ligando de este receptor; de esta forma se inicia una cascada de señalización que restringe el dominio de expresión de *WOX5*. El dominio de expresión de *WOX5* es también delimitado por la proteína *REPRESSOR OF WUSCHEL1 (ROW1)*, una proteína que reprime la expresión de *WOX5* al unirse a la histona 3 trimetilada en la lisina 4 (H3K4me3) en la región promotora de *WOX5*.

## II. JUSTIFICACIÓN

La prevalencia del crecimiento determinado de la raíz primaria de las cactáceas (Dubrovsky, 1997; Dubrovsky and Gómez-Lomelí, 2003; Shishkova *et al.*, 2008), así como de las raíces laterales y de las raíces regeneradas a partir de callo (Shishkova *et al.*, 2007), en adición a su ocurrencia de manera independiente de las condiciones de cultivo (Dubrovsky, 1997), indica que el crecimiento determinado de la raíz primaria mediante el agotamiento del RAM es un programa genético robusto. La raíz primaria de la mayoría de especies de las cactáceas de la subfamilia Cactoideae (Shishkova *et al.*, 2013), entre ellas *P. pringlei* (Dubrovsky and Gómez-Lomelí, 2003), presenta

crecimiento determinado de manera natural. Sin embargo, a pesar de los estudios que se han realizado acerca de la ocurrencia del crecimiento determinado en la familia Cactaceae, no se conoce la regulación genética que subyace al agotamiento del RAM, aun cuando la técnica de hibridación sustractiva en *Stenocereus gummosus* permitió identificar a algunos genes que se expresan diferencialmente en el ápice de la raíz primaria entre las plántulas con RAM activo y RAM agotado (Shishkova *et al.*, 2013).

*P. pringlei* es una cactácea endémica del Desierto de Sonora cuya raíz primaria detiene su crecimiento a los 7-10 días después de la germinación (d.d.g.) como resultado del agotamiento del RAM, por lo cual esta especie resulta un buen modelo para el estudio del agotamiento del meristemo pues permite la colecta de las muestras en una ventana temporal más amplia que el resto de las especies que se han estudiado (Shishkova *et al.*, 2013). El estudio de la regulación genética del agotamiento del RAM, al igual que otros procesos biológicos en los cactus, estuvo limitado debido a que no existía un genoma de referencia para el estudio de los organismos de la familia Cactaceae. Aunado a esto, el número de secuencias nucleotídicas provenientes de cactáceas y disponibles públicamente era escaso. Esta situación mejoró a partir del segundo semestre de 2014, cuando la implementación de esfuerzos colaborativos a través del consorcio internacional OneKP (Matasci *et al.*, 2014) resultó en la generación de una vasta cantidad de secuencias mediante RNAseq a partir de órganos aéreos de tres especies de cactáceas (Yang *et al.*, 2015), sin embargo, el uso de estos datos no es totalmente libre y está sujeto a la aprobación del consorcio para evitar conflictos de interés (Matasci *et al.*, 2014).

Recientemente, la secuenciación masiva de cDNA sintetizado a partir de mRNA (RNAseq) se ha adoptado como una herramienta poderosa que permite estimar la abundancia de los transcritos en una condición, tejido, o estadio de desarrollo específico (Mortazavi *et al.*, 2008). Además, cuando se comparan los datos obtenidos mediante RNAseq de dos o más muestras que difieren en alguna condición, es posible analizar los cambios en la expresión de los genes que están involucrados en la respuesta a tal circunstancia (Oshlack *et al.*, 2010). El ensamblaje del transcriptoma a partir de los datos obtenidos mediante la secuenciación del RNA puede realizarse *de novo*, es decir, no requiere del conocimiento previo del genoma o del transcriptoma de

la especie que se estudia (Grabherr *et al.*, 2011). En el caso de las plantas, la secuenciación del transcriptoma mediante RNAseq se ha utilizado con fines diversos, tales como el estudio de la presión de selección durante la evolución de las especies (Koenig *et al.*, 2013), la identificación de genes nuevos y eventos de “splicing” alternativo (Iorizzo *et al.*, 2011) y el estudio del desarrollo de órganos específicos (Zenoni *et al.*, 2010). De este modo, es posible explorar los mecanismos de regulación genética del agotamiento del RAM de la raíz primaria de las cactáceas a partir de la secuenciación y del ensamblaje *de novo* del transcriptoma del ápice de la raíz en diferentes etapas del desarrollo; al mismo tiempo, esta información permitirá explorar la conservación de los mecanismos regulatorios del desarrollo de la raíz de las plantas angiospermas.

### III. OBJETIVOS

#### Objetivo general

Explorar la regulación genética del agotamiento del meristemo apical de la raíz de las cactáceas al utilizar como modelo de estudio a *Pachycereus pringlei*.

#### Objetivos particulares

- a) Ensamblar *de novo* el transcriptoma del ápice de la raíz primaria de *P. pringlei* en las etapas Inicial, Intermedia y Terminal del desarrollo.
- b) Anotar funcionalmente a los transcritos ensamblados.
- c) Cuantificar los niveles de abundancia, por etapa, de los transcritos ensamblados e identificar a aquellos que se expresan de manera diferencial.
- d) Integrar a algunos de los transcritos ensamblados en una red de regulación transcripcional.
- e) Explorar la existencia de reguladores transcripcionales no descritos de los programas genéticos del agotamiento o mantenimiento del meristemo apical de la raíz.

- f) Validar abundancia de transcritos selectos al cuantificar su expresión relativa mediante RT-qPCR.
- g) Realizar análisis comparativos de transcriptómica entre los datos generados para *P. pringlei* y los datos de secuenciación y de microarreglos que se han obtenido para otras especies vegetales.

## IV. MATERIALES Y METODOS

### IV.1 OBTENCIÓN DE MUESTRAS BIOLÓGICAS

Las semillas de *P. pringlei* que se utilizaron en este trabajo fueron donadas por el Dr. Francisco Molina-Fraener y el Q. B. José F. Martínez Rodríguez (Instituto de Ecología, Unidad Hermosillo; UNAM). Estas semillas fueron colectadas en diferentes localidades del Desierto de Sonora durante el julio de 2014 alrededor de Bahía de Kino (28.831476,-111.803929). La pulpa y las semillas de los frutos y se esparció sobre hojas de papel para ser enviadas a Cuernavaca durante el mismo mes. Al ser recibidas, la pulpa se dejó secar para evitar la contaminación por hongos; posteriormente las semilla se limpiaron al eliminar la pulpa y se almacenaron a temperatura ambiente hasta su uso. Las semillas se esterilizaron por 20 minutos en la solución SuperKill (cloro comercial 60%, Triton-X 0.08%), se lavaron 4 veces por 10 minutos en agua destilada esteril y se germinaron en cajas de Petri con medio LS 0.2x, 0.8% Bacto Agar, sin sacarosa. Las plantas se cultivaron en cajas de Petri en posición vertical en una cámara de crecimiento a 28°C con un fotoperíodo de 12 h luz / 12 h oscuridad e intensidad luminosa de 80  $\mu\text{mol m}^{-2}\text{s}^{-1}$ . Se colectaron muestras de 1 mm del ápice de la raíz primaria de *P. pringlei* en tres etapas de desarrollo: de 1-4 d.d.g., en la etapa inicial (IN), cuando el RAM está presente y activo; de 5-6 d.d.g, en la etapa intermedia (IT), cuando el RAM sigue presente pero es más pequeño, puesto que la zona de diferenciación se aproxima al ápice de la raíz primaria; y de 9-13 d.d.g., en la etapa terminal (TR), cuando todas las células del ápice de la raíz están completamente diferenciadas. Se colectaron aproximadamente 350, 450 y 600 ápices para las muestras de las etapas IN, IT y TR, respectivamente; cada una de las cuales se colectó por duplicado para RNAseq y por triplicado para qRT-PCR. Las muestras

se congelaron inmediatamente en nitrógeno líquido y se mantuvieron a -70°C hasta la extracción del RNA. Hacia la conclusión de este trabajo se colectaron frutos en diferentes sitios alrededor de Bahía de Kino (28.831476,-111.803929) y Guaymas (27.989451,-110.838980) durante junio del 2017.

## **IV.2 RNAseq Y ENSAMBLAJE DEL TRANSCRIPTOMA DE *Pachycereus pringlei***

El RNA total se extrajo mediante el método de TRIzol (Invitrogen) de acuerdo con el protocolo del fabricante a partir del ápice de la raíz primaria de *P. pringlei*. La preparación de bibliotecas y la secuenciación en la plataforma Illumina Hi-Seq 2000 se realizó en el Beijing Genomics Institute (BGI Tech, Hong Kong) en el formato de lecturas apareadas (paired-end) de 100 nucleótidos de longitud (2x100). Se analizó la calidad de las lecturas mediante el software FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) y el paquete CLC Genomics Workbench (CLC, a Qiagen Company). Las lecturas se procesaron de la siguiente manera: los primeros 13 nucleótidos del extremo 5' se removieron; cuando las dos lecturas provenientes del mismo fragmento sobrelapaban en sus extremos, estas lecturas se fusionaron. Las regiones de baja calidad y los adaptadores de secuenciación también fueron removidos. Las lecturas que después de los pasos descritos tuvieron una longitud menor a 40 nt se descartaron. El total de las lecturas fusionadas y apareadas se utilizó para realizar el ensamblaje del transcriptoma. El transcriptoma de la punta de la raíz primaria de *P. pringlei* se ensambló con el software CLC Genomics Workbench al utilizar la herramienta de ensamblaje *de novo*, la cual reconstruye contigs a partir de la generación de gráficas de de Bruijn. Se realizaron diferentes ensamblajes al variar los parámetros de tamaño de palabra ("word size") y de burbuja ("bubble size") y a partir de ellos se escogió al ensamblaje con mejores valores de N50 y menor redundancia en el mapeo de lecturas. Las lecturas no mapeadas durante el ensamblaje del transcriptoma se sometieron a una nueva ronda de ensamblaje bajo los mismos parámetros de tamaño de palabra y de burbuja. Los contigs ensamblados en ambas rondas se almacenaron como un solo set de datos y

aquellos contigs cuya longitud fue >3 kb se extendieron al utilizar el módulo “Genome Finishing” de CLC Genomics Workbench.

### IV.3 ANOTACIÓN DEL TRANSCRIPTOMA

Los contigs que se ensamblaron, en lo subsecuente referidos como transcritos, se tradujeron a las secuencias de aminoácidos en los seis marcos de lectura posibles. Las secuencias de aminoácidos predichas se utilizaron como sonda para identificar mediante tBLASTx al homólogo más cercano en la base de datos RefSeq de plantas dicotiledóneas, la cual está contenida en el servidor del NCBI (<http://www.ncbi.nlm.nih.gov/refseq/>). El mejor hit de la búsqueda mediante tBlastx se utilizó para anotar las ontologías génicas (función molecular, localización subcelular, proceso biológico) mediante Blast2GO (Conesa *et al.*, 2005). Se utilizaron además las bases de datos InterPro Scan (Quevillon *et al.*, 2005) y el servidor KAAS (Kegg Automated Annotation Service, Moriya *et al.*, 2007) de la Kyoto Encyclopedia of Genes and Genomes.

### IV.4 ANÁLISIS DE EXPRESIÓN DIFERENCIAL

Se realizó la cuantificación de la expresión virtual (abundancia) de los transcritos mediante el conteo de lecturas mapeadas y su normalización al utilizar el método de RPKMs (Mortazavi *et al.*, 2008). Los valores crudos se utilizaron en una prueba de detección empírica de expresión diferencial (EDGE test; Robinson *et al.*, 2010), en la cual consideramos como expresados diferencialmente aquellos transcritos cuya abundancia cambia al menos dos veces entre las diferentes etapas y cuyo  $p$ -value fue menor a 0.005 después de la corrección de falsos positivos (“FDR corrected  $p$ -value”). Se generó una lista de los transcritos que se expresan diferencialmente. Estos transcritos fueron agrupados mediante el método de  $k$ -mean clustering (Lloyd, 1982). El número de particiones ( $k$ ) que se utilizó para el análisis se definió a partir de la regla empírica  $k=\sqrt{[n/2]}$ , con  $n$  = número de transcritos expresados diferencialmente. Los conglomerados (o clusters) que se generaron se utilizaron para identificar a las

categorías de proceso biológico enriquecidas en cada conglomerado mediante un análisis hipergeométrico (Falcon and Gentleman, 2007).

#### **IV.5 VALIDACIÓN DEL RNAseq MEDIANTE RT-qPCR**

El RNA total se extrajo con TRIzol, se trató con DNAsa y se limpió con el kit RNeasy Micro (Qiagen). Se sintetizó la primera cadena de DNA complementario al utilizar a la retrotranscriptasa Superscript II y oligos dT de acuerdo con las instrucciones del fabricante (Thermo Fischer). Los ensayos de RT-qPCR se realizaron en un termociclador LightCycler Nano (Roche, Basel, Suiza). Se utilizó SybrGreen (Thermo Fischer Scientific) como sonda fluorescente en reacciones de amplificación con 50 ng de cDNA como templado. Los valores de Cq se utilizaron para calcular la abundancia relativa de cada transcrito al normalizar con múltiples genes de referencia. Los genes de referencia, o normalizadores, que se utilizaron en este trabajo se seleccionaron con base en las secuencias de los genes de referencia propuestos para *A. thaliana* (Czechowski, 2005). La extracción del RNA para los ensayos de RT-qPCR se realizó con el apoyo técnico de la M. en C. Selene Napsucialy Mendivil, y los ensayos de RT-qPCR con el apoyo técnico del Dr. Pedro E. Lázaro Mixteco y la M. en C. Mayra L. López-Valle.

#### **IV.6 TRANSCRIPTÓMICA COMPARATIVA Y REDES DE REGULACIÓN TRANSCRIPCIONAL**

Los valores de abundancia que se obtuvieron a partir del transcriptoma de la punta de la raíz de *P. pringlei* se compararon con los valores de expresión de sus probables ortólogos en otras especies vegetales, específicamente con *A. thaliana*, *Cucumis sativus*, *Glycine max*, *Oryza sativa*, *Solanum lycopersicum* y *Zea mays*. Se utilizaron los valores de expresión que se obtuvieron a partir de la secuenciación del transcriptoma de las tres zonas longitudinales de la raíz (Huang and Schiefelbein, 2015), así como de microarreglos de *A. thaliana* a los cuales se accedió a través del portal eFP Browser de la Universidad de Toronto (Winter *et al.*, 2007). Adicionalmente se utilizaron las secuencias de los unigenes ensamblados para el transcriptoma de la raíz y el brote de *Lophophora williamsii* (Ibarra-Laclette *et al.*, 2015). Se generó una

lista de genes mediante “literature mining” de genes reportados como reguladores importantes del desarrollo de la raíz de *A. thaliana*. Estos genes y sus reguladores, así como sus primeros vecinos se descargaron, junto con sus interacciones, del mapa de regulación de *A. thaliana* (Jin *et al.*, 2015). Los ortólogos putativos en *P. pringlei* de estos genes fueron identificados mediante la técnica de mejor hit bidireccional (BBH) entre *P. pringlei* y *A. thaliana* para cada secuencia, o como hit unidireccional cuando la estrategia de BBH no fue exitosa. Los contigs se consideraron ortólogos putativos cuando la cobertura del BLAST fue mayor al 40%, el e-value fue menor a  $1 \times 10^{-10}$  y cuando los dominios de la secuencia que se utilizó como sonda estuvieron presentes en la secuencia que se obtuvo como resultado. La presencia de los dominios se verificó con la base de datos SuperFamily (Wilson *et al.*, 2009). A partir de esta información se infirió una red de regulación transcripcional para *P. pringlei*. Se preparó una tabla de interacciones y una tabla de propiedades, las cuales fueron cargadas y visualizadas en el programa Cytoscape v3.2.1.(Shannon *et al.*, 2003).



## V. RESULTADOS Y DISCUSIÓN

### V.1 OBTENCIÓN DE MATERIAL BIOLÓGICO

En este trabajo se utilizaron semillas de 13 frutos de *P. pringlei* (Tabla I), las cuales provienen de diferentes individuos que crecen en la región de Bahía de Kino, en el Desierto de Sonora. Las semillas de cada fruto se almacenaron en sobres independientes y se estimó el número aproximado de semillas al considerar que 100 semillas tienen un peso de  $0.55 \pm 0.06$  g.

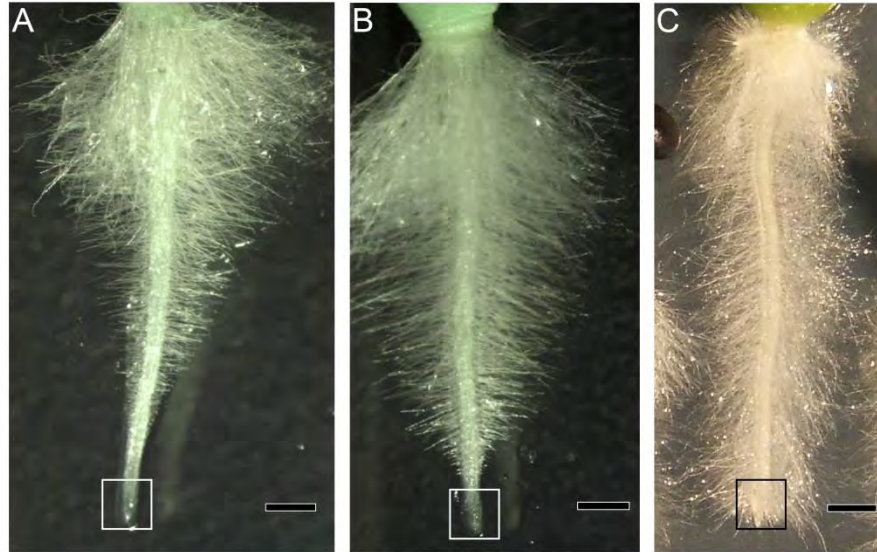
Tabla I. Lote (2014) de semillas de *Pachycereus pringlei*.

Fruto	Color de pulpa	Peso (g)	No. aprox. de semillas
1	Blanca	27.19	4,900
2	Roja	4.28*	780*
3	Roja	21.62	3,900
4	Morada	13.28	2,400
5	Blanca	5.62	1,000
6	Blanca	7.74	1,400
7	Roja	10.4	1,800
8	Roja	24.97	4,500
9	Roja	9.77	1,700
10	Morada	25.5	4,600
11	Roja	20.16	3,600
12	Morada	24.29	4,400
13	Roja	17.63	3,200

\* 50% de las semillas que se recibieron de este fruto se utilizaron antes de realizar la estimación del número de semillas.

Para coleccionar los ápices y asignarlos a la muestra correcta, cada raíz se observó con un estereoscopio para corroborar mediante inspección visual la etapa de crecimiento en la cual se encontraban al momento de la colecta (Fig. 2). Los ápices de la raíz primaria en la etapa IN del desarrollo se coleccionaron entre los días 1-4 d.d.g, en esta etapa la zona de diferenciación, identificada debido a la presencia de los pelos radicales, está alejada del ápice. El ápice de las raíces en la etapa IT se coleccionó entre los 5-6 d.d.g.; esta etapa se distinguió debido a que los pelos radicales se acercan más a la punta de la raíz, aunque no la cubren totalmente. Finalmente, la etapa TR se coleccionó entre los 9-13 d.d.g., y se distinguió porque los pelos radicales cubren a todo el

ápice de la raíz primaria. Los ápices de la raíz primaria se colectaron de manera independiente para dos réplicas biológicas de las etapas IN, IT y TR, en las cuales se incluyeron 350, 450 y 600 ápices por réplica, respectivamente.



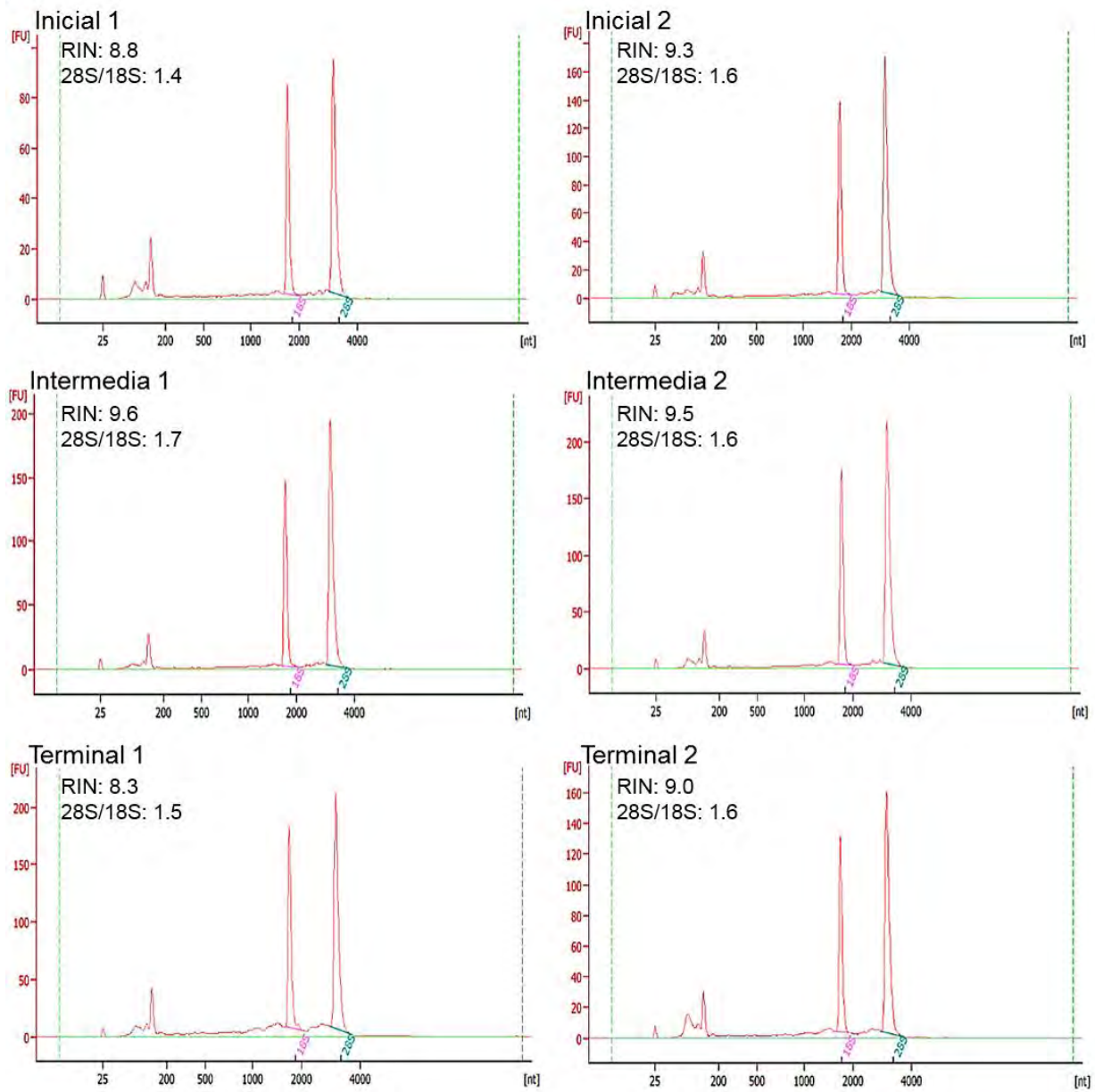
**Figura 2.** Etapas del desarrollo de la raíz primaria de *Pachycereus pringlei*. A) Etapa inicial, el meristemo apical de la raíz (RAM) está presente y activo; B) etapa intermedia, el RAM está presente pero es más pequeño y la zona de diferenciación, visualizada por la formación de pelos radicales, está más próxima al ápice aunque sin cubrirlo totalmente; C) etapa terminal, todas las células del ápice de la raíz, incluyendo a aquellas que formaban al RAM, están diferenciadas. Barra = 1 mm.

El RNA que se extrajo a partir de cada muestra fue caracterizado espectrofotométricamente y se envió al BGI – Hong Kong para la construcción y secuenciación de bibliotecas de RNAseq. Las muestras de RNA tuvieron baja contaminación con proteínas, como se aprecia a partir de los valores cercanos a dos al calcular la relación entre las absorbencias a 260 y 280 nm (Tabla II). Por otro lado, aunque la relación  $A_{260}/A_{230}$  parece bajo, lo cual indicaría contaminación con polisacáridos o polifenoles, este valor es aceptable de acuerdo con lo reportado como mínimo deseable para la preparación de bibliotecas de secuenciación (Johnson *et al.*, 2012), sobre todo al considerar que la contaminación por polisacáridos o polifenoles es un problema común al trabajar con plantas que llevan a cabo el metabolismo ácido de las crasuláceas (Gehrig *et al.*, 2000). Para inferir la integridad del RNA se utilizaron dos parámetros, la relación entre la concentración de los rRNA 28S/18S y el Número de

Integridad del RNA (RIN, por sus siglas en inglés). Aunque el valor de rRNA 28S/18S que se obtuvo para cada muestra fue mayor a 1.5, no es el valor óptimo, el cual tradicionalmente se espera igual o muy cercano a dos. Sin embargo, las revisiones sistemáticas de correlación de este parámetro con la integridad del RNA han revelado que éste no es el mejor indicador debido a la inestabilidad del rRNA 28S y a la formación de diferentes estructuras secundarias ribonucleares, lo cual no permite hacer una correlación lineal entre el valor 28S/18S y la degradación de RNA en la muestra (Imbeaud *et al.*, 2005). El desarrollo de la electroforesis en capilar asistida por la tecnología de microfluídos ha permitido proponer al RIN (Schroeder *et al.*, 2006) como una métrica alternativa a la relación de rRNA 28S/18S. El RIN es un parámetro numérico con valores del 1 al 10, con el cual se califica a las muestras desde totalmente degradadas hasta intactas, respectivamente. De acuerdo con las mediciones realizadas, la muestra TR 1 es la que tiene el RIN más bajo, con un valor de 8.3 (Fig. 3). Sin embargo, se conoce que para las plantas el parámetro RIN subestima la integridad de las extracciones de RNA por uno o dos puntos (Takahashi *et al.*, 2010). De este modo, podemos asumir que las muestras del RNA que obtuvimos no presentan evidencias significativas de degradación.

**Tabla II.** Caracterización espectrofotométrica del RNA de *Pachycerus pringlei* que se extrajo a partir de los ápices de la raíz primaria en tres etapas de desarrollo.

	INICIAL (IN)		INTERMEDIA (IT)		TERMINAL (TR)	
	1	2	1	2	1	2
<b>RNA (ng/μL)</b>	495	813	813	957	1200	843
<b>Absorbencia 260/280</b>	2.09	2.09	2.11	2.11	2.10	2.12
<b>Absorbencia 260/230</b>	1.63	1.52	1.93	1.42	1.37	1.22
<b>28S/18S</b>	1.40	1.60	1.70	1.60	1.50	1.60
<b>RIN</b>	8.80	9.30	9.60	9.50	8.30	9.00



**Figura 3.** Evaluación de la integridad del RNA que se extrajo a partir de las muestras de ápices de la raíz primaria de *Pachycereus pringlei* en tres etapas del desarrollo. Se incluyeron dos réplicas biológicas para cada etapa: inicial, intermedia y terminal. La estimación de integridad se realizó por el BGI mediante la determinación de la relación 28S/18S y del valor RIN al utilizar el equipo Agilent 2100®.

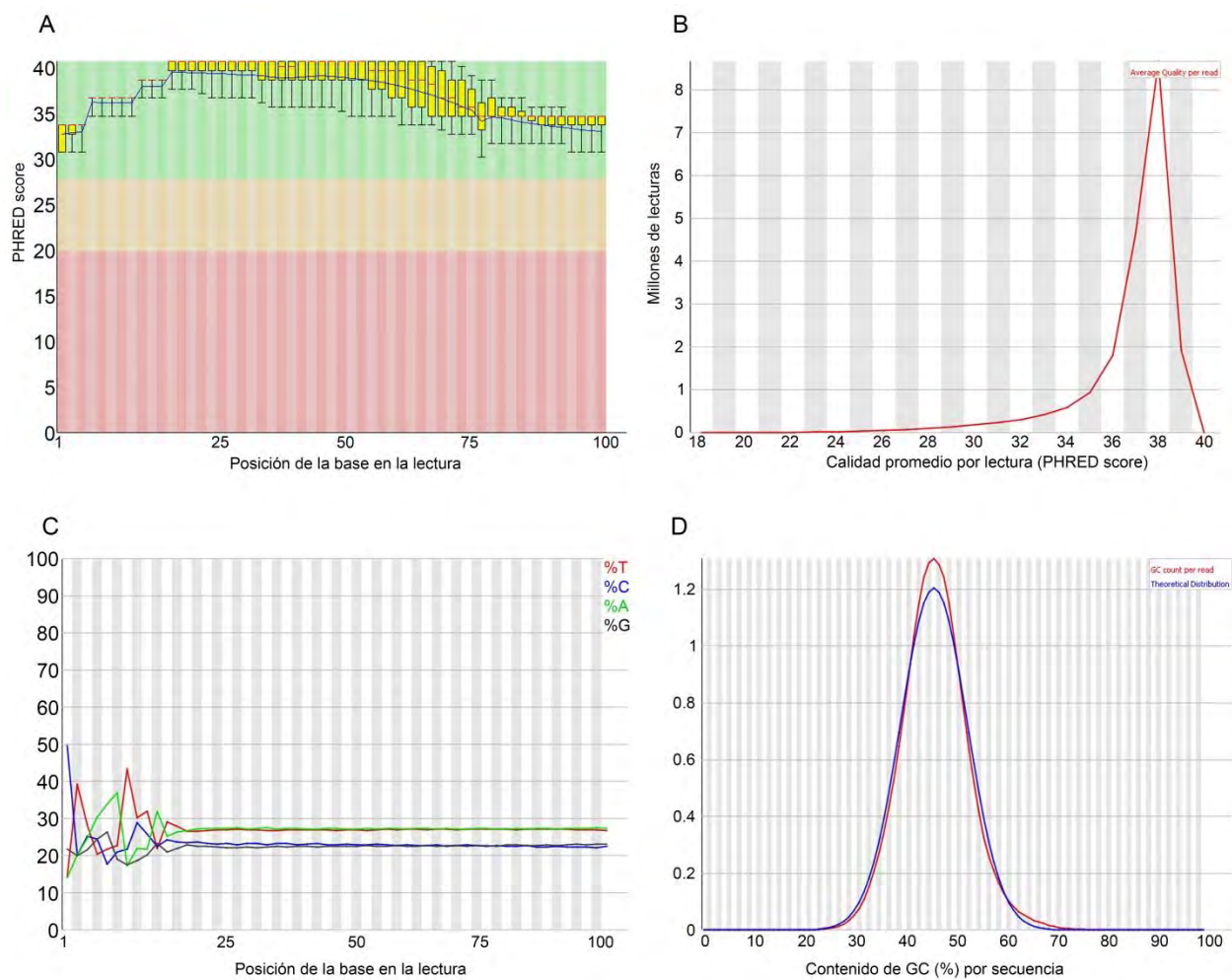
## V.2 SECUENCIACIÓN Y ENSAMBLAJE DEL TRANSCRIPTOMA

### V.2.1 Secuenciación del transcriptoma

La preparación de las bibliotecas para RNAseq, así como la secuenciación de las mismas, fue llevada a cabo por el BGI-Hong Kong al utilizar el protocolo de enriquecimiento de mRNA por selección positiva de RNAs poliadenilados. Los fragmentos de cDNA sintetizados a partir del mRNA del ápice de la raíz primaria de *P. pringlei* se secuenciaron a partir de sus dos extremos; es decir, en el formato de lecturas apareadas, durante 100 ciclos, con un tamaño promedio de inserto de 160 nucleótidos aproximadamente. Como resultado de la secuenciación se obtuvieron aproximadamente 280 millones de lecturas para las 6 muestras, con 46.5 millones de lecturas en promedio por muestra (Anexo 1). La calidad de las lecturas se analizó con el software FastQC. En la figura 4 se presentan los resultados de este análisis para las lecturas de la muestra IN 1, los cuales son representativos de los resultados que se obtuvieron para el resto de las muestras. Los resultados del análisis de calidad para todas las muestras se reportan en el Anexo 2.

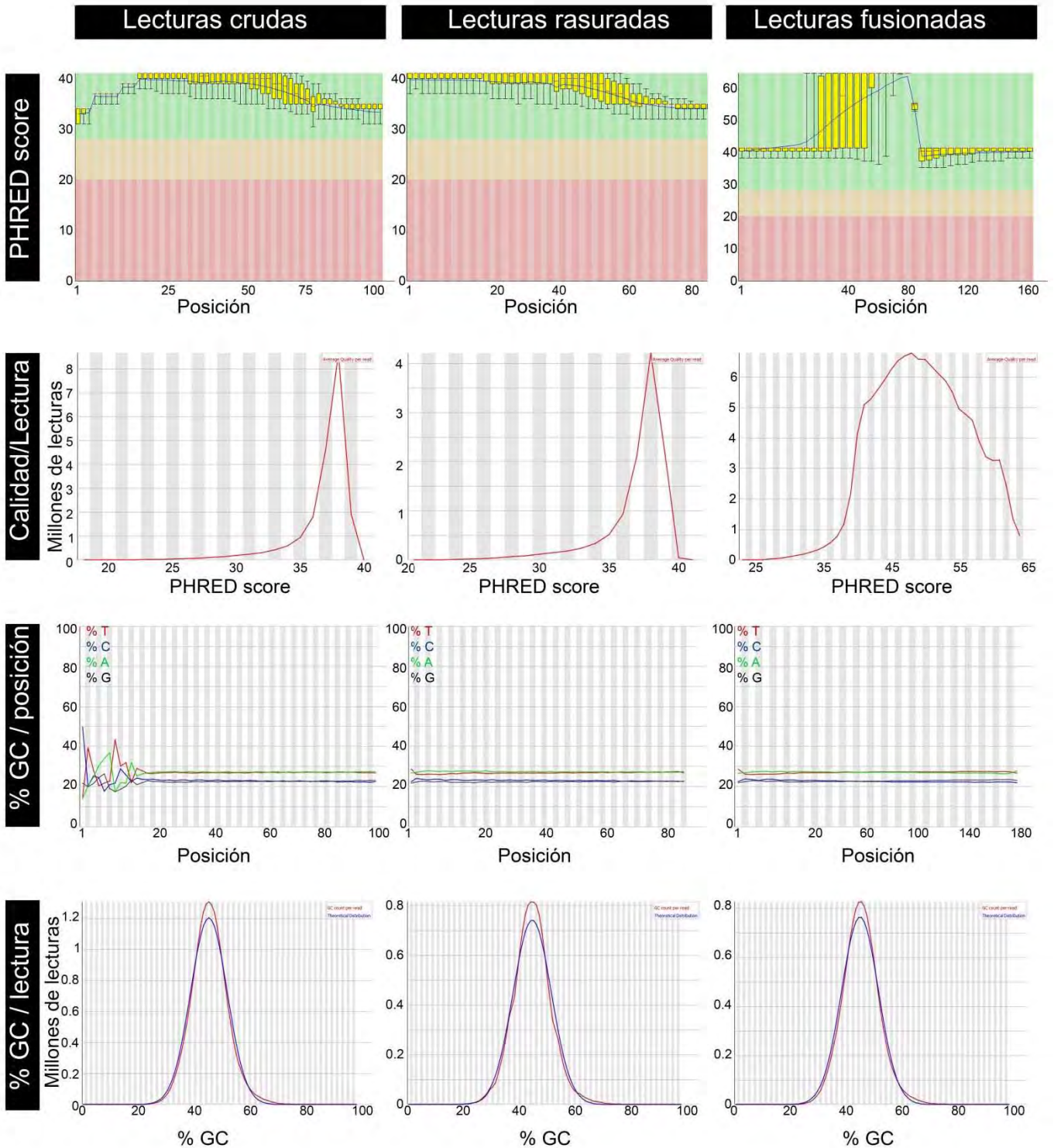
De acuerdo con los resultados de calidad de las lecturas (Fig. 4), es necesario realizar el “trimming” o rasurado de los primeros ~13 nucleótidos, pues la frecuencia de las bases en estas posiciones se desvía del comportamiento esperado (figura 4C, D; Anexo 2D), además de tener scores Phred, o de calidad, más bajos en estas regiones. Adicionalmente, hacia los extremos 5' y 3' de las lecturas se observa un incremento de heptámeros, es decir, en estas regiones existe una sobre-representación de secuencias cortas de 7 nucleótidos hacia los extremos de las lecturas (Anexo 2E). Es probable que estas secuencias correspondan a los adaptadores que se utilizaron durante la preparación de las bibliotecas de RNAseq, por lo cual se realizó la búsqueda y remoción de secuencias adaptadoras utilizadas por la compañía Illumina. Adicionalmente se evaluó si las dos secuencias que componen a un par proveniente del mismo fragmento se sobrelapan en sus extremos; cuando éste fue el caso, ambas lecturas se fusionaron para obtener un solo fragmento con valores Phred más altos, es decir, con probabilidades bajas de resolución incorrecta de bases. Del total de lecturas preprocesadas, 88 millones de pares fueron fusionadas en lecturas sencillas mientras

que 79 millones de pares no sobrelaparon en sus extremos. Una vez que se fusionaron las lecturas sobrelapantes, se realizó un rasurado o trimming adicional al remover a las bases no resueltas y a las bases ambiguas. Posteriormente, se descartó a aquellas lecturas que después de los filtros descritos fueron menores a 40 nucleótidos de longitud. El análisis de calidad de las lecturas después de haber sido procesadas mostró que las posiciones y lecturas de baja calidad se removieron de manera eficiente (Fig. 5). Las lecturas apareadas y fusionadas que cumplieron con los filtros de calidad descritos se utilizaron como datos de entrada para el ensamblaje contigs.



**Figura 4.** Evaluación de la calidad de las lecturas que se obtuvieron para la muestra Inicial 1. A: PHRED score por base; B: PHRED score por secuencia; C: Frecuencia de bases por posición; D: Contenido de GC por secuencia.





**Figura 5.** Análisis de calidad para las lecturas crudas (izquierda) y las lecturas procesadas cuando éstas quedan en pares (columna central) o cuando se fusionaron debido al traslape de sus extremos (derecha).

## V.2.2 Ensamblaje del transcriptoma

El transcriptoma se ensambló al utilizar el módulo Legacy del software CLC Genomics Workbench 7.5.1. Se realizaron diferentes ensamblajes al utilizar distintas combinaciones de los parámetros “tamaño de palabra” y “tamaño de burbuja”. Una explicación detallada de estos parámetros se encuentra en el Anexo 3. Los valores que se utilizaron para tamaño de palabra estuvieron en el intervalo de 50 a 64 y para tamaño de burbuja en el intervalo de 100 a 400. El tamaño mínimo aceptable para los contigs ensamblados se definió en 400 nt. Se escogió como mejor ensamblaje a aquel que maximizó al número de lecturas mapeadas en pares y el tamaño promedio de los contigs ensamblados, además de reducir la proporción de lecturas con mapeo ambiguo. Durante la construcción de los contigs, las lecturas se mapearon sobre los contigs generados y las lecturas no mapeadas se guardaron como un subconjunto independiente. En la primera ronda de ensamblaje se obtuvieron 46,311 contigs. Debido a que *P. pringlei* es una especie tetraploide (Murawski *et al.*, 1994), se realizó un ensamblaje adicional al utilizar al subconjunto de lecturas que no mapearon sobre los contigs durante la construcción de las gráficas de de Bruijn en el ensamblaje del transcriptoma. Este segundo ensamblaje se realizó bajo los mismos parámetros de tamaño de burbuja y de palabra que el primer ensamblaje con el fin de recuperar secuencias redundantes o altamente similares, las cuales se esperan a partir de familias génicas muy grandes y variantes alélicas. Los contigs que se obtuvieron a partir de estas dos rondas de ensamblaje se guardaron como un solo set de datos. Los contigs mayores a 3 kb se extendieron al utilizar el módulo “genome finishing” de CLC-GW. Con esta estrategia se obtuvieron 49,045 contigs en total (Tabla III, columna derecha). En las secciones subsecuentes los contigs serán referidos como transcritos. El transcriptoma que se ensambló para el ápice de la raíz primaria de *P. pringlei* representa un aumento significativo en el número de secuencias nucleotídicas provenientes de la familia Cactaceae. Hasta la generación de este transcriptoma, las bases de datos de acceso público tales como GenBank incluían sólo subsecuencias generadas a partir de bibliotecas de cDNA (ESTs, expressed sequence tags; Mallona *et al.*, 2011), o algunos transcriptomas con valores de N50 más bajos para explorar las rutas de síntesis de metabolitos tales como la mezcalina (Ibarra-Laclette *et al.*, 2015) o



las betalaínas (Qingzhu *et al.*, 2016). Aunque al comparar el número de secuencias ensambladas para el transcriptoma de *L. williamsii* (Ibarra-Laclette *et al.*, 2015) y el de *P. pringlei* generado en este trabajo se aprecia que ambos contienen un número similar de transcritos, el tamaño promedio de los transcritos reconstruidos de *P. pringlei* es aproximadamente dos veces mayor que el de los transcritos ensamblados para *L. williamsii* (Tabla III). Esto es particularmente evidente al comparar la longitud del transcrito más grande ensamblado para ambos transcriptomas; por ejemplo, en el transcriptoma de *L. williamsii* el tamaño del contig más largo es de 4,170 nt, mientras que en *P. pringlei* es de 14,583; similar a lo esperado para el transcrito del gen *BIG* de *A. thaliana* (AT3G02260; 15,294 bp), el cual se traduce en la proteína de mayor tamaño conocida en las especies vegetales.

**Tabla III.** Comparación del transcriptoma ensamblado para *Lophophora williamsii*\* y *Pachycereus pringlei*.

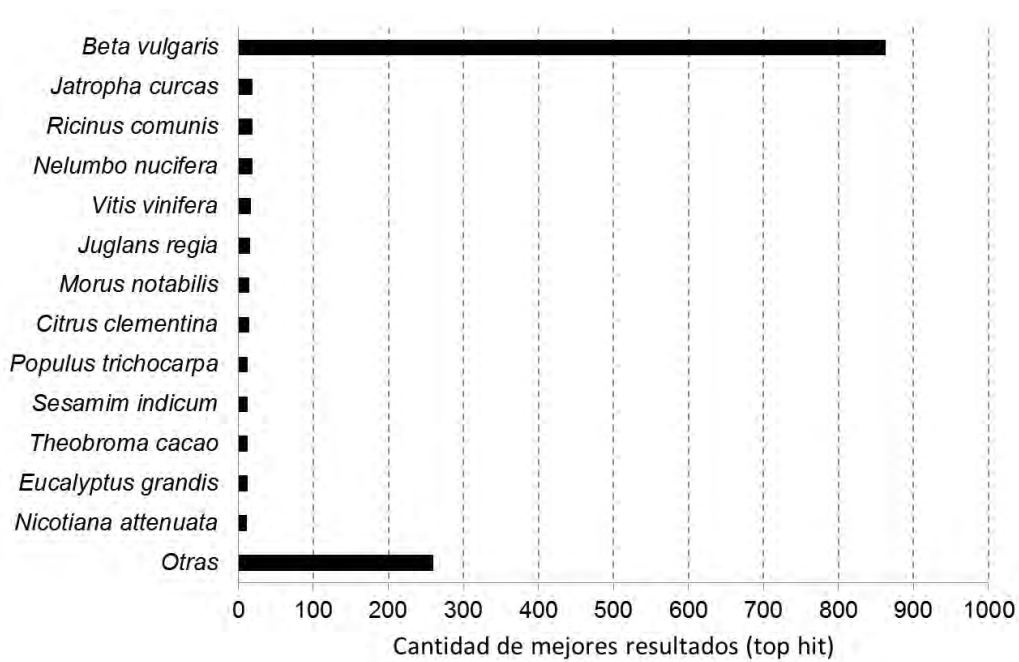
		<i>Lophophora williamsii</i>	<i>Pachycereus pringlei</i>
Secuenciación	Plataforma	GS (FLX, Junior) y PGM	Illumina HiSeq
	Formato de lectura	Single End	Paired End
	Lecturas totales	6x10 <sup>6</sup>	280x10 <sup>6</sup>
	Tamaño de lecturas	50-150 nt	100 (x2) nt
Ensamblaje	Transcritos ensamblados	40,436	49,045
	Transcrito más corto	200 nt	400 nt
	Transcrito más largo	4170 nt	14,583 nt
	Longitud de transcrito (promedio)	572 nt	1081 nt
	Longitud de transcrito (mediana)	454 nt	728 nt

\* Datos tomados de (Ibarra-Laclette *et al.*, 2015)

### V.3 ANOTACIÓN FUNCIONAL DE LOS TRANSCRITOS.

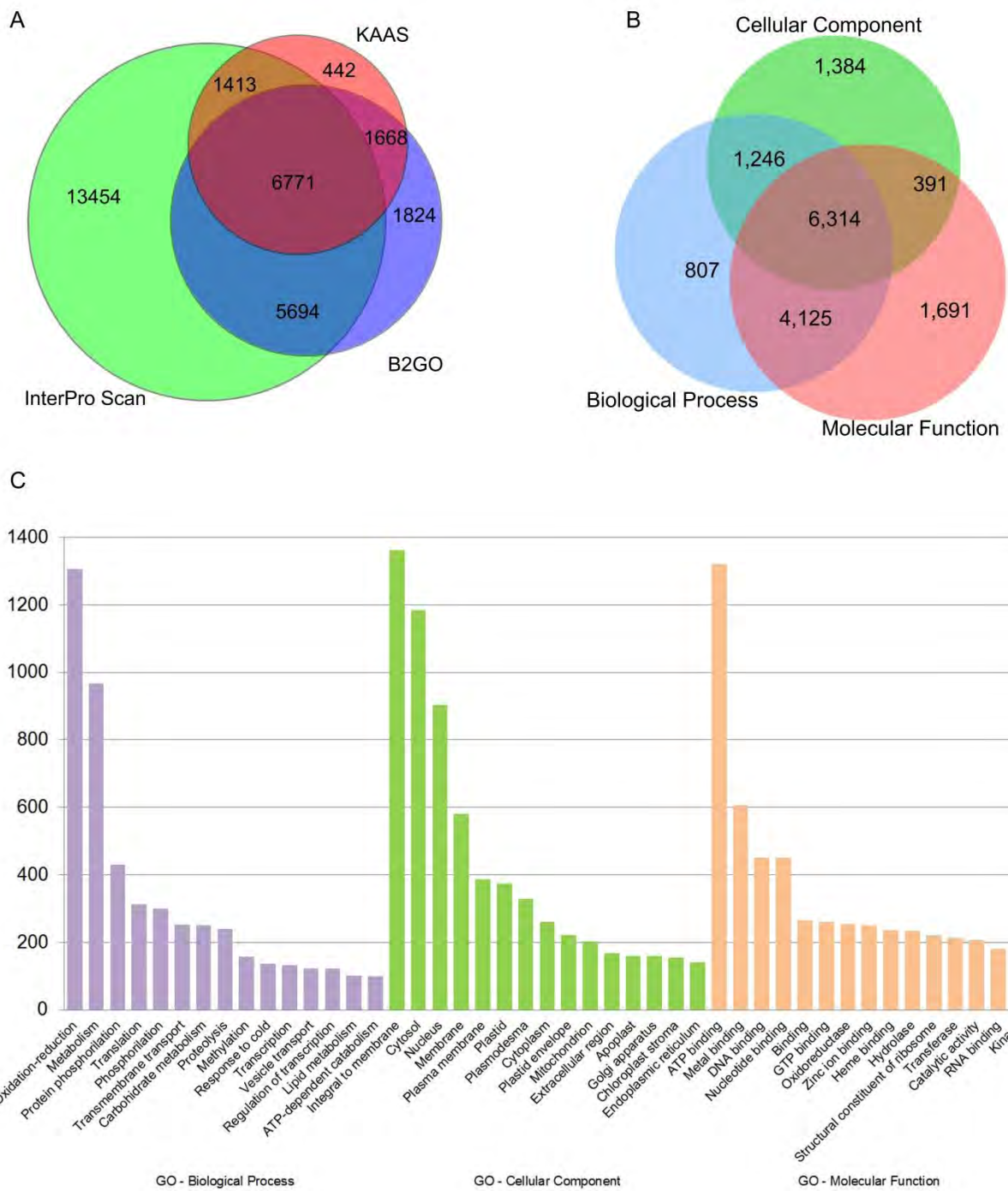
Para cada uno de los contigs que se ensamblaron se realizó la identificación de los posibles marcos abiertos de lectura (ORF) para identificar a la región codificante (CDS)

y a las regiones no traducidas (UTRs) de cada transcrito. A partir del ORF más largo se tradujo la secuencia de la proteína codificada por cada contig y ésta se utilizó como sonda para realizar la anotación funcional, la cual consistió en una búsqueda mediante la herramienta BlastX del NCBI para identificar a los homólogos de los contigs en la base de datos RefSeq de plantas dicotiledóneas. El mejor resultado, o “hit”, de la búsqueda se utilizó para recuperar las diferentes Ontologías Génicas (“Gene Ontologies”: Localización subcelular, Función Molecular y Proceso Biológico) al utilizar el plug-in Blast2GO PRO del software CLC Genomics Workbench. La especie de referencia a partir de la cual se obtuvo el mayor número de resultados para la anotación con ontologías génicas fue *Beta vulgaris* (Fig. 6). Esta especie de la familia Amaranthaceae fue la única del orden Caryophyllales, mismo orden al que pertenece la familia Cactaceae, incluida en la base de datos RefSeq del NCBI al momento de realizar la anotación. La herramienta B2GO permitió anotar 15,958 transcritos con al menos una ontología génica; de éstos, aproximadamente 6,300 contigs fueron anotados con las tres categorías GO (Fig. 7).



**Figura 6.** Especies de referencia contenidas en RefSeq a partir de las cuales se obtuvieron las secuencias más similares a los transcritos de *Pachycereus pringlei* durante la búsqueda de homólogos para la anotación de ontologías génicas.

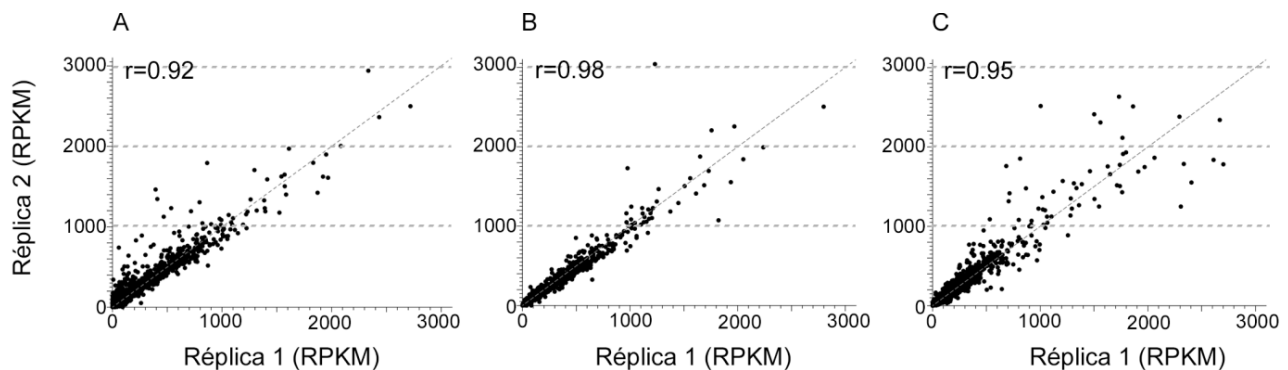
Adicionalmente se anotó a los contigs mediante el método InterPro Scan (Quevillon *et al.*, 2005), el cual identifica firmas de proteínas para asignar anotaciones de acuerdo con diversas bases de datos de acceso público. InterPro Scan permitió anotar 27,333 contigs. Una tercera anotación se realizó al utilizar el método KAAS (KEGG Automated Annotation Service; Moriya *et al.*, 2007) de la “Kyoto Encyclopedia of Genes and Genomes”. La anotación con KAAS permitió el mapeo de 10,294 contigs en rutas metabólicas, incluyendo la reconstrucción completa de las rutas de metabolismo central, señalización por fitohormonas; replicación de DNA eucariótico; factores de transcripción basales; síntesis, supervisión (“surveillance”) y degradación de mRNA; ribogénesis y procesamiento de proteínas en el retículo endoplásmico, entre otros (Anexo 4). Con los tres métodos mencionados se logró anotar alrededor de 31,200 contigs (Fig. 7), lo cual corresponde al 63.75% del total de contigs ensamblados; entre estos se encuentran algunos anotados como “proteínas hipotéticas” o “proteínas expresadas”. Aproximadamente 17,700 contigs no lograron anotarse mediante Blast2GO, InterProScan o KEGG. Debido a que la anotación de nuevas secuencias depende en esencia de las anotaciones generadas para organismos modelo, en general un porcentaje significativo de secuencias no pueden anotarse o se anotan como proteínas predichas o no caracterizadas. Adicionalmente las secuencias de RNA que carecen de ORF, tales como los RNAs largos no codificantes o los precursores de microRNAs, entre otros, presentan tasas de evolución más altas y por lo tanto menos conservación entre especies (Johnsson *et al.*, 2014; Meunier *et al.*, 2013). De esta forma el porcentaje de secuencias no anotadas, o con anotaciones de baja calidad, es directamente proporcional a la distancia evolutiva entre el organismos en estudio y el organismo filogenéticamente más cercano con un genoma bien anotado.



**Figura 7.** Anotación funcional de los transcritos de *Pachycereus pringlei* contenidos en el transcriptoma del ápice de la raíz primaria. (A) Número de contigs o transcritos ensamblados que se anotaron en este trabajo; (B) transcritos anotados por ontología génica; (C) subcategorías más representadas en el transcriptoma por cada ontología génica.

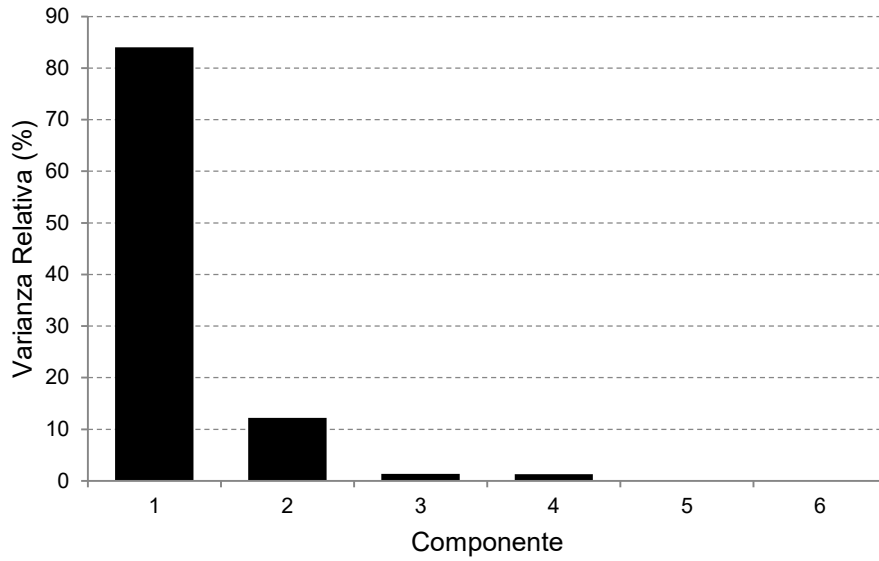
## V.4 CARACTERIZACIÓN DEL TRANSCRIPTOMA DEL ÁPICE DE LA RAÍZ PRIMARIA DE *Pachycereus pringlei*

Los contigs que se ensamblaron se utilizaron como secuencias de referencia para mapear las lecturas de cada etapa y cuantificar la abundancia relativa de cada transcrito en las diferentes etapas del crecimiento de la raíz primaria de *P. pringlei*. Se contó el número total de lecturas que mapearon sobre cada transcrito y este número se normalizó mediante el método de lecturas por kilobase por millón de lecturas mapeadas (RPKMs; Reads Per Kilobase per Million Mapped Reads, Mortazavi *et al.*, 2008) para compensar los posibles sesgos que surgen debido a la variación en el tamaño de las bibliotecas o a la variación de tamaño entre los diferentes transcritos. Se estimó la variabilidad de los niveles de expresión de los transcritos entre cada par de réplicas biológicas al visualizar a los valores de RPKM de cada transcrito en una gráfica de dispersión (Fig. 8). Como se espera, la dispersión de los valores de RPKM entre las réplicas biológicas de cada etapa puede ajustarse a una diagonal de identidad. El factor de correlación  $r$  para cada par de réplicas se encuentra en un intervalo de 0.92-0.98, siendo las réplicas de la etapa IN las que tienen menor índice de correlación y las de la etapa IT las réplicas con la correlación más alta. Esta observación indica que las réplicas biológicas de la etapa IT presentan la menor variabilidad en los niveles de expresión de los transcritos ensamblados, mientras que las de la etapa IN presentan la mayor variabilidad.

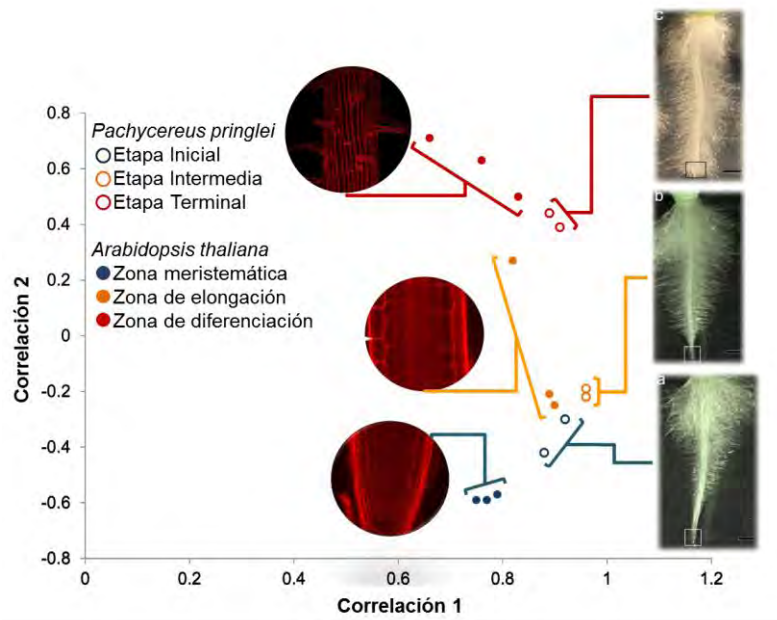


**Figura 8.** Variabilidad entre las réplicas biológicas de las muestras Inicial (A), Intermedia (B) y Terminal (C) del crecimiento de la raíz primaria de *Pachycereus pringlei*.

Adicionalmente se utilizó a los valores de RPKM para realizar un análisis de componente principal (PCA) para estimar el número y peso de los factores más importantes de variabilidad entre las diferentes muestras. La gráfica de Scree (Fig. 9; Cattell, 1966) muestra que la mayor parte de la variabilidad de las muestras se puede atribuir a dos componentes principales, los cuales en su conjunto representan una varianza acumulada del 96.2%. Las proyecciones ortogonales de los dos componentes principales se graficaron para estimar las distancias relativas entre las réplicas biológicas de cada etapa y entre las diferentes etapas. En general, se observa que las dos réplicas biológicas que corresponden a la misma etapa se agrupan con su par y se distinguen de las muestras de otras etapas (Fig. 10); sin embargo, esto no ocurre para una de las réplicas de la etapa IN, la cual se posiciona entre su par biológico y las muestras de la etapa IT (Fig. 10). Las muestras de la etapa TR se agrupan de manera distante a las muestras de las etapas IN e IT; esto se debe, probablemente, a que en la etapa TR del desarrollo de la raíz primaria todas las células del ápice de la raíz, incluyendo a aquellas que componían al RAM, están diferenciadas; mientras que las muestras de las etapas IN e IT incluyen tanto a células meristemáticas como a células de la zona de elongación, aunque estas dos poblaciones están en diferente proporción en cada etapa. Las muestras de la etapa IT son las que exhiben menor variabilidad, como se aprecia debido a la cercanía de los puntos de las réplicas biológicas IT 1 y 2, lo cual es congruente también con el valor de correlación  $r$  obtenido a partir del ajuste de la dispersión de los RPKMs a una recta (Fig. 8). De acuerdo con lo observado en la gráfica de PCA, las muestras de la etapa IN son las que presentan la mayor variabilidad. Esto se podría explicar al considerar que las muestras de ápices de la raíz primaria colectadas para la etapa IN constituyen a la muestra más heterogénea, pues fueron obtenidas a partir de plántulas de entre 1 y 4 d.d.g. La separación de las muestras de acuerdo con la etapa de desarrollo de la raíz primaria sugiere que la variabilidad observada entre las muestras de las distintas etapas se debe al estadio de desarrollo, y apunta a la existencia de programas transcripcionales particulares y específicos de cada etapa durante el desarrollo de la raíz primaria de *P. pringlei*.



**Figura 9.** Gráfica de Scree de los componentes principales estimados para las muestras utilizadas en este estudio. La mayor parte de la variabilidad observada entre las muestras se atribuye a los dos primeros componentes, cuya varianza acumulada es del 96.2%.



**Figura 10.** Análisis de componente principal para las muestras que provienen de ápices de raíces en distintas etapas del desarrollo de la raíz primaria de *Pachycereus pringlei* y de las tres zonas longitudinales de la raíz primaria de *Arabidopsis thaliana*. Los datos crudos de *A. thaliana* fueron obtenidos de Huang y Schiefelbein (2015), procesados y mapeados al genoma de referencia correspondiente. Los ejes son adimensionales (covarianza(A, B)/[d.s.(A)\*d.s.(B)]).

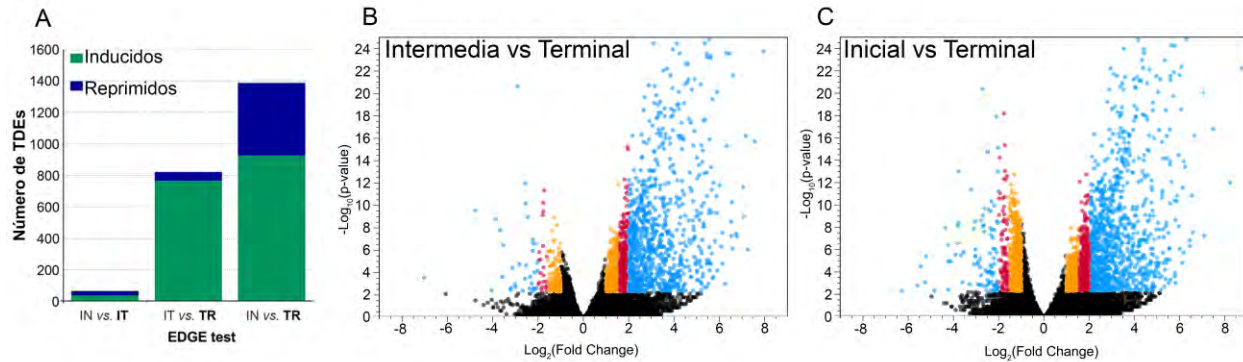
Para analizar si el estado transcripcional de las células del ápice de la raíz primaria de *P. pringlei* en la etapa IN, IT y TR, es comparable al estado transcripcional de las células de la zona meristemática, de elongación y de diferenciación de la raíz primaria de *A. thaliana*, respectivamente, se utilizaron los datos de RNAseq generados por Huang y Schiefelbein (2015). Las lecturas crudas que se generaron para cada zona longitudinal de la raíz primaria de *A. thaliana* se descargaron y se procesaron como se describió para las lecturas que generamos para *P. pringlei*. Una vez procesadas, las lecturas se mapearon al genoma de referencia de *A. thaliana* (Lamesch *et al.*, 2012), se estimó la abundancia relativa de cada transcrito en las diferentes zonas y se utilizó a los valores de RPKM para realizar un análisis de PCA a partir de ellos. Los resultados que se obtuvieron a partir del PCA de *Arabidopsis* se visualizaron en la misma gráfica que los datos de *P. pringlei* (Fig. 10). Como se esperaba, cada una de las tres réplicas biológicas que se analizaron para las zonas longitudinales de la raíz primaria de *A. thaliana*, agrupa más cercanamente con sus pares biológicos que con las muestras provenientes de otras zonas, excepto por una de las réplicas biológicas de la zona de elongación, la cual se agrupa mejor con las muestras que se obtuvieron a partir de la zona de diferenciación. La sobreposición de los resultados del PCA de *A. thaliana* y de *P. pringlei* (Fig. 10) sugiere que las células del ápice de la raíz primaria en la etapa TR son funcionales incluso cuando el RAM está agotado y que llevan a cabo las funciones típicas de las células que se encuentran en la zona de diferenciación de la raíz de otras angiospermas. Estos resultados son congruentes con resultados anteriores en los cuales se demostró que el agotamiento del RAM de las cactáceas *S. gummosus* y *P. pringlei* no involucra a la muerte celular, aunque este fenómeno sí ocurre en algunas células de la cofia (Shishkova and Dubrovsky, 2005). El ligero corrimiento en la posición de las muestras de la etapa TR de *P. pringlei* hacia las muestras de las etapas IT e IN, sugiere que aún cuando las células del ápice están diferenciadas se mantiene la expresión de algunos programas propios de células menos diferenciadas. Esta hipótesis, sin embargo, debe ser explorada con mayor profundidad. Por otro lado, se observó que las muestras de la etapa IT se agrupan con las muestras de la zona de elongación, mientras que las muestras de la etapa IN se encuentran entre las muestras provenientes de las zonas meristemática y de elongación. Esto último puede explicarse



con base en el tamaño del fragmento del ápice colectado, pues al cortar 1 mm del ápice de la raíz primaria de *P. pringlei* en la etapa IN, se incluyen células de la zona meristemática pero también células de la zona de elongación. A partir de este análisis podemos concluir que los programas de diferenciación celular ocurren en dos sitios de la raíz primaria de *P. pringlei*: en las células distales al RAM a lo largo del eje longitudinal de la raíz primaria, tal como se observa en otras angiospermas, pero también en el ápice de la raíz de acuerdo con la progresión del agotamiento del meristemo.

## V.5. EVALUACIÓN DE LA EXPRESIÓN DIFERENCIAL

Se realizó una prueba de detección empírica de la expresión diferencial (EDGE, por sus siglas en inglés; Robinson *et al.*, 2010). *Grosso modo*, este análisis calcula el “fold change”, o cambio en la abundancia de un transcrito entre dos condiciones, a partir de los valores de abundancia relativa de los transcritos en cada réplica biológica de cada condición. Se asigna un valor  $p$  de probabilidad de expresión diferencial para cada transcrito y su variabilidad entre las réplicas de la misma etapa. Al graficar los valores de probabilidad ( $-\log p\text{-val}$ ) vs el cambio en la abundancia relativa entre los transcritos de diferentes etapas ( $-\log_2(\text{Fold Change})$ ) es posible identificar y seleccionar a los transcritos que cambian de manera significativa su expresión entre las diferentes etapas del crecimiento de la raíz primaria de *P. pringlei*. La prueba de EDGE se realizó para cada par de etapas del desarrollo de la raíz primaria de *P. pringlei*. El menor número de transcritos diferencialmente expresados (TDEs) se encontró al comparar a las etapas IN vs IT, debido probablemente a la presencia tanto de células meristemáticas como de células de la zona de elongación, aunque en diferente proporción, en las muestras de ambas etapas (Fig. 11A). Por su parte, al definir los umbrales de cambio en la expresión en  $\text{Log}_2(\text{FC}) \geq 2$  y con  $\text{FDR } p\text{-val} \leq 0.005$ , se encontraron 873 y 1,388 TDEs al comparar a las etapas IT vs TR e IN vs TR, respectivamente (Fig. 11). Estos resultados muestran que el número de TDEs aumenta a medida que el RAM se agota.



**Figura 11.** Análisis de expresión diferencial entre pares de etapas del desarrollo de la raíz primaria de *Pachycereus pringlei*. (A) Número total de transcritos diferencialmente expresados (TDEs). B-C: Gráficas de volcán que se obtienen al comparar a las muestras Intermedia vs Terminal (B) e Inicial vs Terminal (C). La gráfica de volcán para la prueba EDGE entre las muestras Inicial e Intermedia no se incluye debido a que entre ambas existen muy pocos TDEs. TDEs: ●  $2 < \text{Fold Change} < 3$ ; ●  $3 < \text{Fold Change} < 4$ ; ●  $\text{Fold change} > 4$ .

De manera interesante, mientras que el número de genes inducidos (upregulated) y reprimidos (downregulated) en la comparación IN vs IT es similar (Fig. 11A), el número de genes inducidos es aproximadamente 2 y 14 veces más alto que el número de genes que se reprimen al comparar a las etapas IN vs TR e IT vs TR, respectivamente (Fig. 11A). Esto es congruente con la noción general de que las células diferenciadas expresan un mayor número de genes que las células no diferenciadas. Para evaluar si el número de genes que se expresan en las células de la zona de diferenciación de la raíz primaria de las angiospermas es mayor al número de genes que se expresan en las zonas meristemática y de elongación, se descargaron y procesaron las lecturas crudas generadas por Huang y Schiefelbein (2015) para diversas angiospermas. Las lecturas se mapearon sobre el genoma de referencia correspondiente y se identificaron a los TDEs entre las distintas etapas (Tabla IV). Se observa que el número de genes inducidos en la zona meristemática es menor en comparación con los genes que se inducen en las zonas de elongación o de diferenciación. La excepción es *Cucumis sativus*; sin embargo, esta incongruencia aparente se explica con base en la organogénesis de las raíces laterales en *C. sativus*, pues mientras que el desarrollo de los primordios de raíces laterales ocurre en la zona de diferenciación para la mayoría de las angiospermas, en *C. sativus* este proceso empieza desde la zona meristemática (Dubrovsky and Rost, 2003).

**Tabla IV.** Genes diferencialmente expresados entre las zonas longitudinales de la raíz primaria de distintas angiospermas.

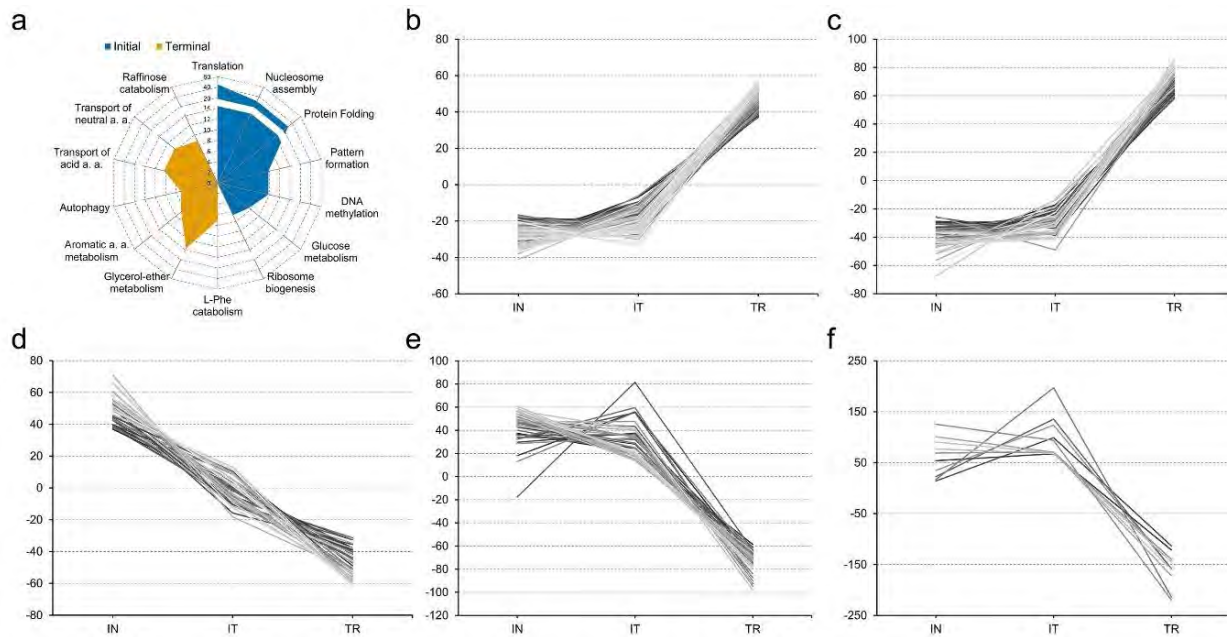
Species	EDGE MZ vs EZ		EDGE MZ vs DZ	
	MZ	EZ	MZ	DZ
<i>A. thaliana</i>	2,683	4,432	3,734	8,947
<i>C. sativus</i>	601	380	1,341	301
<i>G. max</i>	1,357	4,159	3,866	7,685
<i>O. sativa</i>	772	1,655	3,084	7,743
<i>S. lycopersicum</i>	2,905	3,542	4,125	5,661
<i>Z. mays</i>	2,018	3,285	8,012	8,563

Entre los transcritos con mayor abundancia en la etapa TR, es decir, aquellos cuya expresión se aumenta con el agotamiento del RAM, se encuentran algunos transportadores y genes de respuesta a auxinas; varios factores transcripcionales de la familia NAC, los cuales están implicados en distintos procesos de desarrollo de las plantas y en respuestas al estrés abiótico (Olsen *et al.*, 2005); y factores transcripcionales de la familia WRKY, algunos de los cuales están involucrados en las respuestas a distintos tipos de estrés (Rushton *et al.*, 2010). En la etapa TR también se observa una mayor abundancia de transcritos que codifican para proteínas homólogas a TERMINAL FLOWER 1, cuya pérdida de función en *Arabidopsis* resulta en la transición del meristemo de la inflorescencia, normalmente indeterminado, hacia un meristemo determinado al inducir la formación de una flor terminal que abate el crecimiento del brote (Bradley *et al.*, 1997; Hanano and Goto, 2011). Otros transcritos inducidos en la etapa TR incluyen a algunos homólogos de FLOWERING PROMOTING FACTOR, cuya expresión en raíces de *Oryza sativa* está asociada con la formación de un sistema radical más ramificado debido a la inducción de raíces laterales (Ge *et al.*, 2004).

### V.5.1 Análisis de conglomerados

Se tomó al conjunto de contigs que se expresan diferencialmente entre las etapas IN y TR para realizar un análisis hipergeométrico basado en las anotaciones de proceso biológico de las ontologías génicas (Falcon and Gentleman, 2007). Como se observa en la figura 12, procesos tales como la síntesis y el plegamiento de proteínas,

la metilación de DNA, el ensamblaje de nucleosomas y la biogénesis del ribosoma se encuentran sobrerrepresentados en la etapa IN, cuando las muestras incluyen una cantidad significativa de células meristemáticas, es decir, con actividad proliferativa. Por otro lado, en la etapa TR, cuando las células del ápice de la raíz primaria están diferenciadas, encontramos una sobrerrepresentación de procesos biológicos tales como el transporte y metabolismo de amino ácidos y compuestos de glicerol-éster, entre otras.



**Figura 12.** Análisis hipergeométrico (a) y de conglomerados (b-f) de los transcritos que se expresan diferencialmente. (A) Subcategorías de proceso biológico enriquecidas en las etapas inicial o terminal del desarrollo de la raíz primaria de *Pachycereus pringlei*. El eje muestra el  $-\log_{10}(p\text{-val})$  que se obtuvo en las pruebas de enriquecimiento. (b-f) Perfiles de expresión representativos en los cuales se puede agrupar a los transcritos expresados de manera diferencial. El eje de las ordenadas muestra a los valores de expresión (RPKMs), centralizados y normalizados, para las etapas inicial (IN), intermedia (IT) y terminal (TR).

Adicionalmente se realizó un análisis de conglomerados para agrupar a los TDEs entre las etapas IN y TR en conjuntos con perfiles de expresión similar. Este tipo de análisis agrupa a los transcritos de acuerdo con la “distancia” entre sus variables; en este caso, los niveles de expresión virtual en cada etapa. De este modo cada “cluster”,

o conglomerado, contiene transcritos cuyos cambios de expresión entre las etapas de desarrollo son similares, al mismo tiempo que maximiza la distancia entre perfiles de expresión de distintos conglomerados. Para realizar este análisis se utilizó el método de *K-means* (Lloyd, 1982). En total se obtuvieron 27 conglomerados de los cuales se presentan cinco perfiles representativos en la figura 12b-f. Los 27 conglomerados pueden consultarse en el Anexo 5. Este análisis muestra que los 27 conglomerados se pueden agrupar en tres perfiles de expresión (algunos perfiles representativos se muestran en la figura 12b-f, mientras que el total de conglomerados puede consultarse en el Anexo 5): 14 conglomerados de transcritos cuya expresión aumenta en la etapa TR (Fig. 12b-c; Anexo 5); 12 conglomerados de transcritos cuya expresión disminuye conforme progresa la diferenciación de las células del ápice (Fig. 12d-e; Anexo 5); y un único conglomerado que exhibe un ligero pico en la etapa IT en comparación con las etapas IN y TR (Fig. 12f; Anexo 5). Este análisis resulta útil pues se espera que varios transcritos que se encuentran bajo la influencia del mismo regulador transcripcional, o que codifican para proteínas involucradas en procesos biológicos relacionados, se agrupen en el mismo conglomerado. De manera interesante, encontramos que los transcritos que codifican para proteínas involucradas en la autofagia están agrupadas en un conglomerado (Fig. 12b); mientras que aquellos que codifican para proteínas del metabolismo y transporte de aminoácidos se agrupan en un conglomerado distinto (Fig. 12c). Entre los transcritos cuya abundancia disminuye hacia la etapa TR, los que codifican para proteínas involucradas en el ensamblaje de nucleosomas, la traducción y la metilación del DNA están agrupadas en el mismo conglomerado (Fig. 12d); lo mismo que los transcritos que se traducen en enzimas involucradas en el plegamiento de proteínas y su desubiquitilación, así como en el procesamiento de rRNA (Fig. 12e). Por último, el conglomerado que exhibe un pico en la etapa IT contiene a transcritos que codifican para proteínas involucradas en la traducción y la respuesta a algunos tipos de estrés (Fig. 12f). Al considerar a los 27 conglomerados (Anexo 5) que se construyeron, observamos que varios transcritos que codifican para proteínas ribosomales de la subunidad grande o pequeña, cuya expresión disminuye con el progreso del agotamiento del meristemo, están contenidas en el mismo conglomerado. Otros transcritos cuya expresión está disminuída en la etapa TR, en comparación con las

etapas IN o IT, codifican para histonas, histon-desacetilasas, algunas peroxidasas, y proteínas reguladas por giberelinas, tales como los ortólogos putativos de GASA1 y GASA4/6. Por otro lado, entre los transcritos que se inducen hacia la etapa TR se encuentran varios transcritos que codifican para proteínas involucradas en distintos aspectos de la señalización por ácido abscísico. Esta observación, en conjunto con la inducción de algunos factores transcripcionales de las familias WRKY y NAC, sugiere que los programas genéticos para contender contra el estrés por déficit hídrico se inducen en las células del ápice de la raíz primaria de forma independiente de las condiciones de crecimiento, dado que las plántulas crecidas *in vitro* a partir de las cuales se realizó la colecta de las muestras no experimentaron estrés hídrico.

## **V.6 VALIDACIÓN DEL RNAseq MEDIANTE RT-QPCR**

### **V.6.1 Selección de genes normalizadores para ensayos de RT-qPCR**

Para corroborar los datos de expresión virtual que se obtuvieron mediante la secuenciación del transcriptoma se realizaron ensayos de RT-qPCR. Debido a que los análisis de RT-qPCR dependen de manera fundamental de la selección de genes normalizadores cuya expresión sea estable en todas las condiciones experimentales bajo estudio, identificamos primero a los transcritos de *P. pringlei* que podrían ser utilizados como genes de referencia. Para ello consideramos a los genes propuestos como genes normalizadores superiores de *A. thaliana* propuestos por Czechowski *et al.* (2005). Se realizó la búsqueda de los ortólogos putativos de estos genes mediante un análisis de mejor hit bidireccional al tomar como sonda a las secuencias de *A. thaliana* y como base de datos al transcriptoma ensamblado para el ápice de la raíz primaria de *P. pringlei*. En total, se encontraron 23 ortólogos putativos de los genes de referencia propuestos por Czechowski (Tabla V). Para seleccionar entre estos a los genes que se utilizarán como normalizadores, se compararon los valores de RPKMs de estos transcritos en *P. pringlei* y se generó una lista de aquellos que de acuerdo con los resultados del transcriptoma no cambian su expresión (Fig. 13). A partir de lo observado se seleccionó a *TRANSLATION ELONGATION FACTOR 1 $\alpha$*  (*EF1 $\alpha$* ), *UBIQUITIN CONJUGATING ENZYME9* (*UBC9*) y *YELLOW-LEAF-SPECIFIC GENE 8* (*YLS8*) como genes de referencia para normalizar a las abundancias relativas de los

genes selectos en los ensayos de RT-qPCR. Los oligonucleótidos que se diseñaron para la amplificación de los genes de referencia están reportados en la Tabla VI.

### **V.6.2 Ensayo de eficiencia para oligonucleótidos de RT-qPCR**

Los oligonucleótidos que se reportan en la Tabla VI se sintetizaron en la Unidad de Síntesis y Secuenciación de DNA (Instituto de Biotecnología – UNAM). Se utilizó el cDNA sintetizado a partir de RNA de raíces completas para calcular la eficiencia de amplificación con estos oligonucleótidos. Los resultados de estos ensayos se reportan en la figura 14, la cual confirma que con cada par de oligonucleótidos se amplifica un sólo producto (panel dF/dT, Fig. 14). Los resultados que se obtuvieron en las reacciones de amplificación, al utilizar diferentes diluciones de cDNA para cada par de oligonucleótidos, muestran que la eficiencia de amplificación es aceptable (entre 1.87 y 2; panel Cq, figura 14), al igual que los valores de correlación  $r^2$  de la ecuación de la curva ( $\sim 0.99$ , panel Cq; figura 14). Con base en estas observaciones consideramos que los oligonucleótidos que se diseñaron (Tabla VI) pueden ser utilizados para amplificar a los genes que se utilizarán como normalizadores, los cuales, a su vez, permitirán cuantificar la abundancia relativa del transcrito de distintos transcritos selectos.

**Tabla V.** Valores de expresión virtual (RPKM) para los ortólogos putativos de *P. pringlei* identificados a partir de los genes normalizadores superiores de *A. thaliana*.

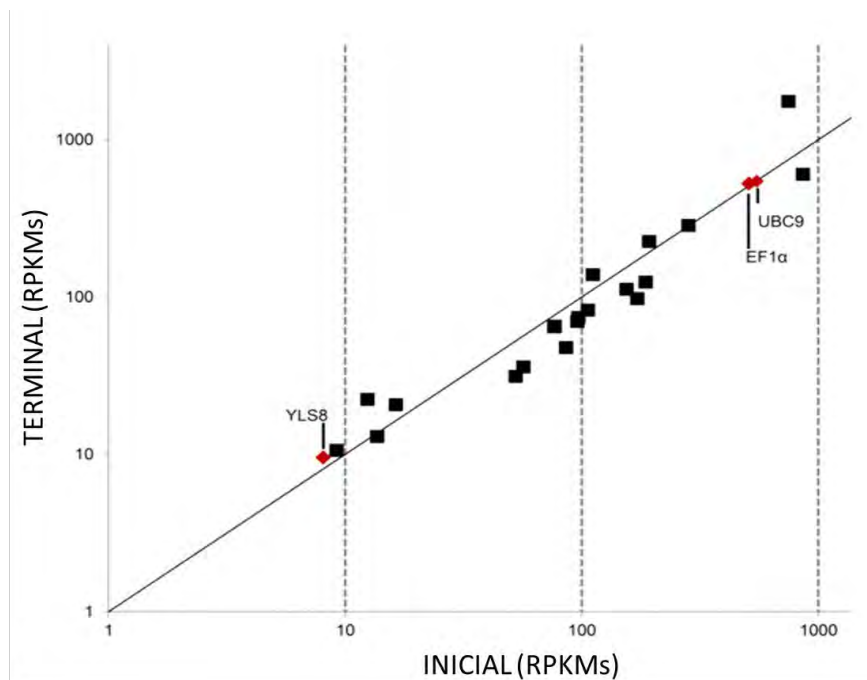
Descripción	Contig	INICIAL		INTERMEDIA		TERMINAL	
		1	2	1	2	1	2
AcylCoA BP	3269	73.21	58.6	66.48	66.63	38.02	38.08
PP2A	5248	142.42	121.42	123.2	123.98	77.23	78.74
AcylCoA BP_2	2079	43.93	36.76	39.08	40.8	24.55	25.43
CLA	9503	44.57	42.57	45.06	40.6	26.73	30.67
PP2A_2	9902	149.25	136.93	142.45	131.55	99.33	99.38
E2F $\alpha$	5001	717.61	604.83	684.04	727.78	545	415.74
PP2A_3	8913	112.74	124.83	112.59	100.72	88.43	90.68
CLA_2	12003	76.83	70.33	70.34	74.13	51.9	55.87
PP2A_4	7793	73.44	75.34	66.15	55.63	53.3	65.22
PP2A_5	4543	84.65	78.54	76.9	77.21	61.8	70.07
UBC	22624	59.02	58.65	56.54	54.67	48.15	55.25
KOR1	3527	58.84	59.27	64.18	72.66	54.75	49.39
KOR1_2	22902	10.52	10.46	14.94	17.09	10.13	10.62
UBC_2	24691	414.88	427.47	396.49	377.1	402.53	470.79
UBC9	15035	229.12	204.45	210.96	195.78	226.4	228.57
E1F $\alpha$	6759	401.27	379.59	363.6	388.22	428.72	413.51
YLS8	17282	145.98	150.19	156.32	155.68	183.81	175.92
GalaOx	3600	7.44	6.67	8.34	8.38	9.1	7.8
MADS	7058	5.9	6.49	5.29	7.16	7.85	7.45
UBC_4	17222	80.47	90.87	89.78	90.51	100.3	121.37
PTB	48438	12.67	12.51	16.71	20.22	16.55	16.46
YLS8_2	30952	9.48	9.62	7.07	7.67	19.24	16.42
UBC_3	35678	525.63	622.83	655.12	672.13	1546.79	1249.75

Los genes que se utilizaron como normalizadores para los ensayos de RT-qPCR están señalados dentro de la caja con bordes dobles y corresponden a UBC9, E1F $\alpha$  y YLS8.

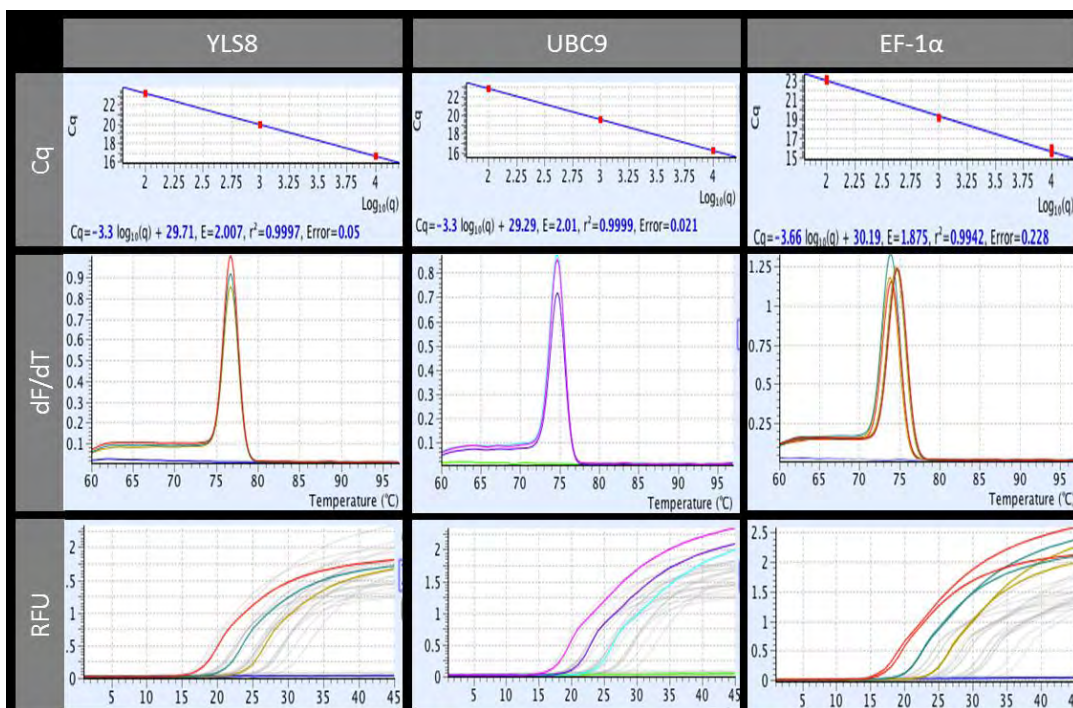
**Tabla VI.** Oligonucleótidos diseñados para la amplificación de los genes de referencia que se utilizaron en este trabajo.

Clave	Secuencia	Longitud	Tm NN	GC (%)
E1F $\alpha$ _FWD	GAGGTCTGAAGTTATGGGTCGTG	23 nt	63°C	52.2
E1F $\alpha$ _RV	AGCACACCGACGCTACTAATG	21 nt	62.2°C	52.4
UBC9_FWD	GGTGTTCAGGTATTGAGCTG	21 nt	61.4°C	52.4
UBC9_RV	AGCCATCCTCACAAAAGCAACC	22 nt	63°C	50
YLS8_FWD	ACCGTTCAGGTTGAGGCGAAG	21 nt	64.2°C	57.1
YLS8_RV	GATACCCCAAATAAGCATAGGC	23 nt	63.1°C	47.8





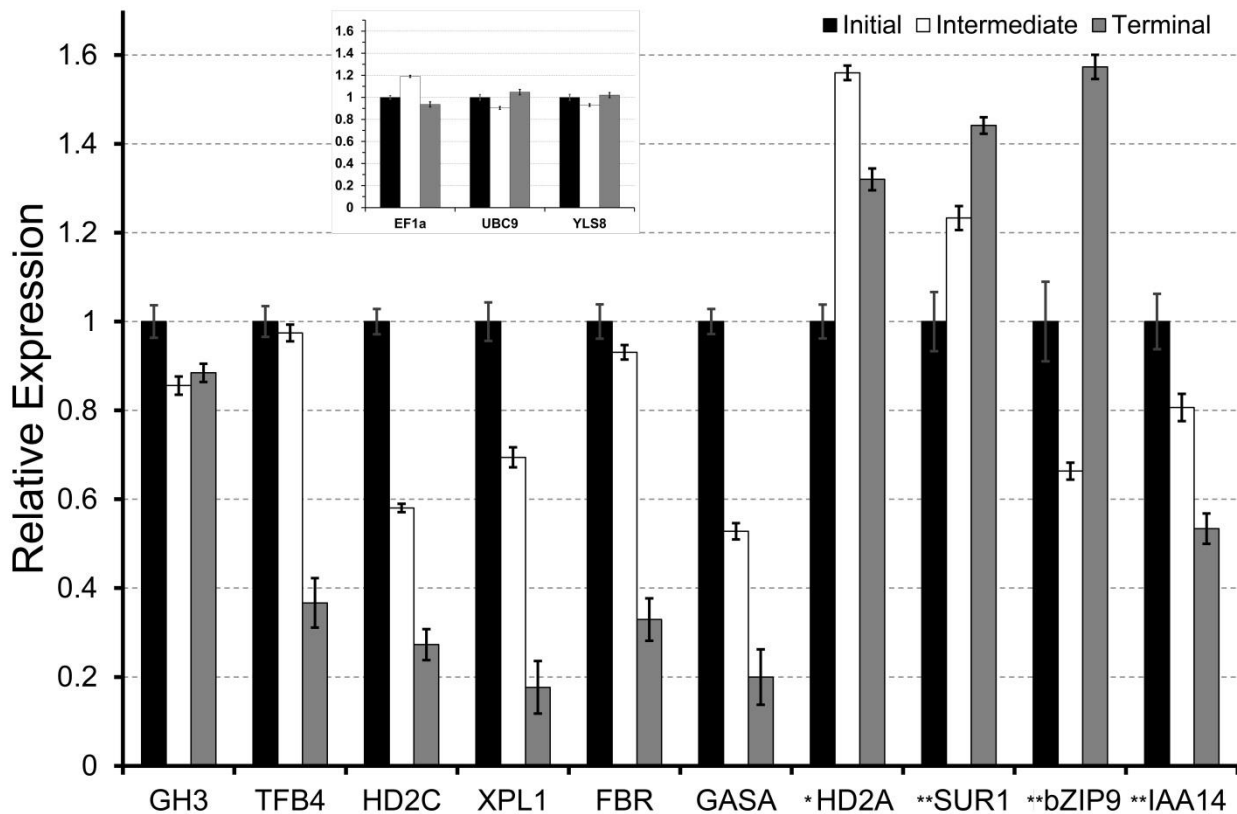
**Figura 13.** Dispersión de los valores de RPKM para los genes normalizadores candidatos identificados en *Pachycreus pringlei* a partir de los genes normalizadores superiores propuestos por Czechowski *et al.* (2005). Los transcritos que se seleccionaron como genes de referencia en este trabajo se encuentran señalados en rojo y etiquetados con su nombre.



**Figura 14.** Determinación de la eficiencia de amplificación para los oligonucleótidos utilizados para la amplificación de los genes de referencia seleccionados en este trabajo.

### V.6.3. Cuantificación de la expresión relativa de genes de interés

Anteriormente se identificaron en *S. gummosus*, mediante hibridación sustractiva, a algunos genes que se expresan de manera diferencial en el ápice de la raíz primaria cuando el RAM está activo y cuando está agotado (Shishkova *et al.*, 2013). Para explorar si los genes ortólogos de éstos en *P. pringlei* tienen regulación diferencial, y si los perfiles de expresión de los genes de *P. pringlei* son comparables al ser cuantificados mediante RNAseq y mediante RT-qPCR, se determinó la abundancia relativa del transcrito de estos genes mediante ensayos de RT-qPCR. Las reacciones de qPCR se realizaron con 50 ng de cDNA sintetizado a partir del RNA que se extrajo de muestras independientes de las etapas IN, IT y TR del desarrollo de la raíz primaria. La expresión diferencial de los ortólogos para los cinco genes que se habían identificado en *S. gummosus* fue comprobada mediante el ensayo de RT-qPCR y sus perfiles de expresión fueron congruentes con lo observado a partir de la cuantificación por RNAseq (Fig. 15). Esto muestra que los patrones de expresión en el crecimiento determinado de la raíz primaria están conservados al menos para estos genes y en la tribu Pachycereeae de la subfamilia Cactoideae, a la cual pertenecen *S. gummosus* y *P. pringlei*. Adicionalmente se evaluó la abundancia relativa de transcrito de siete genes más: los tres genes que se utilizaron como genes normalizadores: *PpEF1 $\alpha$* , *PpYLS8*, *PpUBC9*; *GRETCHEN HAGEN 3 (PpGH3)*, un gen cuya expresión es estable durante las diferentes etapas del desarrollo de la raíz primaria de *P. pringlei*; *INDOLE-3-ACETIC ACID INDUCIBLE14 (PpIAA14)*, cuya expresión disminuye hacia la etapa TR; y *BASIC LEUCINE ZIPPER9 (PpbZIP9)*, así como *SUPERROOT1 (PpSUR1)*, los cuales tienen valores de expresión más altos en la etapa TR que en la etapa IN. Con la excepción de *HISTONE DEACETYLASE 2A (PpHD2A)*, los perfiles de expresión de 12 de los 13 genes evaluados fueron reproducidos cualitativamente mediante RT-qPCR.



**Figura 15.** Abundancia relativa de distintos genes de interés de *Pachycereus pringlei*. Los niveles de expresión relativa se cuantificaron mediante ensayos de RT-qPCR y fueron normalizados con respecto a los genes *PpEF1 $\alpha$* , *PpUBC9* y *PpYLS8* (inserto). \*El perfil de expresión de HD2A determinado mediante el RNAseq no fue congruente con lo cuantificado para *PpHD2A*. \*\*La cuantificación de los perfiles de expresión de estos genes se realizó a partir de dos réplicas biológicas, mientras que para el resto de los genes se utilizaron tres réplicas biológicas. Las barras de error corresponden al error estandar.

## V.7 CONSERVACIÓN DE GENES EN LA FAMILIA CACTACEAE

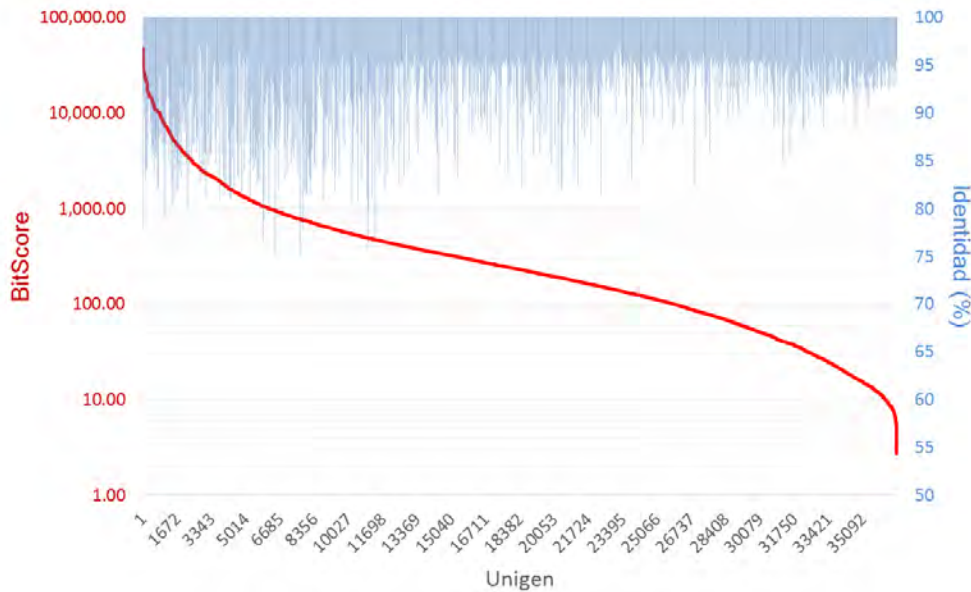
Recientemente se reportó la secuenciación por RNAseq del transcriptoma de *Lophophora williamsii* (Ibarra-Laclette *et al.*, 2015), la cual es una cactácea de amplia distribución desde el centro de México (Querétaro) hasta el norte de la República Mexicana (Coahuila y Chihuahua). El transcriptoma de *L. williamsii* se secuenció mediante RNAseq con tres plataformas distintas, GS-FLX, GS-Junior (Life Technologies) y PGM (Thermo Fisher Scientific); al utilizar tejido de raíces y “botones”

(buttons; tops of the stem (sic)) y fue ensamblado *de novo* en 40,436 contigs, los cuales están disponibles como unigenes en formato fasta en el material suplementario de la publicación original. Para comparar a las secuencias que Ibarra-Laclette *et al.* obtuvieron para *L. williamsii* con las secuencias que obtuvimos nosotros para *P. pringlei*, se interrogó a la base de datos de transcritos de *P. pringlei* al utilizar al transcriptoma completo de *L. williamsii* como sonda (Tabla VII, Fig. 16). Se observó que la mediana del valor E es de  $2.25 \times 10^{-118}$  entre las secuencias nucleotídicas de *P. pringlei* y *L. williamsii*, es decir existe un alto grado de conservación en la secuencia de los transcritos ensamblados *de novo* entre estas dos especies. Debido a que los transcritos ensamblados para *P. pringlei* son significativamente más largos que los ensamblados para *L. williamsii*, se obtuvieron en promedio 21 secuencias de *Lophophora* por cada secuencia de *Pachycereus* (Tabla VII).

**Tabla VII.** Resultados del BlastN del transcriptoma de *Lophophora williamsii*\* vs el transcriptoma de *Pachycereus pringlei*.

	Mínimo	Máximo	Promedio	Mediana
Identity	75	100	99.26	100
Bit Score	20.02	47,628.06	1,054.18	261.38
HSP*	1	270	21.6	21
E-value	0	9.75	0.02	0

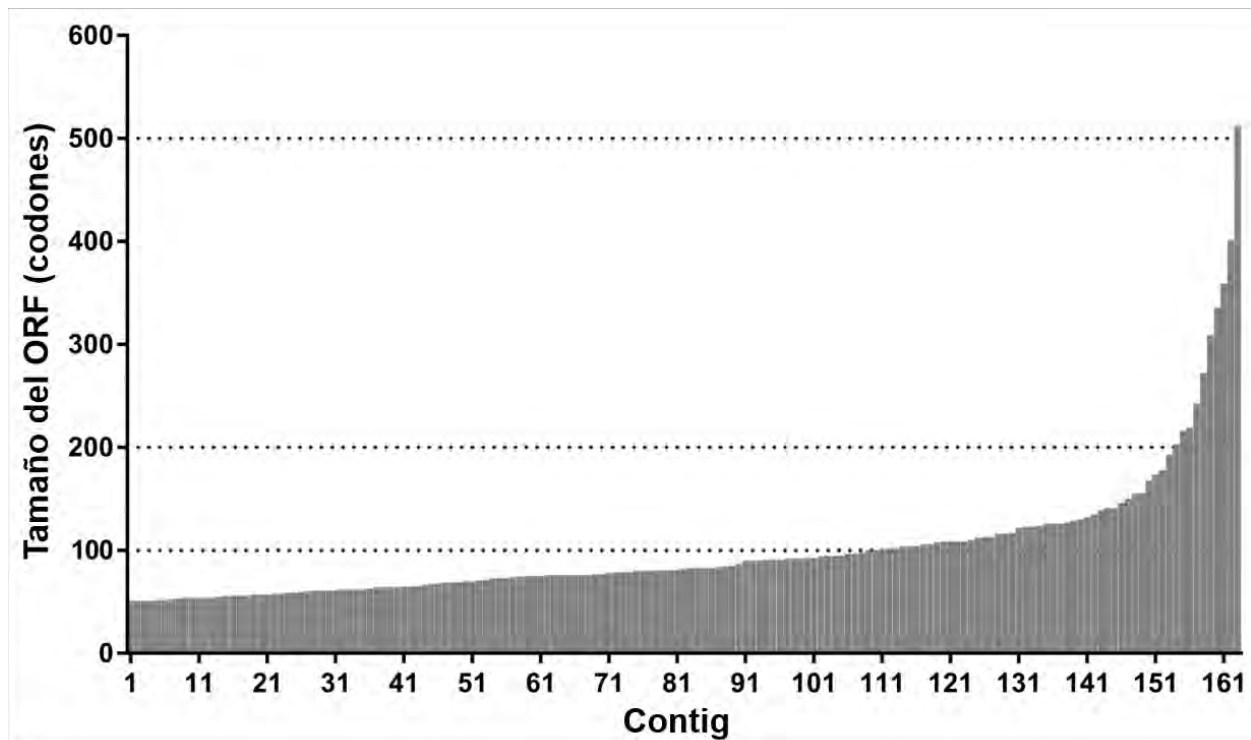
\*HSP: High-scoring segment pair; alineamiento de dos fragmentos de tamaño arbitrario cuyo "score" representa un máximo local.



**Figura 16.** Distribución de los valores de BitScores y de identidad a nivel de nucleótidos obtenidos a partir del blastN del transcriptoma de *Lophophora williamsii* vs el transcriptoma de *Pachycereus pringlei*.

Como se mencionó en la sección V.3, durante la anotación funcional de los transcritos de *P. pringlei* alrededor de 17,700 contigs no lograron anotarse con ninguna de las bases de datos o servidores que se utilizaron. De estos, 244 transcritos se identificaron como TDEs entre la etapa IN y TR, por lo cual se puede proponer que algunos de ellos pueden representar transcritos específicos de las cactáceas involucrados en la regulación del agotamiento del RAM. Para explorar esta hipótesis realizamos una búsqueda mediante BlastN de los TDEs no anotados de *P. pringlei* contra el transcriptoma de *L. williamsii*. Se encontró que 130 de los TDEs no anotados tienen al menos una secuencia homóloga en el transcriptoma de *Lophophora*. Además, se encontraron secuencias homólogas de algunos de estos TDEs en el transcriptoma de la zona cambial de otras especies de cactáceas, específicamente en *Ariocarpus retusus*, *Ferocactus pilosus*, *Pereskia lychnidiflora* y *Echinocactus platyacanthus* (Petroné Mendoza, 2018). 163 de los 244 TDEs no anotados contienen ORFs que pueden ser traducidos en proteínas putativas con más de 50 aminoácidos. De éstos, 10, 45 y 108 transcritos contienen ORFs que se podrían traducir a proteínas mayores a 200 aa, entre 100-200 aa, y de 50-100 aa, respectivamente (Fig. 17). Esta observación, en conjunto con la conservación de estas secuencias en diferentes especies de

cactáceas, así como su regulación diferencial en *P. pringlei* durante el desarrollo de la raíz primaria, sugiere la existencia de reguladores no descritos en el desarrollo de la raíz y propios de la familia Cactaceae. Esta hipótesis requiere una exploración más minuciosa pues abre nuevas oportunidades de investigación, pero sólo será posible en la medida que se establezcan estrategias para el análisis funcional de genes en las plantas cactaceas.



**Figura 17.** Distribución de la longitud de los marco abierto de lectura putativos identificados en los transcritos con expresión diferencial que no pudieron ser anotados.

## V.8 RED DE REGULACIÓN TRANSCRIPCIONAL PARA EL ÁPICE DE LA RAÍZ DE *Pachycereus pringlei*

Los procesos celulares que ocurren durante el desarrollo o en respuesta a los diferentes estímulos ambientales no están controlados por genes aislados, sino por una red que conecta a múltiples elementos, incluyendo a los reguladores transcripcionales y sus blancos (Istrail *et al.*, 2007). Con el objetivo de identificar a los principales mecanismos que controlan al crecimiento determinado de la raíz primaria de *P. pringlei*,

integramos a algunos de los transcritos ensamblados en redes de regulación genética. Debido a que *P. pringlei* no es un organismo modelo y que la información genética sobre cactáceas es escasa, estos análisis se realizaron a partir de su confrontación con datos generados para *A. thaliana* (Azpeitia *et al.*, 2013; Jin *et al.*, 2015). Inferimos una red de regulación transcripcional para *P. pringlei* a partir de la literatura disponible y del mapa de regulación genética de *Arabidopsis* (Jin *et al.*, 2015; <http://atrm.cbi.pku.edu.cn/>). Inicialmente se identificó en la base de datos de *P. pringlei* a algunos probables ortólogos de factores transcripcionales previamente descritos en la literatura como reguladores importantes para el desarrollo de la raíz o para el mantenimiento del RAM. La lista de reguladores se expandió al incluir a 130 genes adicionales cuya expresión está aumentada y conservada en la zona meristemática de la raíz primaria de las angiospermas (Huang and Schiefelbein, 2015). De esta lista, que incluye a 260 genes en total, se identificaron a 81 genes que funcionan como reguladores transcripcionales, o cuyo regulador transcripcional está contenido en el mapa de regulación de *Arabidopsis*. Para cada uno de los genes que se obtuvieron se buscó a su ortólogo putativo en el transcriptoma de *P. pringlei* mediante la estrategia de BBH al utilizar el algoritmo de tBlastN; o bien, de hit unidireccional cuando el BBH no fue exitoso. Se generó una tabla de interacciones en donde cada fila incluye tres columnas: el nodo fuente, la direccionalidad y tipo de la interacción, y el nodo destino. El tipo y dirección de las interacciones se estableció con base en las interacciones descritas para *A. thaliana*. Adicionalmente se generó una tabla de propiedades en donde se integran las características de cada nodo de acuerdo con los resultados que se obtuvieron del transcriptoma, por ejemplo, el nivel de expresión del nodo de manera global y por etapas; y el nombre de su ortólogo en *A. thaliana*, entre otros. Las tablas de interacción y de propiedades, las cuales se pueden consultar en el Anexo 6, fueron cargadas y visualizadas en el software Cytoscape v3.2.1. De este modo inferimos una red que incluye 81 nodos y 111 arcos, o interacciones (Fig. 18). Para visualizar a la red se especificó el tamaño de cada nodo de forma proporcional al nivel de expresión ( $\text{Log}_2(\text{RPKM})$ ) del transcrito, y con anaranjado y azul a la expresión relativa en las etapas IN y TR, respectivamente. 48 de los 81 nodos corresponden a ortólogos putativos identificados por BBH, mientras que 30 nodos se recuperaron como hits

unidireccionales. Sólo para tres nodos no se encontraron ortólogos en nuestro transcriptoma, éstos son YAB1, FEZ y LFY, probablemente debido a su baja o nula expresión en el ápice de la raíz de *P. pringlei*. Se observa que los nodos que corresponden a factores transcripcionales tienen un tamaño menor que el resto de los nodos en la red (ver por ejemplo el nodo MYB26 en comparación con el nodo HD2B); esto es un resultado esperado, pues los factores transcripcionales no requieren niveles de expresión altos para ejercer su función; en cambio, otras proteínas cuya función es estructural o enzimática requieren niveles de expresión más altos para mantener su función en la célula. La topología de la red que se muestra en la figura 18 permite observar que existe modularidad en la regulación genética del crecimiento de la raíz. Se pueden apreciar, por ejemplo, clusters que integran señales mediadas por hormonas, tales como las auxinas, y que incluyen a nodos con alta conectividad como los ARF.

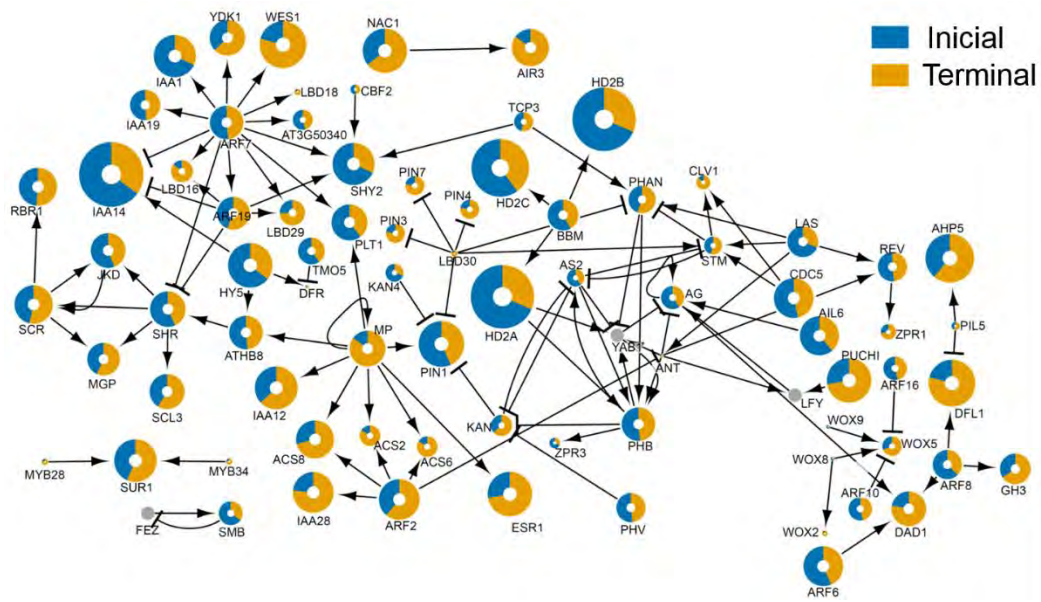
Para facilitar el análisis de la red, se extrajeron tres módulos a partir de ella, los cuales incluyen a nodos representantes de cada una de las tres vías de regulación principales reportadas para *A. thaliana*; es decir, a la vía mediada por los factores de transcripción PLT; la vía que incluye a los reguladores transcripcionales SHR y SCR; y la vía mediada por WOX5 (Fig. 19). En la red inferida para *P. pringlei*, el módulo PpPLT está representado en por los nodos PpPLT1 y PpBBM, e incluye además a su interacción con las proteínas ARF y algunas desacetilasas de histonas (Fig. 19a). Al momento de inferir la red de regulación transcripcional, PLT1 y BBM eran las únicas proteínas PLT para las cuales se habían identificado con seguridad a sus reguladores transcripcionales. La reciente identificación de algunos blancos específicos para cada una de los factores de transcripción PLT (Santuari *et al.*, 2016) proporciona información valiosa que permitirá la inclusión de nodos adicionales a este módulo en la red de regulación inferida para *P. pringlei*.

El segundo módulo (Fig. 19b) incluye a los cofactores transcripcionales PpSCR y PpSHR, de la familia GRAS, y a otros interactores tales como PpJKD. Recientemente fue reportado que SHR, SCR y JKD de *Arabidopsis* interactúan para formar un complejo macromolecular que regula la transcripción de diferentes genes de acuerdo

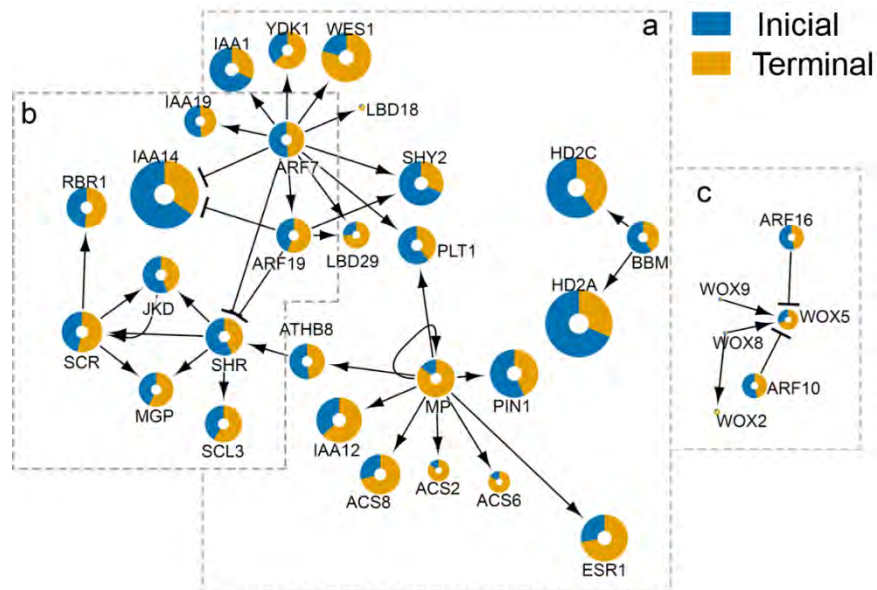


con su contexto celular (Hirano *et al.*, 2017; Long *et al.*, 2017). De este modo, aunque los nodos PpSCR, PpSHR y PpJKD parecen tener una expresión relativamente estable en el ápice de la raíz primaria de *P. pringlei*, esto podría deberse a la participación de los genes PpSCR-PpSHR en procesos adicionales al mantenimiento de la identidad del CQ y del crecimiento de la raíz; por ejemplo, al establecimiento del patrón radial de la raíz (Helariutta *et al.*, 2000), la diferenciación del tejido vascular (Nakajima *et al.*, 2001) y la transición de las células de la zona de elongación a la zona de diferenciación (Moubayidin *et al.*, 2013). El módulo PpSHR-PpSCR también incluye a otros interactores, tales como PpMGP y PpRBR, los cuales se han identificado como reguladores importantes del mantenimiento del RAM (Cruz-Ramírez *et al.*, 2012; Welch *et al.*, 2007b). El tercer módulo de regulación incluye a las proteínas WOX y a algunos de sus interactores, tales como los ARF (Fig. 19c). Aunque fue posible identificar transcritos de *P. pringlei* que codifican para proteínas WOX putativas, debido a la alta similitud entre los miembros de esta familia no fue posible asignar a un ortólogo putativo mediante la técnica de BBH. Aunado a esto, los niveles de expresión de los transcritos de *PpWOX* son muy bajos, con valores de RPKM <4 en el transcriptoma de *P. pringlei*. Estos valores son congruentes con lo esperado, pues WOX5 por ejemplo, sólo se expresa en las células del CQ, mientras que en los ápices colectados para las muestras de la etapa IN las células del CQ representan una mínima proporción de las células de la zona meristemática y en la etapa intermedia el CQ desaparece.

Estos resultados sugieren que los programas de regulación transcripcional que operan en el ápice de la raíz primaria de *P. pringlei* son similares a los programas de desarrollo de la raíz primaria de otras angiospermas, aunque las diferencias en los niveles de expresión de los transcritos de Cactáceas diverge de lo observado en el ápice de otras plantas. En *Arabidopsis*, por ejemplo, los genes *PLT* se mantienen con un máximo de expresión en el RAM, mientras que en *P. pringlei* el nivel de expresión de *PpPLT1* y *PpBBM*, así como de otros miembros de la familia *PLT*, decrece en el ápice de la raíz en la etapa terminal.



**Figura 18.** Red de regulación transcripcional inferida para el ápice de la raíz primaria de *Pachycereus pringlei*. El tamaño de los nodos es proporcional a la suma de los valores RPKM del transcrito en las tres etapas del desarrollo ( $\text{Log}_2(\text{RPKM})$ ); La abundancia relativa de cada transcrito en la etapa inicial y terminal se indica con los colores azul y anaranjado, respectivamente. Los nodos de color gris son aquellos para los cuales no se encontraron secuencias en el transcriptoma de *P. pringlei*.



**Figura 19.** La red de regulación transcripcional inferida para *Pachycereus pringlei* incluye a las tres vías más importantes del desarrollo de la raíz descritas para *Arabidopsis thaliana*: (a) el módulo PpPLT, el cual incluye a PpPLT1 y PpBBM como clusters separados; (b) PpSHR-PpSCR; y (c) PpWOX. El tamaño de los nodos es proporcional a la suma de los valores de RPKM del transcrito en las tres etapas del desarrollo ( $\text{Log}_2(\text{RPKM})$ ); La abundancia relativa de cada transcrito en la etapa inicial y terminal se indica con los colores azul y anaranjado, respectivamente.

## VI. CONCLUSIONES

Como parte de este trabajo secuenciamos y ensamblamos *de novo* el transcriptoma del ápice de la raíz primaria de *Pachycereus pringlei* en las etapas inicial, intermedia y terminal del desarrollo. El transcriptoma ensamblado incluye 49,045 transcritos, de los cuales fue posible anotar al 63.7%. La anotación de los transcritos así como su mapeo en distintas rutas metabólicas, de señalización y en diferentes procesos celulares indica que el transcriptoma es robusto y confiable, además de amplio.

A partir del análisis de expresión diferencial de los transcritos en las diferentes etapas del desarrollo de la raíz primaria de *P. pringlei*, encontramos 873 y 1,388 transcritos diferencialmente expresados al comparar a las etapas IN vs TR e IT vs TR, respectivamente. Entre los TDEs, el número de transcritos que se inducen hacia la etapa TR es mayor que el número de transcritos que se reprimen, lo cual corrobora la noción general de que las células no diferenciadas expresan un conjunto menor de genes que las células diferenciadas. El análisis hipergeométrico mostró que entre los transcritos más abundantes en la etapa IN, es decir, cuando las células conservan su actividad proliferativa, se encuentra una sobrerrepresentación de transcritos asociados con el ciclo celular, síntesis del DNA, traducción de proteínas y biogénesis del ribosoma, así como el ensamblaje de nucleosomas, entre otros. Por otro lado, las células en la etapa TR expresan una mayor cantidad de genes relacionados con el metabolismo, principalmente con la síntesis y transporte de aminoácidos y otros compuestos. De manera interesante, en la etapa TR también existe una sobrerrepresentación de algunos factores transcripcionales de la familia WRKY y de la familia NAC, los cuales están involucrados en las respuestas a distintos tipos de estrés, particularmente al estrés por déficit hídrico. Esta observación sugiere que *P. pringlei* monta de manera constitutiva los mecanismos moleculares para contender con la sequía que caracteriza a las zonas áridas a las cuales está adaptado, y que esta respuesta se puede observar de manera concomitante con la progresión del agotamiento del RAM de la raíz primaria.

Respecto a los programas transcripcionales que operan en el ápice de la raíz primaria en distintas etapas de desarrollo, observamos que en general estos son comparables a aquellos que ocurren en el eje longitudinal de la raíz de las angiospermas; es decir, que el programa genético de agotamiento del meristemo conduce a la diferenciación de las células del ápice de la raíz mediante los programas de diferenciación que se observan en otras angiospermas, cuando las células de la raíz pasan por las zonas meristemática, de elongación y de diferenciación. Sin embargo, dado que la divergencia en la secuencia de aminoácidos entre las proteínas codificadas por los transcritos de *P. pringlei* y los de otras angiospermas es baja, aunado a la variación en la abundancia de los transcritos en las etapas del desarrollo de la raíz primaria de *P. pringlei*, es probable que la adquisición del crecimiento determinado como carácter adaptativo a las zonas áridas o semi áridas se deba a la divergencia en las secuencias promotoras de los genes que constituyen nodos en la red de regulación transcripcional del desarrollo de la raíz primaria. Esta hipótesis debe ser explorada de manera más exhaustiva.

A partir del análisis comparativo entre las secuencias ensambladas para el transcriptoma de *P. pringlei* y de *L. williamsii*, demostramos que existe un alto grado de conservación de secuencias a nivel de nucleótidos entre estas dos especies. Durante la anotación de los transcritos un porcentaje significativo de secuencias no pudieron anotarse. Algunos de los transcritos no anotados se expresan diferencialmente entre las etapas IN y TR del desarrollo de la raíz primaria de *P. pringlei*. La identificación de secuencias homólogas en *L. williamsii* de los TDEs no anotados de *P. pringlei* sugiere la existencia de transcritos linaje específicos de la familia Cactaceae, algunos de los cuales pudieran estar involucrados en la regulación del agotamiento del RAM de la raíz primaria. Esta información resulta valiosa, pues permite la exploración de reguladores del desarrollo de la raíz no descritos hasta el momento, abriendo oportunidades interesantes para entender el origen de mecanismos de desarrollo como producto de la evolución y adaptación de las especies a sus nichos ecológicos. La secuenciación y el ensamblaje, muy recientemente, del genoma nuclear de *Carnegiea gigantea* (Cactaceae), y la liberación de los "draft genomes" de baja cobertura y aun muy fragmentados de tres especies de cactáceas columnares adicionales, entre ellas *P.*

*pringlei* (Copetti *et al.*, 2017), permitirá una mayor profundidad en la exploración de las hipótesis generadas con la conclusión de este trabajo.

La presente tesis permitió la escritura y publicación de dos artículos científicos, los cuales pueden consultarse en su versión en línea (DOI: 10.1038/s41598-018-26897-1 y 10.3389/fpls.2017.00218) y en el Anexo 7.

## VII. PERSPECTIVAS.

Los datos generados como parte de esta tesis permitirán, en combinación con la reciente liberación de un genoma de referencia para la familia Cactaceae, extraer información adicional de relevancia para el estudio del agotamiento del RAM en las cactáceas. Entre las directrices a corto y mediano plazo que se desprenden de este trabajo, podemos mencionar:

- \* Extender el análisis con base en las anotaciones de los transcritos. Particularmente, identificar a aquellos transcritos que fueron anotados únicamente mediante InterPro Scan (por contener firmas proteicas) pero que no tienen homólogos entre las secuencias provenientes del clado Viridiplantae. Estas secuencias, aunque se consideraron como “anotadas” en este trabajo, podrían también representar transcritos linaje específicos.
- \* Explorar familias génicas de interés; por ejemplo, extender el análisis de expresión y de divergencia evolutiva de los factores de transcripción PLT.
- \* Integrar, en la medida de lo posible, los datos generados con nuestro transcriptoma y el genoma preliminar de *P. pringlei* (Copetti *et al.*, 2017) para explorar, por ejemplo, la existencia de variantes de “splicing”, o empalme, alternativo.
- \* Explorar la posible expansión y reducción de familias génicas, así como la neofuncionalización o subfuncionalización de genes en el contexto de la tetraploidía de *P. pringlei*.

## VIII. BIBLIOGRAFÍA

Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nussaume, L., Noh, Y.-S., Amasino, R., and Scheres, B. (2004). The *PLETHORA* Genes Mediate Patterning of the *Arabidopsis* Root Stem Cell Niche. *Cell* *119*, 109–120.

Azpeitia, E., Weinstein, N., Benítez, M., Mendoza, L., and Alvarez-Buylla, E.R. (2013). Finding Missing Interactions of the *Arabidopsis thaliana* Root Stem Cell Niche Gene Regulatory Network. *Front. Plant Sci.* *4*.

Benfey, P.N., Linstead, P.J., Roberts, K., Schiefelbein, J.W., Hauser, M.-T., and Aeschbacher, R.A. (1993). Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* *119*, 57–70.

Berg, C. van den, Willemsen, V., Hage, W., Weisbeek, P., and Scheres, B. (1995). Cell fate in the *Arabidopsis* root meristem determined by directional signalling. *Nature* *378*, 62–65.

Boke, N.H. (1989). Root Glochids and Root Spurs of *Opuntia arenaria* (Cactaceae). *American Journal of Botany* *66*, 1085–1092.

Bolle, C. (2004). The role of GRAS proteins in plant signal transduction and development. *Planta* *218*, 683–692.

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. (1997). Inflorescence Commitment and Architecture in *Arabidopsis*. *Science* *275*, 80–83.

Cattell, R.B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research* *1*, 245–276.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* *21*, 3674–3676.

Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J.L.M., Childs, K.L., Eguiarte, L.E., Lee, S., Liu, T.L., McMahon, M.M., Whiteman, N.K., *et al.* (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *PNAS* *114*, 12003–12008.

Cruz-Ramírez, A., Díaz-Triviño, S., Blilou, I., Grieneisen, V.A., Sozzani, R., Zamioudis, C., Miskolczi, P., Nieuwland, J., Benjamins, R., Dhonukshe, P., *et al.* (2012). A Bistable Circuit Involving SCARECROW-RETINOBLASTOMA Integrates Cues to Inform Asymmetric Stem Cell Division. *Cell* *150*, 1002–1015.

Cruz-Ramírez, A., Díaz-Triviño, S., Wachsman, G., Du, Y., Arteága-Vázquez, M., Zhang, H., Benjamins, R., Blilou, I., Neef, A.B., Chandler, V., *et al.* (2013). A SCARECROW-RETINOBLASTOMA Protein Network Controls Protective Quiescence in the *Arabidopsis* Root Stem Cell Organizer. *PLOS Biology* *11*, e1001724.

Cui, H., Levesque, M.P., Vernoux, T., Jung, J.W., Paquette, A.J., Gallagher, K.L., Wang, J.Y., Blilou, I., Scheres, B., and Benfey, P.N. (2007). An Evolutionarily Conserved Mechanism

Delimiting SHR Movement Defines a Single Layer of Endodermis in Plants. *Science* 316, 421–425.

Czechowski, T., Stitt M., Altmann T., Udvardi MK., Scheible WR. (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*. *PLANT PHYSIOLOGY* 139, 5–17.

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A., and Benfey, P.N. (1996). The *SCARECROW* Gene Regulates an Asymmetric Cell Division That Is Essential for Generating the Radial Organization of the *Arabidopsis* Root. *Cell* 86, 423–433.

Dietz, K.-J., Vogel, M.O., and Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma* 245, 3–14.

Dinkelaker, B., Hengeler, C., and Marschner, H. (1995). Distribution and Function of Proteoid Roots and other Root Clusters. *Botanica Acta* 108, 183–200.

Doerner, P. (1998). Root development: Quiescent center not so mute after all. *Current Biology* 8, R42–R44.

Dolan, L., Janmaat, K., Willemsen, V., Linstead, P., Poethig, S., Roberts, K., and Scheres, B. (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., and 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., and Rost, T.L. (2003). Lateral root initiation. *Encyclopedia of Applied Plant Sciences* (Thomas B., Murphy D.J., and Murray B.G. Academic Press).

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* 23, 257–258.

Feldman, L.J. (1976). The *de novo* origin of the quiescent center regenerating root apices of *Zea mays*. *Planta* 128, 207–212.

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

Gallagher, K.L., and Benfey, P.N. (2009). Both the conserved GRAS domain and nuclear localization are required for SHORT-ROOT movement. *The Plant Journal* 57, 785–797.

Ge, L., Chen, H., Jiang, J.-F., Zhao, Y., Xu, M.-L., Xu, Y.-Y., Tan, K., Xu, Z.-H., and Chong, K. (2004). Overexpression of *OsRAA1* Causes Pleiotropic Phenotypes in Transgenic Rice Plants,

including Altered Leaf, Flower, and Root Development and Root Response to Gravity. *Plant Physiology* 135, 1502–1513.

Gehrig, H.H., Winter, K., Cushman, J., Borland, A., and Taybi, T. (2000). An improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides. *Plant Mol Biol Rep* 18, 369–376.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29, 644–652.

Groot, E.P., Sweeney, E.J., and Rost, T.L. (2003). Development of the adhesive pad on climbing fig (*Ficus pumila*) stems from clusters of adventitious roots. *Plant and Soil* 248, 85–96.

Hanano, S., and Goto, K. (2011). *Arabidopsis* *TERMINAL FLOWER1* Is Involved in the Regulation of Flowering Time and Inflorescence Development through Transcriptional Repression. *The Plant Cell* 23, 3172–3184.

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.-T., and Benfey, P.N. (2000). The *SHORT-ROOT* Gene Controls Radial Patterning of the *Arabidopsis* Root through Radial Signaling. *Cell* 101, 555–567.

Hirano, Y., Nakagawa, M., Suyama, T., Murase, K., Shirakawa, M., Takayama, S., Sun, T., and Hakoshima, T. (2017). Structure of the SHR–SCR heterodimer bound to the BIRD/IDD transcriptional factor JKD. *Nature Plants* 3, 17010.

Huang, L., and Schiefelbein, J. (2015). Conserved Gene Expression Programs in Developing Roots from Diverse Plants. *The Plant Cell* 27, 2119–2132.

Ibarra-Laclette, E., Zamudio-Hernández, F., Pérez-Torres, C.A., Albert, V.A., Ramírez-Chávez, E., Molina-Torres, J., Fernández-Cortés, A., Calderón-Vázquez, C., Olivares-Romero, J.L., Herrera-Estrella, A., *et al.* (2015). *De novo* sequencing and analysis of *Lophophora williamsii* transcriptome, and searching for putative genes involved in mescaline biosynthesis. *BMC Genomics* 16, 657.

Imbeaud, S., Gaudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., Mueller, O., Schroeder, A., and Auffray, C. (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res* 33, e56–e56.

Iorizzo, M., Senalik, D.A., Grzebelus, D., Bowman, M., Cavagnaro, P.F., Matvienko, M., Ashrafi, H., Van Deynze, A., and Simon, P.W. (2011). *De novo* assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics* 12, 389.

Istrail, S., De-Leon, S.B.-T., and Davidson, E.H. (2007). The regulatory genome and the computer. *Developmental Biology* 310, 187–195.

Jeong, S., Trotochaud, A.E., and Clark, S.E. (1999). The *Arabidopsis* *CLAVATA2* Gene Encodes a Receptor-like Protein Required for the Stability of the *CLAVATA1* Receptor-like Kinase. *The Plant Cell* 11, 1925–1933.



- Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J., and Gao, G. (2015). An *Arabidopsis* Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. *Molecular Biology and Evolution* 32, 1767–1773.
- Jofuku, K.D., Boer, B.G. den, Montagu, M.V., and Okamoto, J.K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *The Plant Cell* 6, 1211–1225.
- Johnson, M.T.J., Carpenter, E.J., Tian, Z., Bruskiwich, R., Burris, J.N., Carrigan, C.T., Chase, M.W., Clarke, N.D., Covshoff, S., dePamphilis, C.W., *et al.* (2012). Evaluating Methods for Isolating Total RNA and Predicting the Success of Sequencing Phylogenetically Diverse Plant Transcriptomes. *PLOS ONE* 7, e50226.
- Johnsson, P., Lipovich, L., Grandér, D., and Morris, K.V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1840, 1063–1071.
- Kamiya, N., Nagasaki, H., Morikami, A., Sato, Y., and Matsuoka, M. (2003). Isolation and characterization of a rice *WUSCHEL*-type homeobox gene that is specifically expressed in the central cells of a quiescent center in the root apical meristem. *The Plant Journal* 35, 429–441.
- Koenig, D., Jiménez-Gómez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., *et al.* (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *PNAS* 110, E2655–E2662.
- Krizek, B.A. (2011). Auxin regulation of *Arabidopsis* flower development involves members of the AINTEGUMENTA-LIKE/PLETHORA (AIL/PLT) family. *J Exp Bot* 62, 3311–3319.
- Kumpf, R.P., and Nowack, M.K. (2015). The root cap: a short story of life and death. *J Exp Bot* 66, 5651–5662.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., *et al.* (2012). The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* 40, D1202–D1210.
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 129–137.
- Long, Y., Smet, W., Cruz-Ramírez, A., Castelijns, B., Jonge, W. de, Mähönen, A.P., Bouchet, B.P., Perez, G.S., Akhmanova, A., Scheres, B., *et al.* (2015). *Arabidopsis* BIRD Zinc Finger Proteins Jointly Stabilize Tissue Boundaries by Confining the Cell Fate Regulator SHORT-ROOT and Contributing to Fate Specification. *The Plant Cell* 27, 1185–1199.
- Long, Y., Stahl, Y., Weidtkamp-Peters, S., Postma, M., Zhou, W., Goedhart, J., Sánchez-Pérez, M.-I., Gadella, T.W.J., Simon, R., Scheres, B., *et al.* (2017). *In vivo* FRET–FLIM reveals cell-type-specific protein interactions in *Arabidopsis* roots. *Nature* 548, 97–102.
- López-Valle, M.L. (2015). Análisis molecular de los genes *PpPLT* en el desarrollo de la raíz primaria del cardón *Pachycereus pringlei*. Tesis de maestría. Universidad Nacional Autónoma de México.

- Mähönen, A.P., Tusscher, K. ten, Siligato, R., Smetana, O., Díaz-Triviño, S., Salojärvi, J., Wachsman, G., Prasad, K., Heidstra, R., and Scheres, B. (2014). PLETHORA gradient formation mechanism separates auxin responses. *Nature* 515, 125–129.
- Mallona, I., Egea-Cortines, M., and Weiss, J. (2011). Conserved and Divergent Rhythms of Crassulacean Acid Metabolism-Related and Core Clock Gene Expression in the Cactus *Opuntia ficus-indica*. *Plant Physiology* 156, 1978–1989.
- Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., *et al.* (2014). Data access for the 1,000 Plants (1KP) project. *GigaScience* 3, 17.
- Mayer, K.F.X., Schoof, H., Haecker, A., Lenhard, M., Jürgens, G., and Laux, T. (1998). Role of WUSCHEL in Regulating Stem Cell Fate in the *Arabidopsis* Shoot Meristem. *Cell* 95, 805–815.
- Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., Hu, H., Khaitovich, P., and Kaessmann, H. (2013). Birth and expression evolution of mammalian microRNA genes. *Genome Res.* 23, 34–45.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35, W182–W185.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5, 621–628.
- Moubayidin, L., Di Mambro, R., Sozzani, R., Pacifici, E., Salvi, E., Terpstra, I., Bao, D., van Dijken, A., Dello Iorio, R., Perilli, S., *et al.* (2013). Spatial Coordination between Stem Cell Activity and Cell Differentiation in the Root Meristem. *Developmental Cell* 26, 405–415.
- Murawski, D.A., Fleming, T.H., Ritland, K., and Hamrick, J.L. (1994). Mating system of *Pachycereus pringlei*: an autotetraploid cactus. *Heredity* 72, 86–94.
- Nakajima, K., Sena, G., Nawy, T., and Benfey, P.N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* 413, 307–311.
- Okamoto, J.K., Caster, B., Villarroel, R., Montagu, M.V., and Jofuku, K.D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *PNAS* 94, 7076–7081.
- Olsen, A.N., Ernst, H.A., Leggio, L.L., and Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. *Trends in Plant Science* 10, 79–87.
- Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. *Genome Biology* 11, 220.
- Perilli, S., Perez-Perez, J.M., Mambro, R.D., Peris, C.L., Díaz-Triviño, S., Bianco, M.D., Pierdonati, E., Moubayidin, L., Cruz-Ramírez, A., Costantino, P., *et al.* (2013). RETINOBLASTOMA-RELATED Protein Stimulates Cell Differentiation in the *Arabidopsis* Root Meristem by Interacting with Cytokinin Signaling. *The Plant Cell* 25, 4469–4478.

- Petrone Mendoza, E. (2018). Inferencias funcionales de la madera dimórfica mediante el análisis transcriptómico en Cactaceae. Tesis de Maestría. Universidad Nacional Autónoma de México.
- Prasad, K., Grigg, S.P., Barkoulas, M., Yadav, R.K., Sanchez-Perez, G.F., Pinon, V., Blilou, I., Hofhuis, H., Dhonukshe, P., Galinha, C., *et al.* (2011). *Arabidopsis* PLETHORA Transcription Factors Control Phyllotaxis. *Current Biology* 21, 1123–1128.
- Qingzhu, H., Chengjie, C., Zhe, C., Pengkun, C., Yuewen, M., Jingyu, W., Jian, Z., Guibing, H., Jietang, Z., and Yonghua, Q. (2016). Transcriptomic Analysis Reveals Key Genes Related to Betalain Biosynthesis in Pulp Coloration of *Hylocereus polyrhizus*. *Front. Plant Sci.* 6.
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res* 33, W116–W120.
- Riechmann, J.L., and Meyerowitz, E.M. (1998). The AP2/EREBP family of plant transcription factors. *Biol chem* 379, 633 - 646.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rodríguez-Rodríguez, J.F., Shishkova, S., Napsucialy-Mendivil, S., and 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.
- Rushton, P.J., Somssich, I.E., Ringler, P., and Shen, Q.J. (2010). WRKY transcription factors. *Trends in Plant Science* 15, 247–258.
- Sabatini, S., Heidstra, R., Wildwater, M., and Scheres, B. (2003). SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev.* 17, 354–358.
- Sánchez-Calderón, L., López-Bucio, J., Chacón-López, A., Cruz-Ramírez, A., Nieto-Jacobo, F., Dubrovsky, J.G., and Herrera-Estrella, L. (2005). Phosphate Starvation Induces a Determinate Developmental Program in the Roots of *Arabidopsis thaliana*. *Plant Cell Physiol* 46, 174–184.
- Santuari, L., Sanchez-Perez, G.F., Luijten, M., Rutjens, B., Terpstra, I., Berke, L., Gorte, M., Prasad, K., Bao, D., Timmermans-Hereijgers, J.L.P.M., *et al.* (2016). The PLETHORA Gene Regulatory Network Guides Growth and Cell Differentiation in *Arabidopsis* Roots. *The Plant Cell* 28, 2937–2951.
- Sarkar, A.K., Luijten, M., Miyashima, S., Lenhard, M., Hashimoto, T., Nakajima, K., Scheres, B., Heidstra, R., and Laux, T. (2007). Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* 446, 811–814.
- Scheres, B., Benfey, P., and Dolan, L. (2002). Root Development. *The Arabidopsis Book* e0101.
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* 7, 3.

- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* *13*, 2498–2504.
- Shishkova, S., and Dubrovsky, J.G. (2005). Developmental programmed cell death in primary roots of Sonoran Desert Cactaceae. *American Journal of Botany* *92*, 1590–1594.
- Shishkova, S., García-Mendoza, E., Castillo-Díaz, V., Moreno, N.E., Arellano, J., and Dubrovsky, J.G. (2007). Regeneration of roots from callus reveals stability of the developmental program for determinate root growth in Sonoran Desert Cactaceae. *Plant Cell Rep* *26*, 547–557.
- Shishkova, S., Rost, T.L., and Dubrovsky, J.G. (2008). Determinate Root Growth and Meristem Maintenance in Angiosperms. *Ann Bot* *101*, 319–340.
- Shishkova, S., Peñas, L., Laura, M., Napsucialy-Mendivil, S., Matvienko, M., Kozik, A., Montiel, J., Patiño, A., and 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. *Ann Bot* *112*, 239–252.
- Stahl, Y., and 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.
- Stahl, Y., Wink, R.H., Ingram, G.C., and Simon, R. (2009). A Signaling Module Controlling the Stem Cell Niche in *Arabidopsis* Root Meristems. *Current Biology* *19*, 909–914.
- Stone, J.M., Trotochaud, A.E., Walker, J.C., and Clark, S.E. (1998). Control of Meristem Development by CLAVATA1 Receptor Kinase and Kinase-Associated Protein Phosphatase Interactions. *Plant Physiology* *117*, 1217–1225.
- Taiz, L., and Zeiger, E. (2010). *Plant Physiology* (Sinauer Associates, Inc).
- Takahashi, H., Kamakura, H., Sato, Y., Shiono, K., Abiko, T., Tsutsumi, N., Nagamura, Y., Nishizawa, N.K., and Nakazono, M. (2010). A method for obtaining high quality RNA from paraffin sections of plant tissues by laser microdissection. *J Plant Res* *123*, 807–813.
- Trotochaud, A.E., Jeong, S., and Clark, S.E. (2000). CLAVATA3, a Multimeric Ligand for the CLAVATA1 Receptor-Kinase. *Science* *289*, 613–617.
- Varney, G.T., and McCully, M.E. (1991). The branch roots of *Zea*. II. Developmental loss of the apical meristem in field-grown roots. *New Phytologist* *118*, 535–546.
- Watt, M., Silk, W.K., and Passioura, J.B. (2006). Rates of Root and Organism Growth, Soil Conditions, and Temporal and Spatial Development of the Rhizosphere. *Ann Bot* *97*, 839–855.
- Welch, D., Hassan, H., Blilou, I., Immink, R., Heidstra, R., and Scheres, B. (2007a). *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes Dev.* *21*, 2196–2204.
- Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res* *37*, D380–D386.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLOS ONE* 2, e718.

Yang, Y., Moore, M.J., Brockington, S.F., Soltis, D.E., Wong, G.K.-S., Carpenter, E.J., Zhang, Y., Chen, L., Yan, Z., Xie, Y., *et al.* (2015). Dissecting Molecular Evolution in the Highly Diverse Plant Clade Caryophyllales Using Transcriptome Sequencing. *Mol Biol Evol* 32, 2001–2014.

Zenoni, S., Ferrarini, A., Giacomelli, E., Xumerle, L., Fasoli, M., Malerba, G., Bellin, D., Pezzotti, M., and Delledonne, M. (2010). Characterization of Transcriptional Complexity during Berry Development in *Vitis vinifera* Using RNA-Seq. *Plant Physiology* 152, 1787–1795.

## VIII. ANEXOS

### 1. ACCESO PÚBLICO DE LOS DATOS GENERADOS EN ESTA TESIS

#### A) REPORTE GENERAL DE LA SECUENCIACIÓN POR RNAseq

Muestra	Ápices	Lecturas (x10 <sup>6</sup> )	Q20 (%)	Contenido de GC (%)
Inicial 1	354	40.35	99.07	45.72
Inicial 2	350	49.28	99.04	46.14
Intermedia 1	470	50.51	99.06	45.90
Intermedia 2	458	43.49	99.06	45.99
Terminal 1	607	46.00	99.07	45.51
Terminal 2	610	49.58	99.04	45.71

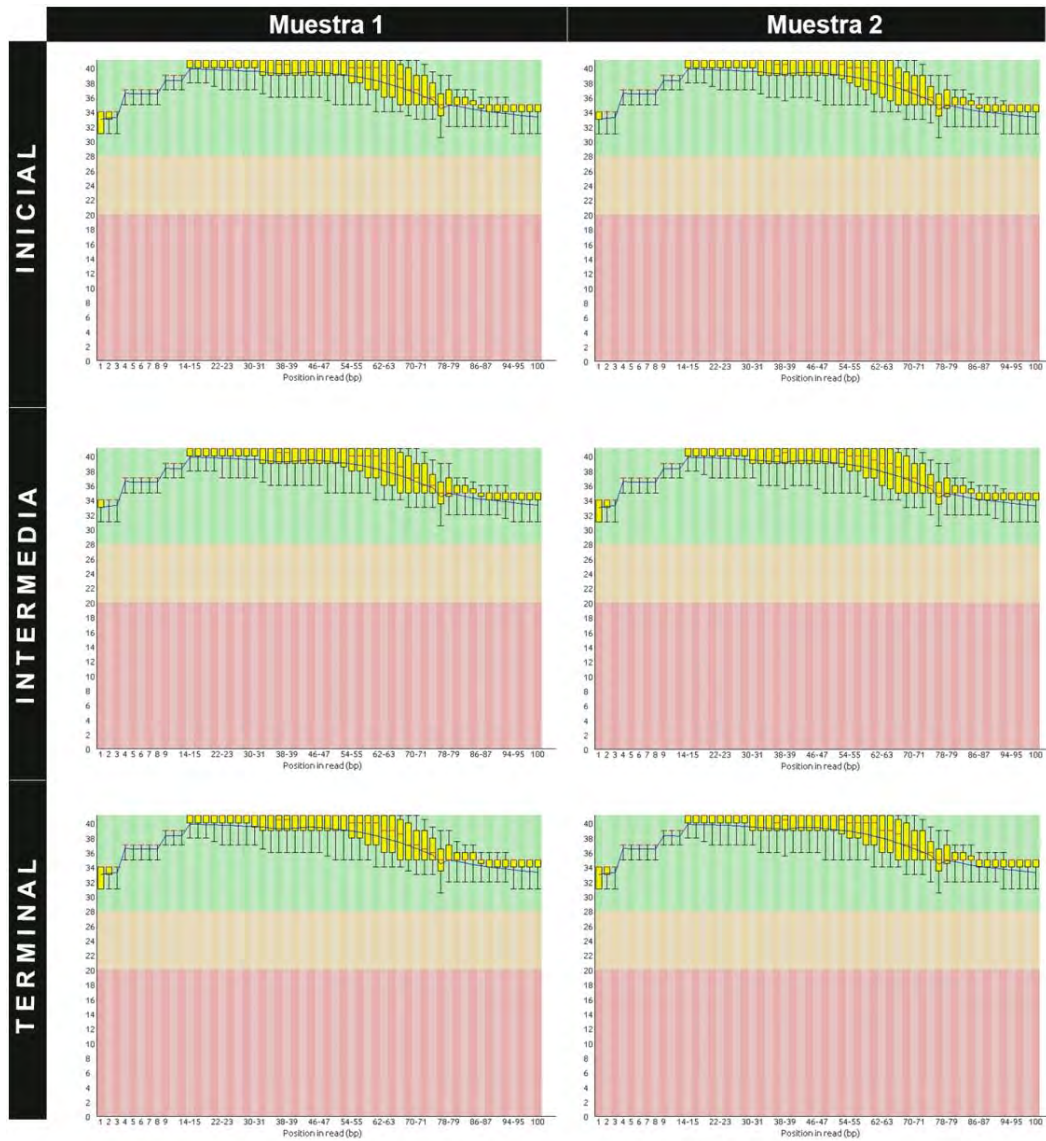
#### B) ACCESIÓN EN GEO-OMNIBUS

Series GSE104832		UPDATE	Query DataSets for GSE104832
Status	Public on May 25, 2018		
Title	RNA-seq insights into primary root apical meristem exhaustion and determinate root growth in <i>Pachycereus pringlei</i> (Cactaceae)		
Organism	<a href="#">Pachycereus pringlei</a>		
Experiment type	Expression profiling by high throughput sequencing		
Summary	<p>Many Cactaceae species exhibit determinate growth of the primary root as a consequence of root apical meristem (RAM) exhaustion. The genetic regulation of this root growth pattern is unknown. We de novo assembled and annotated the root apex transcriptome of the <i>Pachycereus pringlei</i> primary root at three developmental stages with active or exhausted RAM. The assembled transcriptome was characterized and used to evaluate differential transcript expression, identify RT-qPCR reference genes, and infer a transcriptional regulatory network. We generated a robust and comprehensive transcriptome of the primary root apex of <i>P. pringlei</i> and identified putative orthologs of <i>Arabidopsis</i> regulators of RAM maintenance, as well as putative lineage-specific transcripts. Furthermore, the transcriptome contained putative orthologs of most proteins involved in housekeeping processes, hormone signaling, and metabolic pathways. Specific transcriptional programs operate in the root apex at specific developmental time points. The transcriptional state of the <i>P. pringlei</i> root apex as the RAM becomes exhausted is comparable to that of cells from the meristematic, elongation, and differentiation zones of <i>Arabidopsis</i> roots. We suggest that the genetic program underlying the drought stress response is induced during development of the Cactaceae root, and that lineage-specific transcripts could contribute to root apical meristem exhaustion in Cactaceae.</p>		
Overall design	1 mm of the apex of the <i>Pachycereus pringlei</i> primary root was collected. 350, 450, and 600 root tips were pooled per samples (developmental stage) at the initial, intermediate, or terminal growth stage, respectively. Two biological replicates were included per developmental stage. Total RNA was extracted and mRNA-seq was performed on the Illumina HiSeq Platform at BGI-Tech, Hong Kong.		
Contributor(s)	<a href="#">Rodríguez-Alonso G</a> , <a href="#">Matvienko M</a> , <a href="#">Shishkova S</a>		
Citation(s)	Rodríguez-Alonso G, Matvienko M, López-Valle ML, Lázaro-Mixteco PE et al. Transcriptomics insights into the genetic regulation of root apical meristem exhaustion and determinate primary root growth in <i>Pachycereus pringlei</i> (Cactaceae). <i>Sci Rep</i> 2018 Jun 4;8(1):8529. PMID: <a href="#">29867103</a>		

## 2. ANÁLISIS DE CALIDAD DE LAS LECTURAS QUE SE GENERARON A PARTIR DE LA SECUENCIACIÓN POR RNAseq

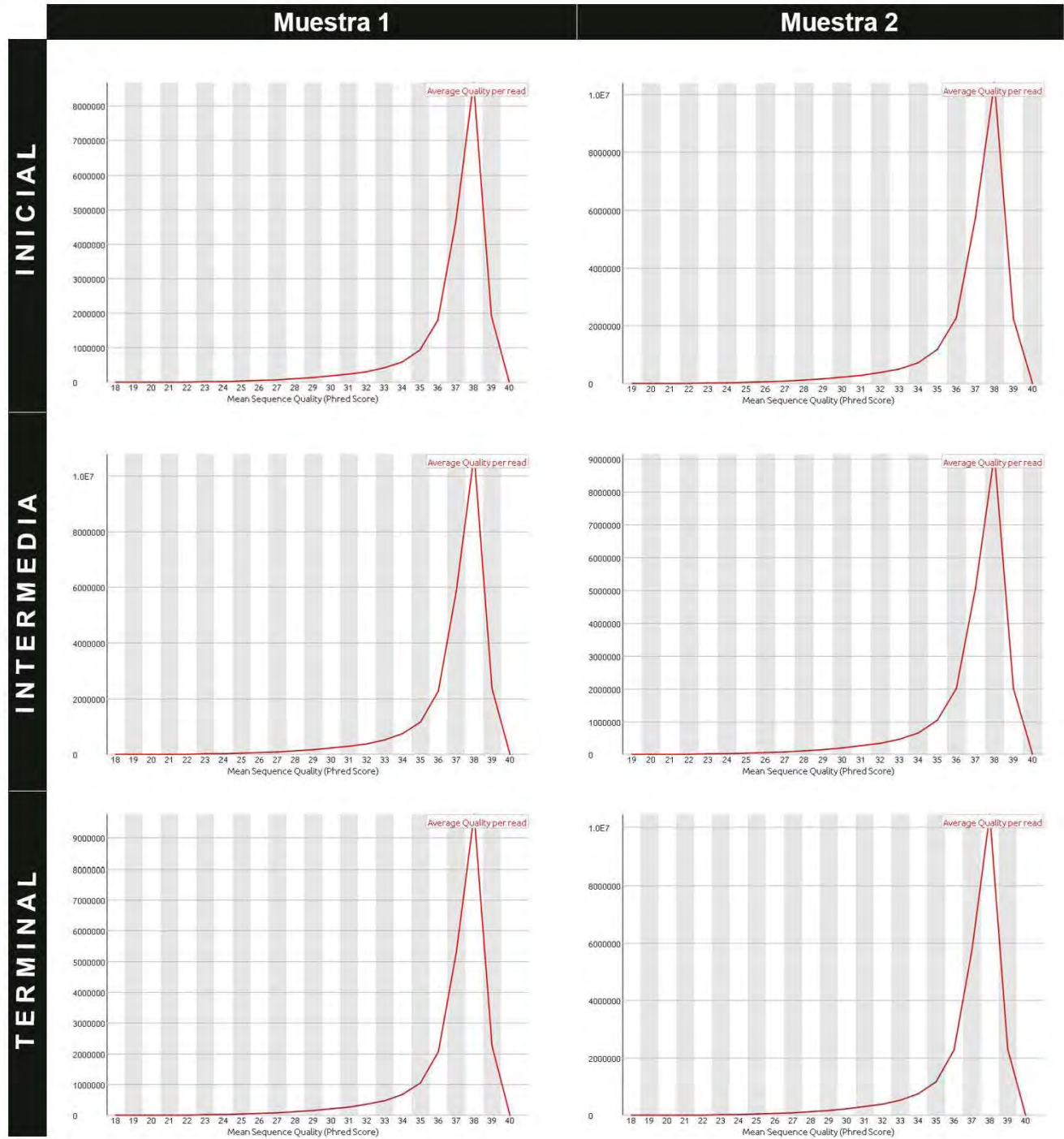
Los archivos de la secuenciación de ARN se almacenan en el formato FastQ, el cual incluye a la secuencia de nucleótidos de cada lectura y un score de calidad asociado a cada base (PHRED score), para indicar la confiabilidad de esa posición. A partir de estos archivos se realizó un análisis de control de calidad de las lecturas que se recibieron. Los resultados se muestran a continuación.

### A) PHRED Score por posición.



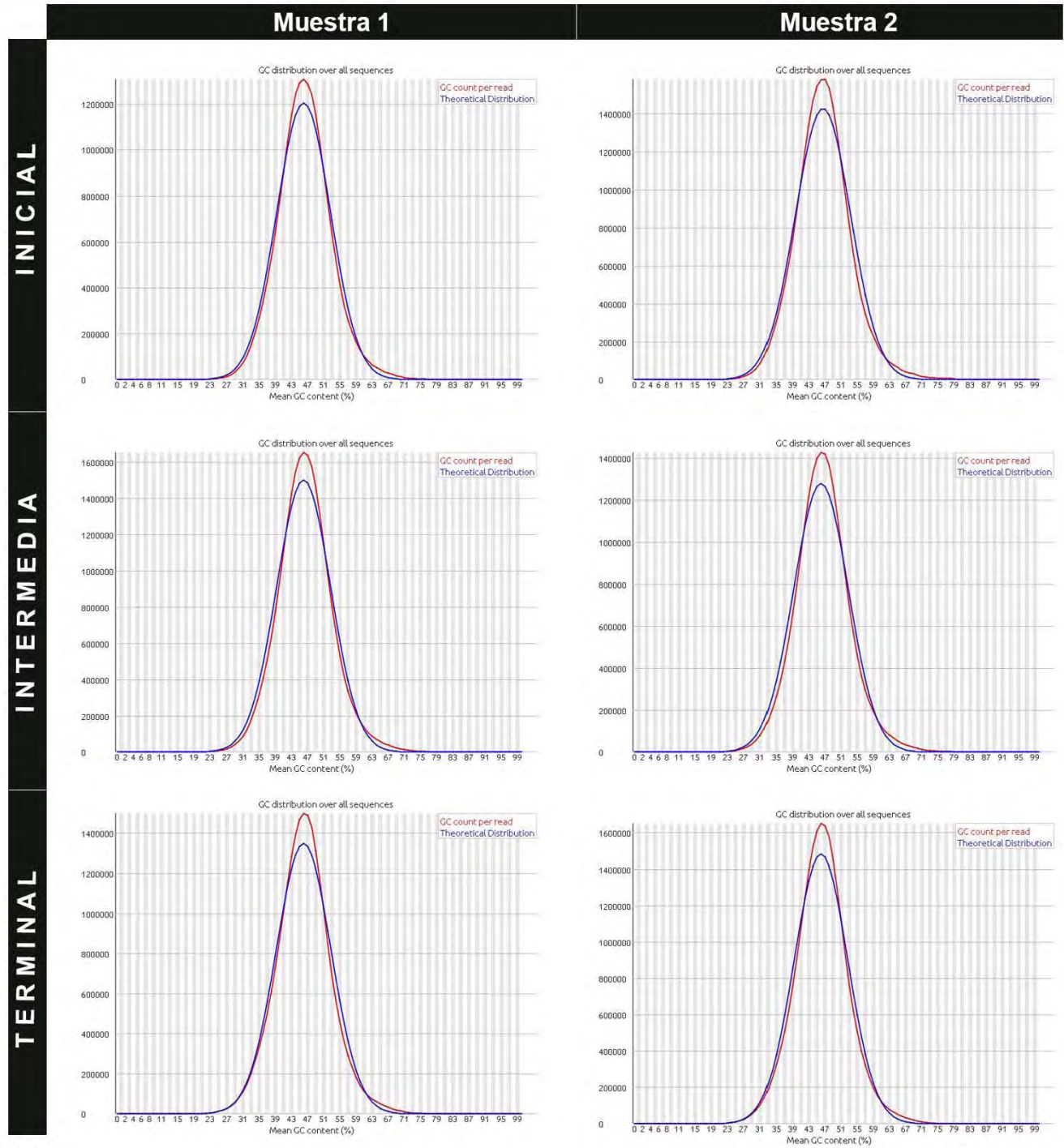


## B) Calidad por secuencia

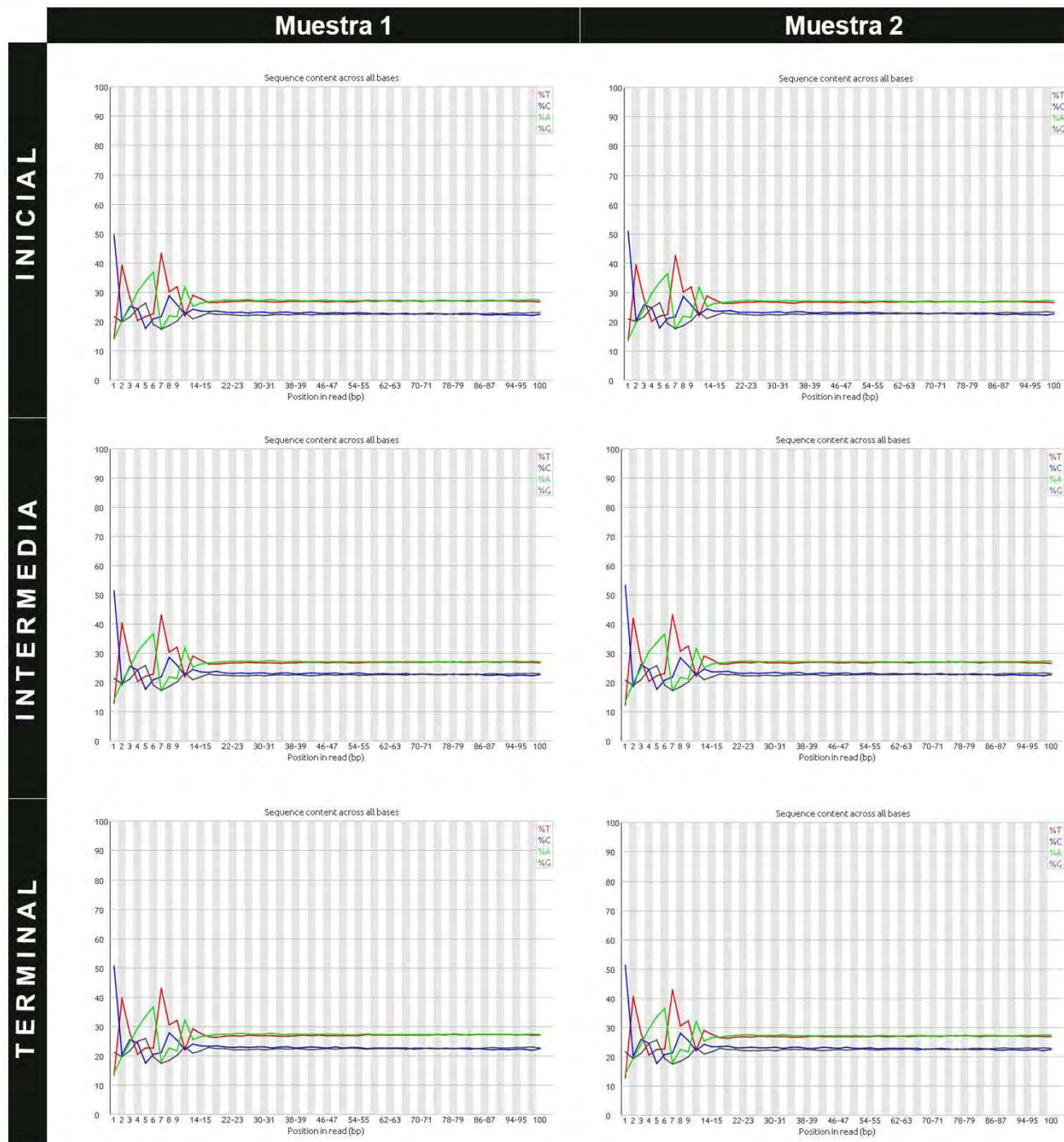




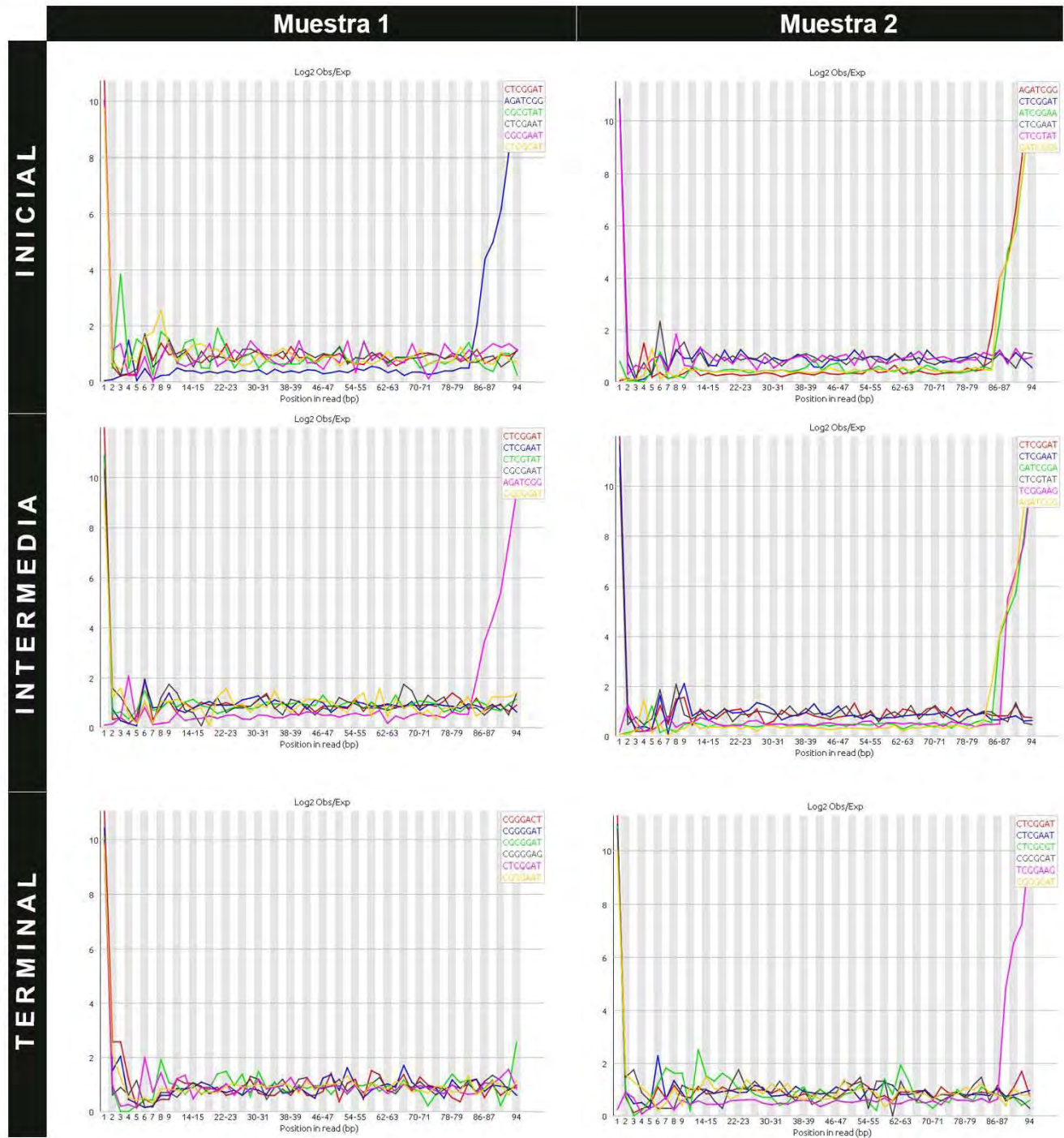
### C) Contenido de GC por secuencia



## D) Distribución de nucleótidos por base



## E) Secuencias cortas (k-mers) sobre-representadas.



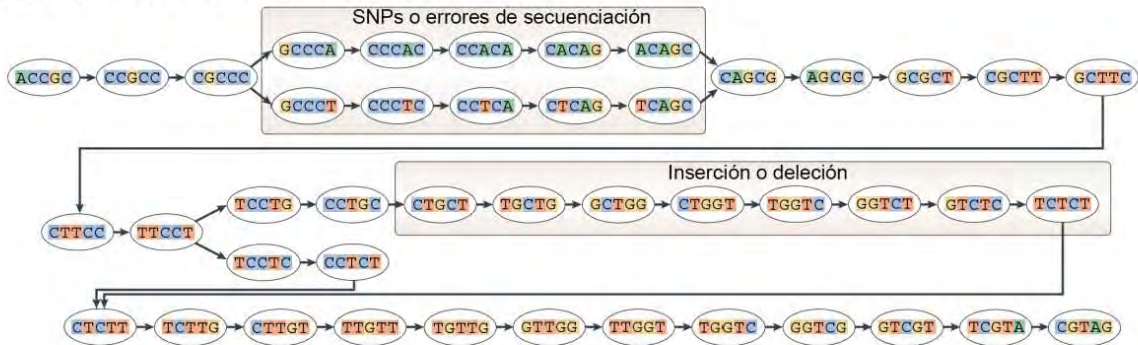


### 3. DEFINICIÓN DE PARÁMETROS DURANTE EL ENSAMBLAJE DE-NOVO DEL TRANSCRIPTOMA

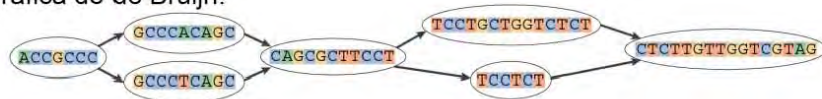
a) Se generan todas las palabras de tamaño k.



b) Construcción de la gráfica de de Bruijn.



c) Colapso de la gráfica de de Bruijn.



d) Explorar la gráfica



e) Isoformas ensambladas



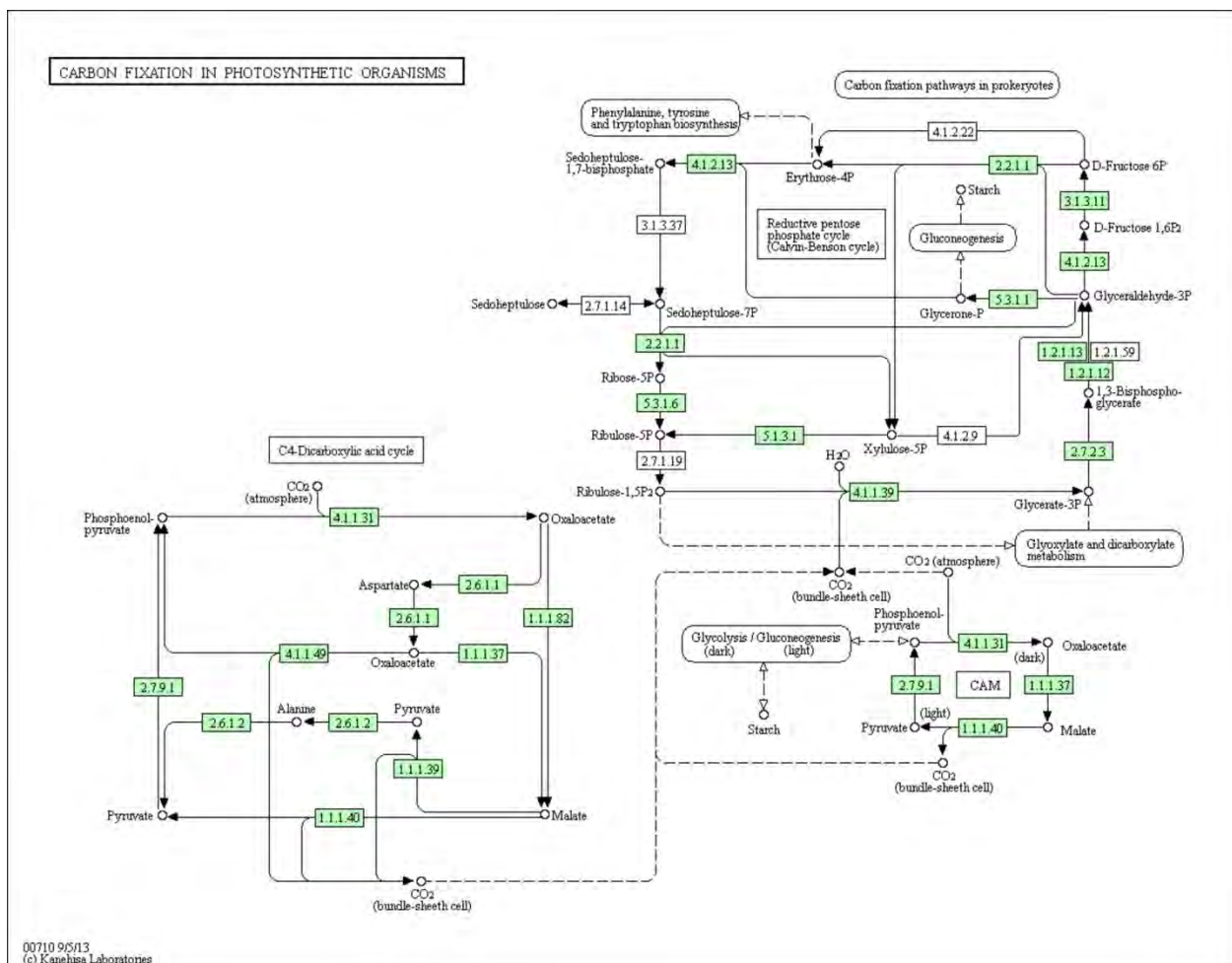
Tomado de Martin & Wang 2011. DOI: 10.1038/nrg3068

El ensamblaje *de novo* de un transcriptoma no utiliza a un genoma de referencia para la reconstrucción de transcritos. En su lugar, se toma ventaja del hecho de que los fragmentos secuenciados se generaron al azar y que por consiguiente existirá redundancia de secuencia entre algunos de ellos. Para la reconstrucción de transcritos las lecturas se fragmentan virtualmente en diferentes secuencias cortas llamadas palabras, o  $k$ -meros, en donde  $k$  define el número de bases que contendrá cada palabra (WordSize). A partir de cada lectura se obtienen todas las palabras de tamaño  $k$  que están contenidas en ella (Panel A), de esta manera se genera una lista con todas las posibles palabras contenidas en todas las lecturas que se ensamblarán. Para el ensamblaje se escoge a una de estas palabras como semilla, o “seed”, y se busca entre la lista de todas las palabras generadas a aquellas cuya secuencia en los extremos es idéntica a la anterior. En el caso ideal, sólo existe un  $k$ -mero hacia cada extremo de la palabra que se utiliza como semilla. Cada una de las palabras que se ensamblan en contigs se utiliza para representar un nodo en un arreglo conocido como gráfica de de Bruijn (Panel B). Cuando dos palabras pueden continuar la extensión del contig, se abre una dicotomía, o burbuja. La burbuja se colapsa cuando en una dirección dada sólo existe una palabra posible y la dicotomía se cierra. El tamaño máximo de la burbuja (bubble size), es decir, la capacidad de resolución de ambigüedades durante el ensamblaje de los contigs, puede controlarse y varía en función de la astringencia que se requiera durante la reconstrucción de los transcritos. Cuando para las palabras no se encuentra ningún nuevo  $k$ -mero que continúe la extensión del contig, el proceso se detiene. La gráfica de de Bruijn se abrevia al eliminar las redundancias y empalmar las secuencias en sus regiones idénticas (Panel C). Se realiza entonces un recorrido a través de la gráfica de de Bruijn abreviada (Panel D) que contiene a las burbujas y los  $k$ -meros fusionados para identificar a todas las posibles variantes de un contig (Panel E), las cuales pueden corresponder a isoformas, SNPs, o incertidumbres en la secuenciación.

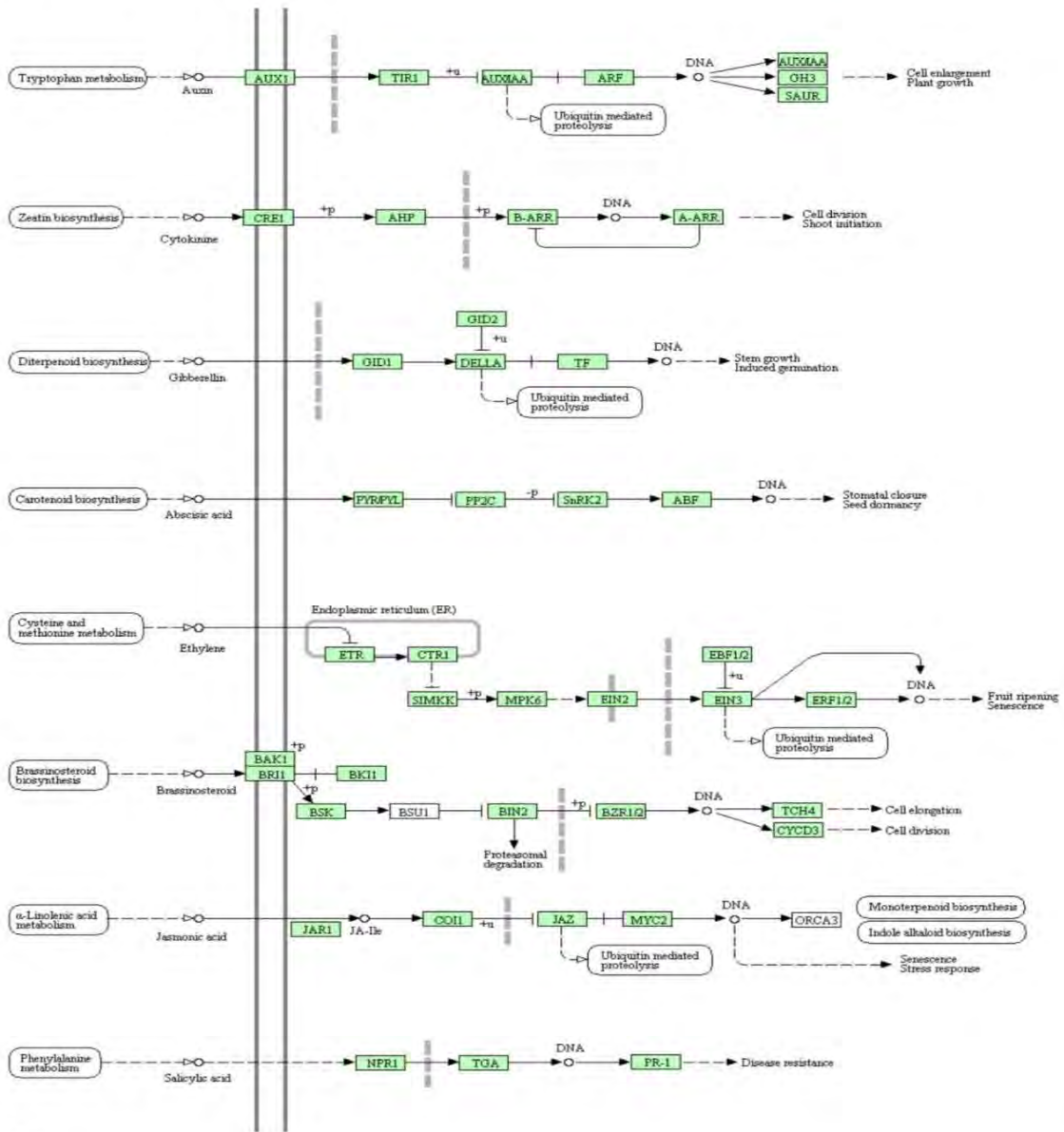
## 4. ALGUNAS RUTAS Y PROCESOS BIOLÓGICOS CONTENIDOS EN EL TRANSCRIPTOMA DEL ÁPICE DE LA RAÍZ PRIMARIA DE *Pachycereus pringlei*

El transcriptoma ensamblado *de novo* para el ápice de la raíz primaria de *Pachycereus pringlei* es amplio, como se muestra a partir del mapeo de los transcritos en rutas metabólicas, de señalización, o en componentes de complejos multiproteicos, mediante el uso del servidor KAAS-KEGG. Los recuadros en los esquemas siguientes representan enzimas o proteínas estructurales. Los recuadros verdes indican que se encontró al menos a un ortólogo putativo de esa proteína en el transcriptoma de *P. pringlei*; los recuadros blancos indican que no se encontró ningún ortólogo putativo correspondiente a esa proteína.

Los diagramas presentados aquí se tomaron con consentimiento de Kanehisa Laboratories.

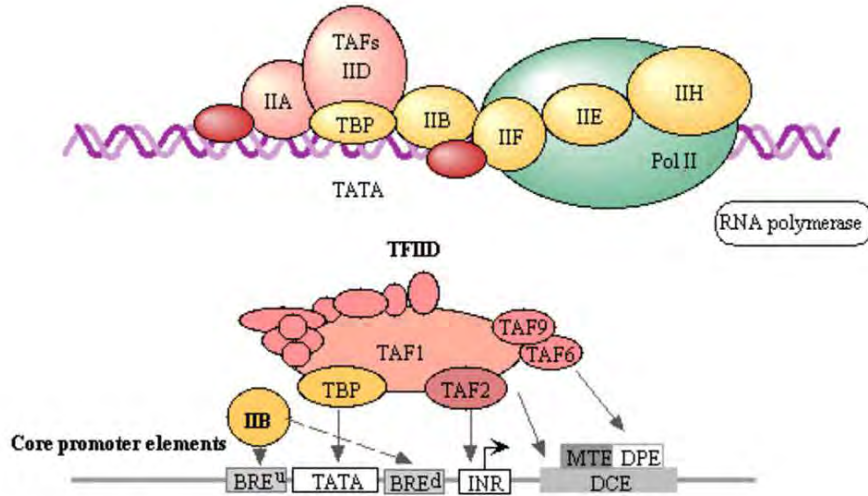


PLANT HORMONE SIGNAL TRANSDUCTION

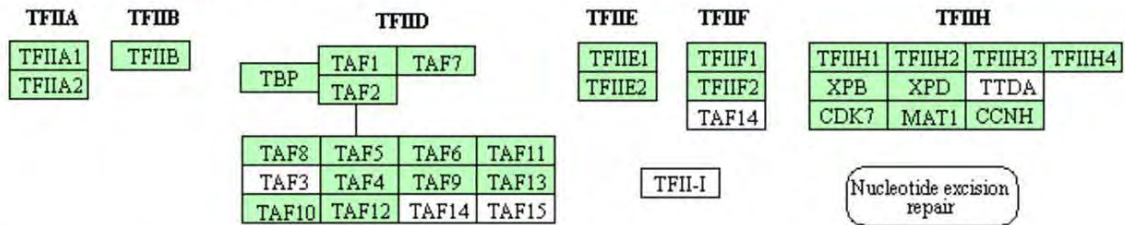


04075 9/6/16  
© Kanhan Laboratories

**BASAL TRANSCRIPTION FACTORS (EUKARYOTES)**



**General transcription factors for RNA polymerase II**

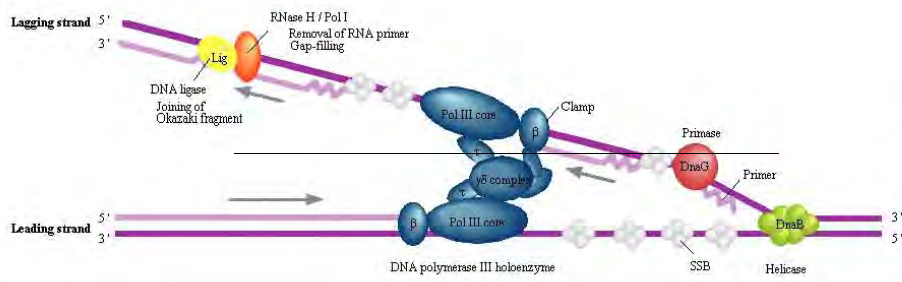


03022 3/15/11  
(c) Kanehisa Laboratories



**DNA REPLICATION**

**Replication complex (Bacteria)**



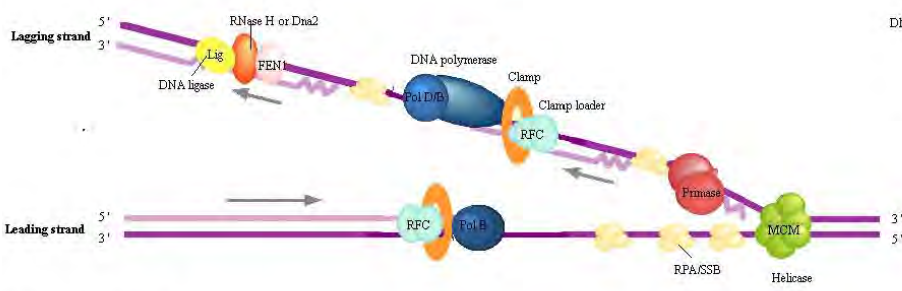
DNA polymerase III holoenzyme

β	Pol III core		
ε			
α			
γ & τ			
Clamp	ψ	δ'	γδ complex
β	κ	δ	

Helicase: DnaB  
Primase: DnaG  
SSB

RNaseH: RNaseHI  
DNA polymerase I: Dpol  
DNA ligase: Lig

**Replication complex (Archaea)**



DNA polymerase B DNA polymerase D

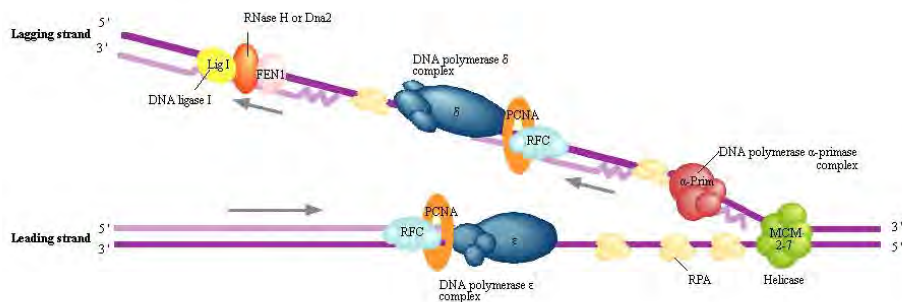
PolB	PolD1
	PolD2

Helicase: MCM  
Primase: Pri1, Pri2  
RFA/SSB: RPA

Clamp: PCNA  
Clamp loader: RFC5, RFC6  
RNaseH: RNaseHI, RNaseHII

Helicase: Dna2  
Fen1  
DNA ligase: Lig

**Replication complex (Eukaryotes)**



DNA polymerase α-primase complex

α1	α2	Pri1	Pri2
----	----	------	------

DNA polymerase δ complex

δ1	δ2	δ3	δ4
----	----	----	----

DNA polymerase ε complex

ε1	ε2	ε3	ε4
----	----	----	----

MCM complex (helicase)

Mcm2	Mcm3	RFA
Mcm4	Mcm5	RFA1
Mcm6	Mcm7	RFA2/4
		RFA3

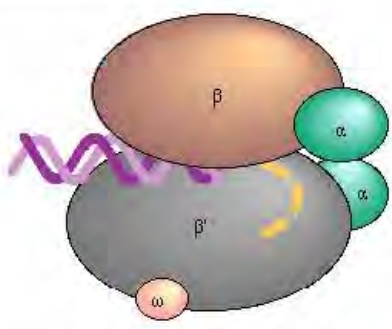
Clamp: PCNA  
Clamp loader: RFC1, RFC2/4, RFC3/5

RNaseHI: RNaseH1  
RNaseHII: RNaseHA, RNaseHB, RNaseHC

Helicase: Dna2  
Fen1  
DNA ligase: Lig1

03030 1/23/15  
(c) Kanehisa Laboratories

RNA POLYMERASE



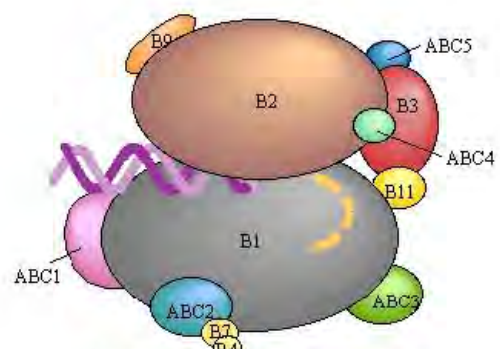
RNA polymerase (*Thermus aquaticus*)

Bacterial

$\beta$	$\alpha$	$\omega$	$\delta$
$\beta'$			

Archaeal

B	D	F	H	K	E
A	G		N	L	P



RNA polymerase II (*Saccharomyces cerevisiae*)

Eukaryotic Pol II

Core subunits		Pol II specific subunits	Pol I, II, and III common subunits
B2	B3	B4	ABC1
B1	B11	B7	ABC2
		B9	ABC3
			ABC4
			ABC5

Eukaryotic Pol III

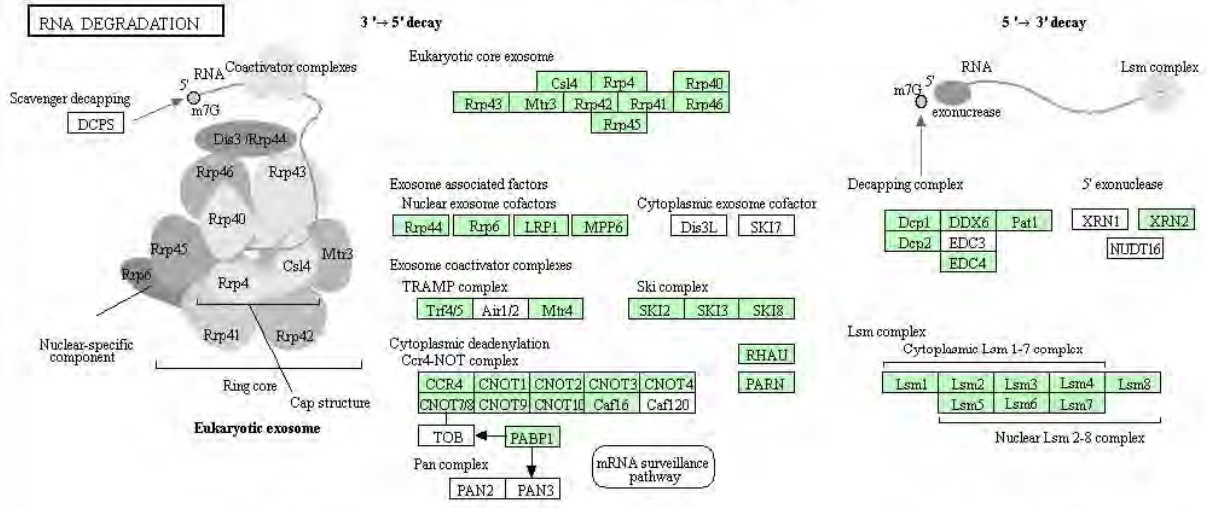
Core subunits		Pol III specific subunits		
C2	AC2	C3	C4	C11
C1	AC1	C25	C31	C34
				C37

Eukaryotic Pol I

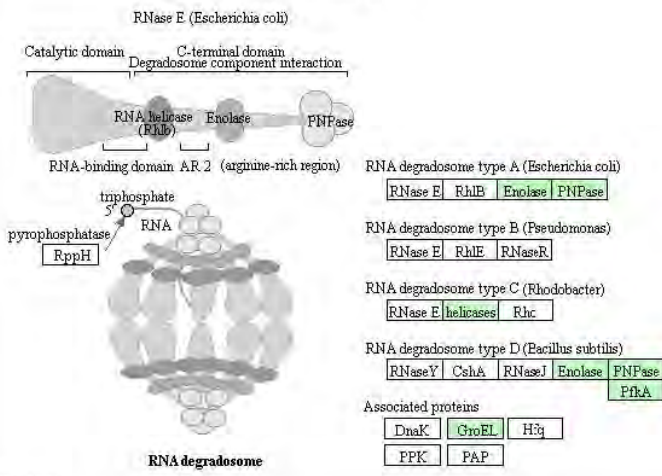
Core subunits		Pol I specific subunits	
A2	AC2	A12	A14
A1	AC1	A34	A49
			A43

03020 3/25/11  
(c) Kanehisa Laboratories

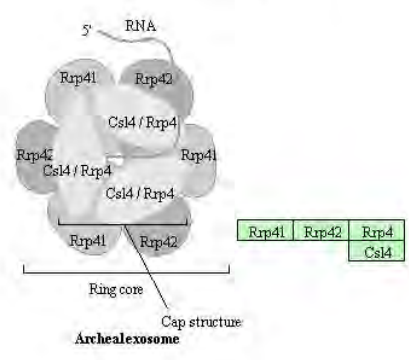
**Eukaryotic RNA degradation**



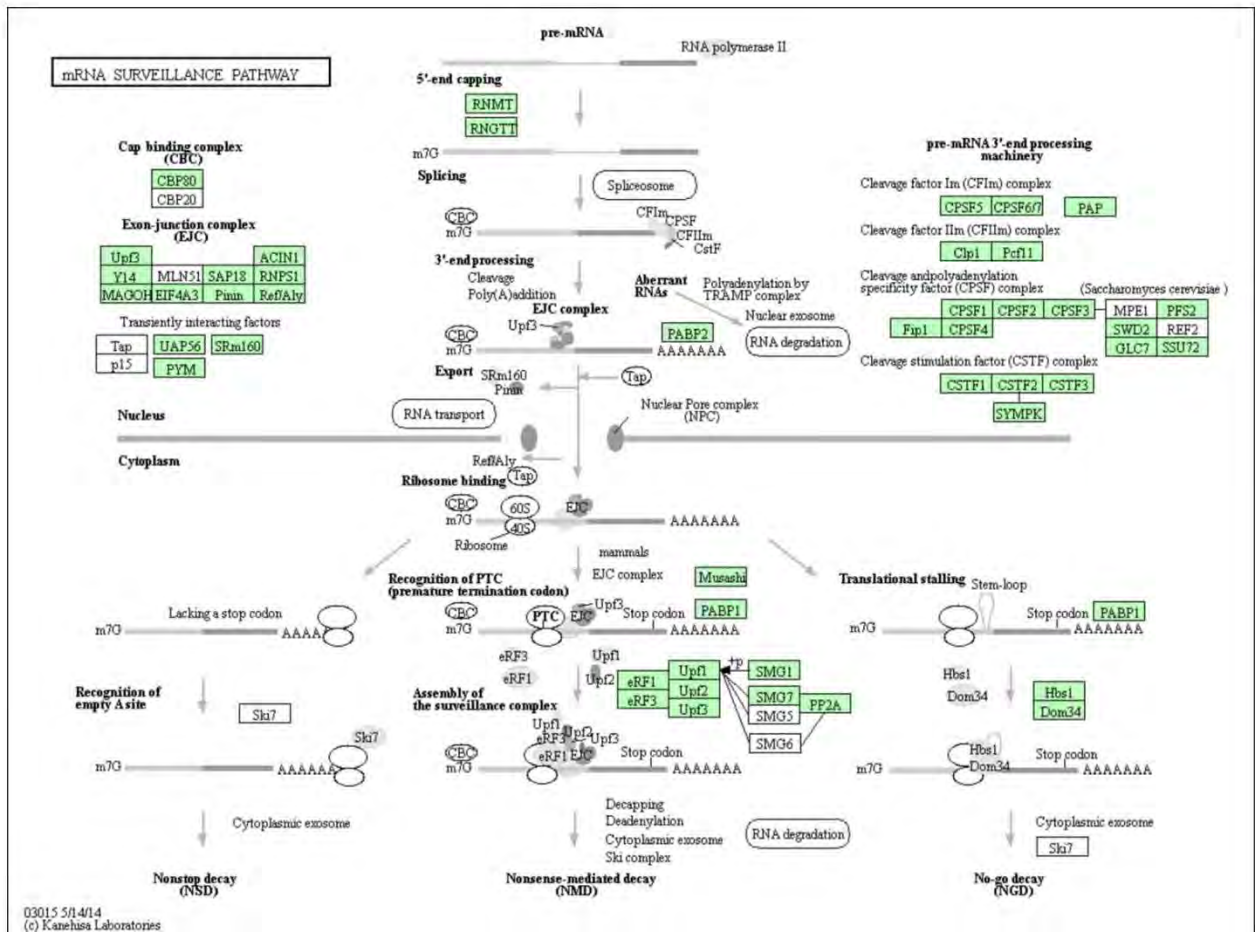
**Bacterial RNA degradation**



**Archeal RNA degradation**



03018 11/6/14  
(c) Kanehisa Laboratories





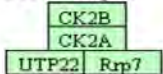
# RIBOSOME BIOGENESIS IN EUKARYOTES

## Ribosomal RNAs

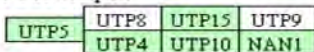


## 90S pre-ribosome components

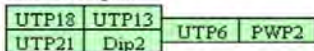
### UTP-C complex



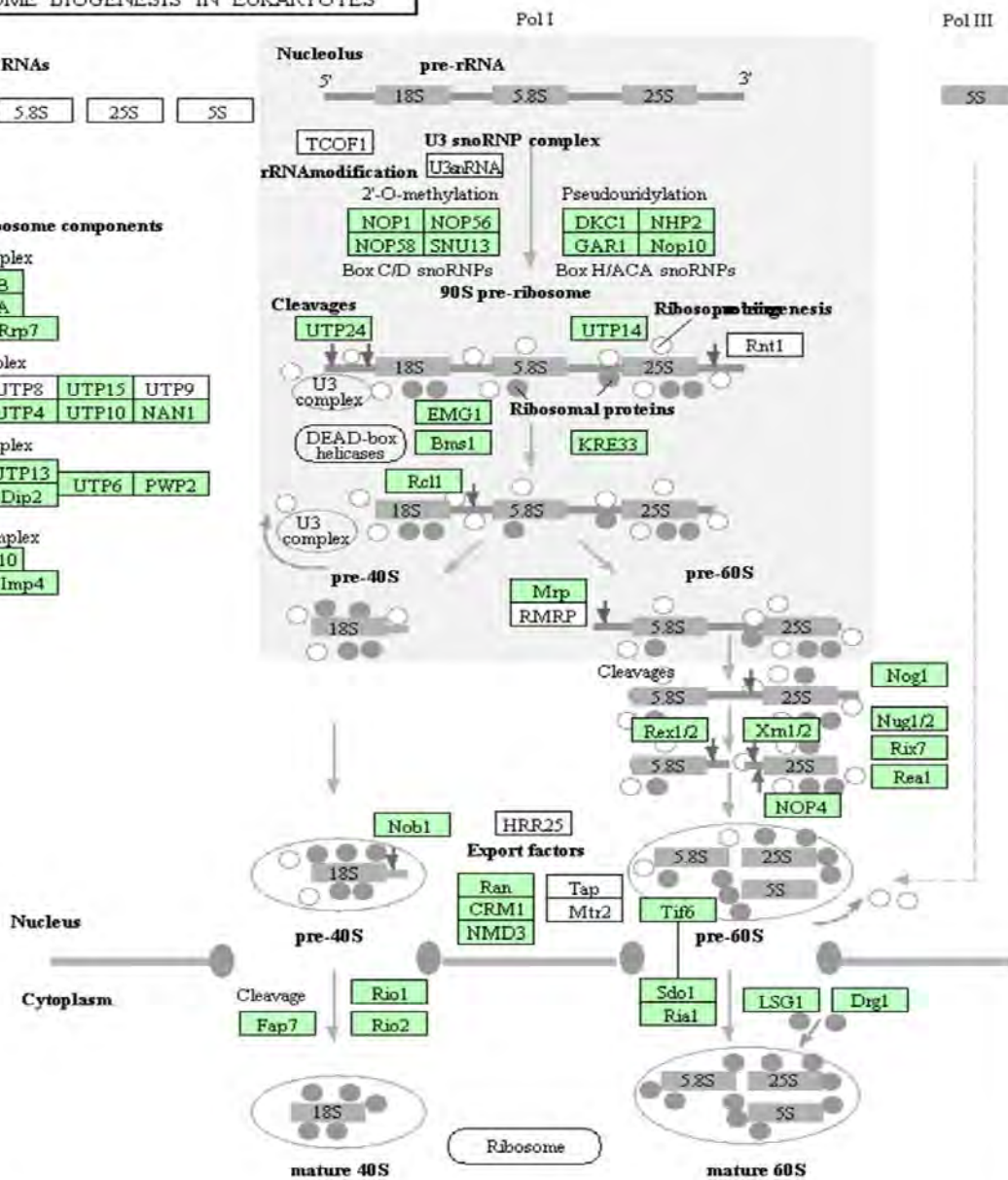
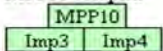
### t-UTP complex



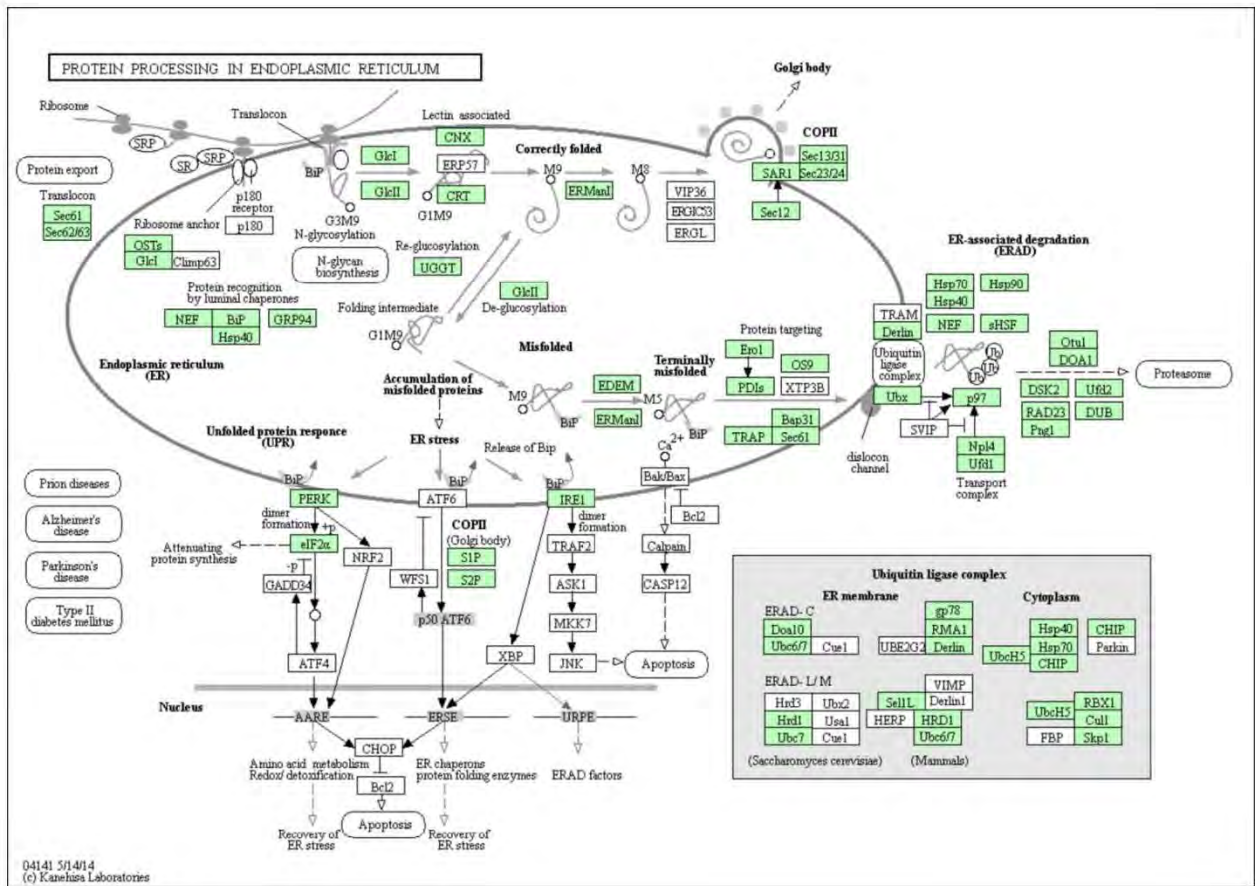
### UTP-B complex



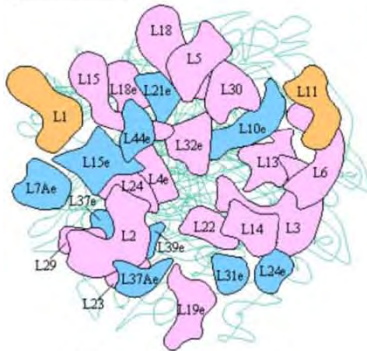
### MPP10 complex



03008 9/10/14  
 (c) Kanehisa Laboratories



**RIBOSOME**



Large subunit (*Haloarcula marismortui*)



Small subunit (*Thermus aquaticus*)

**Ribosomal RNAs**

Bacteria / Archaea	23S	5S	16S
Eukaryotes	25S	5.8S	18S

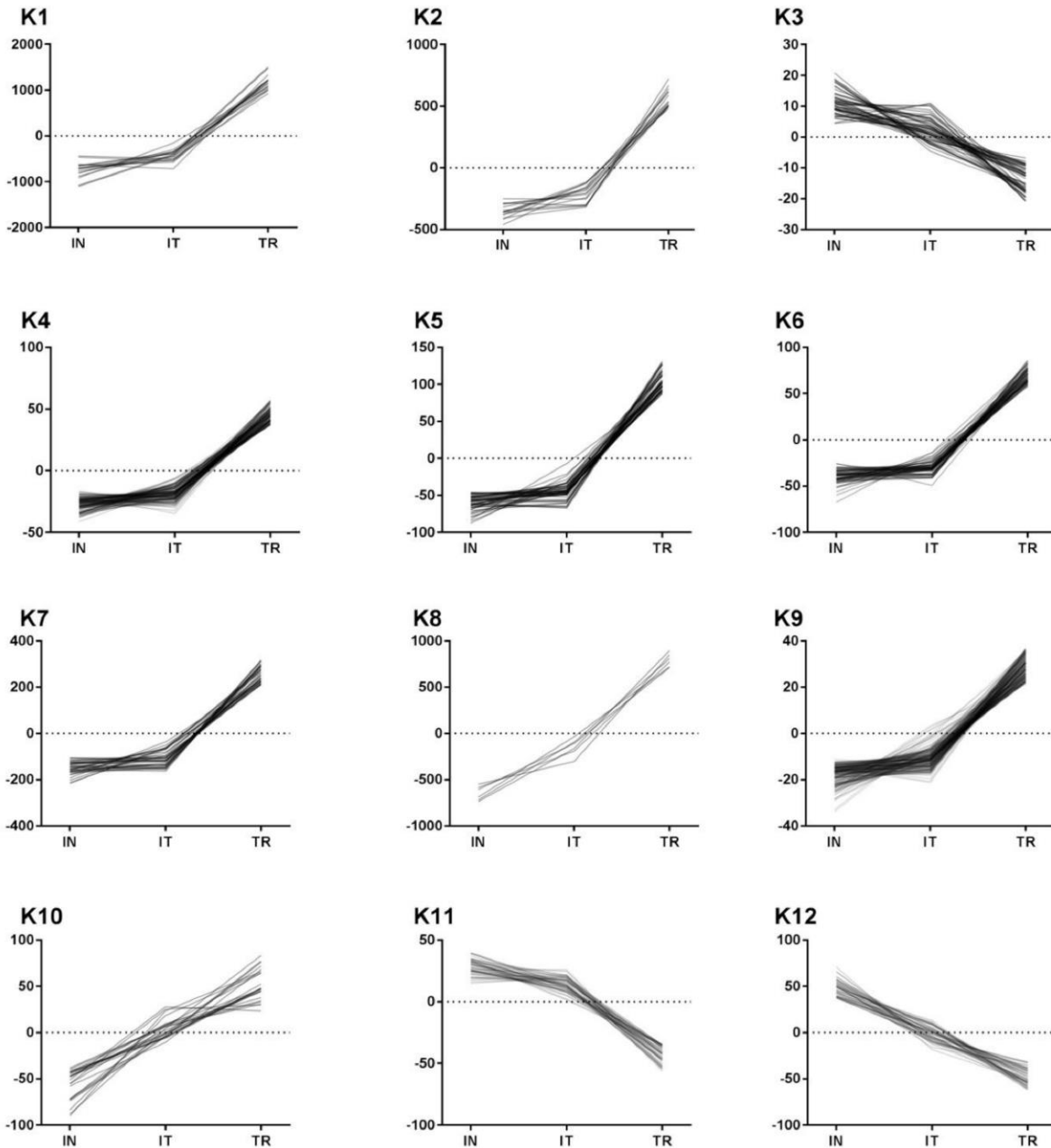
**Ribosomal proteins**

EF-Tu	S10	L3	L4	L23	L2	S19	L22	S3	RP-L16	L29					
	S20e	L3e	L4e	L23Ae	L8e	S15e	L17e	S3e	L10e	L35e	L7/L12 stalk				
	S17	L14	L24	L5	S14	S8	L6	L18	S5	L30	L15				
	S11e	L23e	L26e	S4e	L11e	S29e	S15Ae	L9e	L32e	L19e	L5e	S2e	L7e	L27Ae	SecY
IF1	L36	S13	S11	S4	RpoA	L17	L13	S9							
	L34e	L14e	S18e	S14e	S9e	L18e	L13Ae	S16e							
EF-Tu,G	S7	S12	L7A	RpoC,B	L7/L12	L12	L10	L1	L11						
	S5e	S23e	L30e	L7Ae	L7/L12	L12	L10	L1	L11	L1/LP2	L10Ae	L12e			
EF-Ts	S2	IF2	S15	IF3	L35	L20	L34	RF1	L31	L32	L9	S18	S6		
	S4e	S13e	S13e												
	L28	L33	L21	L27	FtsY,Ffh	S16	L19	S1	S20	S21	L25				
	L10e	L13e	L15e	L21e	L24e	L31e	L35Ae	L37e	L37Ae	L39e	L40e	L41e	L44e		
	S3Ae	S6e	S8e	S17e	S19e	S24e	S25e	S26e	S27e	S27Ae	S28e	S30e	LX		
	L6e	L18Ae	L22e	L27e	L28e	L29e	L36e	L38e							
	S7e	S10e	S12e	S21e											

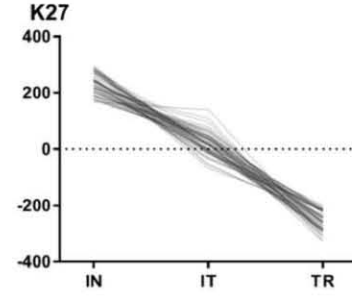
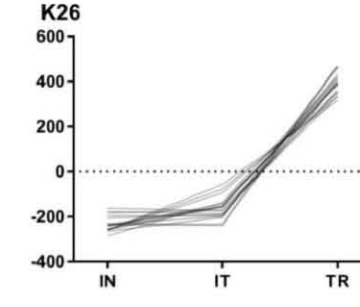
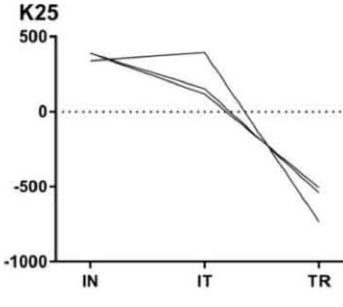
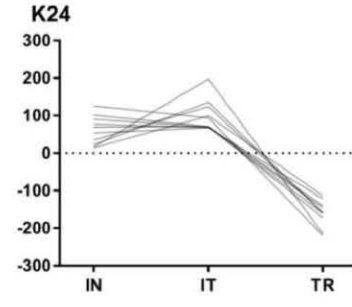
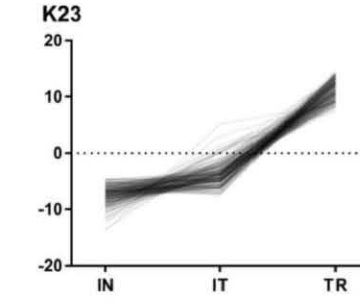
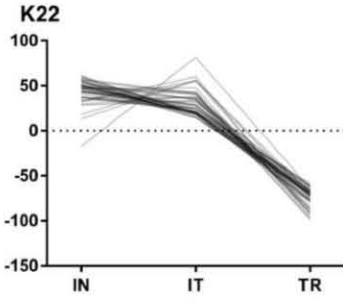
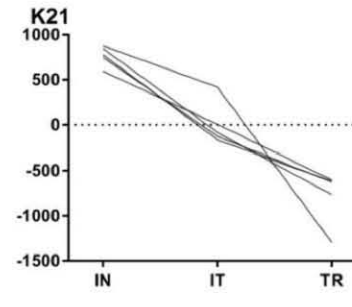
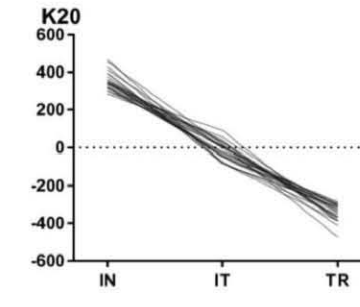
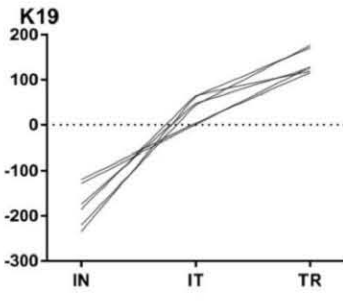
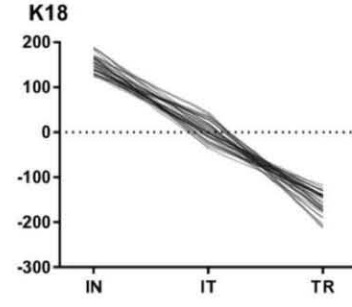
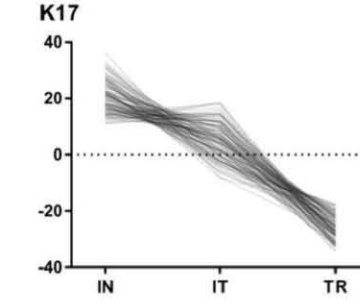
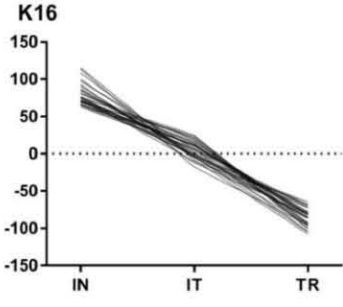
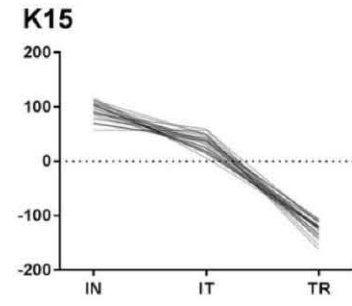
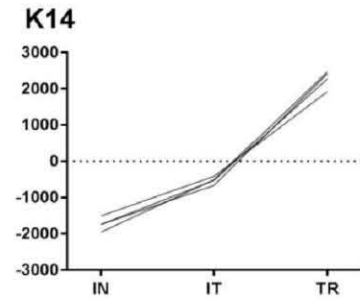
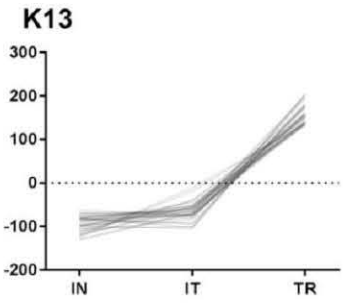
03010 7/29/13  
(c) Kanehisa Laboratories

## 5. PERFILES DE TRANSCRIPCIÓN IDENTIFICADOS MEDIANTE K-MEAN CLUSTERING.

Perfiles transcripcionales que se identificaron mediante el método de k-mean clustering de los transcritos diferencialmente expresados (DETs). Cada gráfica muestra los valores normalizados y centrados de los niveles de expresión de los DETs en las tres etapas del desarrollo de la raíz primaria de *P. pringlei*: IN, inicial; IT, Intermedia; TR, Terminal.







## 6. RED DE REGULACIÓN TRASCRIPCIONAL

### A) Tabla de interacciones

Node A	Interaction type	Node B
AT1G30330	Activation	AT2G44810
AT1G30490	Repression	AT5G16560
AT5G16560	Repression	AT1G65620
AT5G16560	Repression	AT1G73590
AT1G09770	Activation	AT1G62360
AT1G09770	Activation	AT1G75820
AT1G62360	Repression	AT2G37630
AT1G62360	Repression	AT1G65620
AT1G62360	Activation	AT1G75820
AT2G45190	Repression	AT4G18960
AT2G45190	Activation	AT2G34710
AT4G18960	Activation	AT2G44810
AT4G18960	Activation	AT4G18960
AT2G34710	Activation	AT1G65620
AT2G34710	Repression	AT5G16560
AT2G34710	Activation	AT2G45190
AT2G34710	Activation	AT3G52770
AT2G37630	Activation	AT2G34710
AT2G37630	Repression	AT2G45190
AT1G65620	Repression	AT2G45190
AT1G65620	Activation	AT2G34710
AT1G65620	Repression	AT1G62360
AT1G65620	Repression	AT5G16560

AT4G37750	Activation	AT5G61850
AT4G37750	Activation	AT5G60690
AT4G37750	Repression	AT4G18960
AT4G37750	Activation	AT2G45190
AT4G37750	Activation	AT2G34710
AT5G42630	Repression	AT1G73590
AT5G45980	Activation	AT5G59340
AT5G45980	Activation	AT3G11260
AT5G62000	Activation	AT1G01480
AT5G62000	Activation	AT4G11280
AT5G62000	Repression	AT4G37750
AT5G62000	Activation	AT4G37770
AT5G62000	Activation	AT5G25890
AT2G28350	Repression	AT3G11260
AT5G61850	Activation	AT4G18960
AT5G60690	Activation	AT2G45450
AT3G44750	Activation	AT2G34710
AT3G44750	Activation	AT2G45190
AT1G55580	Activation	AT1G62360
AT1G55580	Activation	AT2G37630
AT1G55580	Activation	AT5G60690
AT1G55580	Activation	AT4G37750
AT1G79580	Repression	AT1G26870
AT1G26870	Activation	AT1G79580
AT1G56010	Activation	AT2G04160
AT3G25710	Repression	AT5G42800

AT4G30080	Repression	AT3G11260
AT4G25470	Activation	AT1G04240
AT4G37650	Activation	AT5G03150
AT4G37650	Activation	AT3G54220
AT4G37650	Activation	AT1G03840
AT4G37650	Activation	AT1G50420
AT5G03150	Activation	AT3G54220
AT3G54220	Activation	AT1G03840
AT3G54220	Activation	AT3G12280
AT3G54220	Activation	AT5G03150
AT5G11260	Activation	AT4G32880
AT5G11260	Activation	AT4G14550
AT5G11260	Activation	AT5G42800
AT4G32880	Activation	AT4G37650
AT5G10510	Activation	AT4G18960
AT1G19850	Activation	AT3G20840
AT1G19850	Activation	AT4G11280
AT1G19850	Activation	AT4G32880
AT1G19850	Activation	AT4G37770
AT1G19850	Activation	AT1G04550
AT1G19850	Activation	AT1G12980
AT1G19850	Activation	AT1G19850
AT1G19850	Activation	AT1G73590
AT1G19850	Activation	AT1G01480
AT5G61420	Activation	AT2G20610
AT2G33880	Activation	AT3G11260

AT1G53230	Activation	AT2G37630
AT1G53230	Activation	AT1G04240
AT1G19220	Repression	AT4G14550
AT1G19220	Repression	AT4G37650
AT1G19220	Activation	AT2G42430
AT1G19220	Activation	AT3G58190
AT1G19220	Activation	AT1G04240
AT2G20180	Activation	AT1G03430
AT2G20180	Repression	AT5G54510
AT5G20730	Activation	AT3G20840
AT5G20730	Activation	AT3G15540
AT5G20730	Activation	AT3G58190
AT5G20730	Activation	AT3G50340
AT5G20730	Activation	AT1G19220
AT5G20730	Activation	AT1G04240
AT5G20730	Activation	AT2G45420
AT5G20730	Activation	AT2G42430
AT5G20730	Repression	AT4G37650
AT5G20730	Activation	AT4G14560
AT5G20730	Repression	AT4G14550
AT5G20730	Activation	AT4G37390
AT5G20730	Activation	AT4G27260
AT5G17430	Activation	AT3G44750
AT5G17430	Activation	AT5G03740
AT5G17430	Activation	AT5G22650
AT5G18560	Activation	AT5G61850

AT5G37020	Activation	AT1G28130
AT5G37020	Activation	AT2G44810
AT5G37020	Activation	AT5G54510
AT5G60890	Activation	AT2G20610
AT4G00220	Repression	AT2G37630
AT4G00220	Repression	AT2G01420
AT4G00220	Repression	AT1G73590
AT4G00220	Repression	AT1G70940
AT4G00220	Activation	AT1G62360
AT4G00220	Repression	AT1G23080

## B) Tabla de propiedades

Node	<i>Arabidopsis</i> name	<i>P. pringlei</i> contig	<i>P. pringlei</i> annotation	E-value	BBH	RPKM Initial	RPKM Interm.	RPKM Term.
AT1G30330	ARF6	0001038	auxin response factor 6-like	0	Y	24.12	27.34	18.65
AT2G44810	DAD1	0005103	phospholipase a1-chloroplastic-like	2.52E-75	N	7.27	9.93	25.76
AT1G30490	PHV	0001618	homeobox-leucine zipper protein revoluta	0	Y	7.93	9.77	7.39
AT5G16560	KAN	0005141	probable transcription factor kan2-like	4.24E-47	N	2.62	4.36	4.6
AT1G09770	CDC5	0000721	cell division cycle 5-like isoform x1	0	Y	23.08	26.9	20.04
AT1G62360	STM	0009327	homeobox protein knotted-1-like 6	3.66E-72	N	2.52	2.83	3.05
AT1G75820	CLV1	0001781	receptor protein kinase clavata1-like	0	Y	0.6	1.5	3.68
AT2G45190	YAB1	NSH	NSH	3.51	NSH	0.001	0.001	0.001
AT4G18960	AG	0032278	mads-box transcription factor 27-like isoform x1	2.62E-44	N	6.33	3.48	4.25
AT2G34710	PHB	0000849	homeobox-leucine zipper family protein	0	Y	12.14	14.87	11.42
AT2G37630	PHAN	0008775	transcription factor as1	1.94E-118	Y	5.7	8.01	7.17
AT1G65620	AS2	0027369	lob domain-containing protein 6	1.71E-62	Y	2.89	3.22	1.86
AT1G73590	PIN1	0002874	auxin efflux carrier component 1-like	0	Y	44.07	46	33.59
AT3G52770	ZPR3	0030394	PREDICTED: uncharacterized protein LOC101514572	1.15E-08	Y	1.74	1.66	0.78
AT4G37750	ANT	0014765	ap2-like ethylene-responsive transcription factor ant-like	1.99E-113	Y	0.53	0.34	1.28

AT5G42630	KAN4	0013106	probable transcription factor kan2-like	6.47E-35	N	3.3	3.75	1.42
AT5G45980	WOX8	0045416	wuschel-related homeobox 11-like	7.83E-16	N	1.15	0.48	0.54
AT5G59340	WOX2	0033393	wuschel-related homeobox 4	1.52E-21	N	0.33	0.39	1.91
AT3G11260	WOX5	0019483	wuschel-related homeobox 5-like	2.28E-51	Y	1.77	3.98	3.94
AT5G62000	ARF2	0001054	auxin response factor 2	0	N	19.42	26.39	30.43
AT1G01480	ACS2	0006764	1-aminocyclopropane-1-carboxylate synthase-like	0	Y	1.28	2.89	6.82
AT4G11280	ACS6	0006764	1-aminocyclopropane-1-carboxylate synthase-like	0	Y	1.28	2.89	6.82
AT4G37770	ACS8	0007505	1-aminocyclopropane-1-carboxylate synthase-like	0	Y	9.92	25.33	24.34
AT5G25890	IAA28	0005636	auxin-responsive protein iaa26-like	5.59E-35	N	13.32	20.16	43.38
AT2G28350	ARF10	0001279	auxin response factor 18-like	0	Y	5.09	5.04	4.42
AT5G61850	LFY	0000040	dna polymerase theta	4.14	NSH	0.001	0.001	0.001
AT5G60690	REV	0001618	homeobox-leucine zipper protein revoluta	0	Y	7.93	9.77	7.39
AT3G44750	HD2A	0027114	histone deacetylase isoform 2	8.55E-30	Y	286.77	270.64	131.72
AT2G45450	ZPR1	0041901	PREDICTED: uncharacterized protein LOC101244019	0.03	N	1.46	1.48	3.71
AT1G55580	LAS	0001820	scarecrow-like protein 28	2.58E-50	N	10.83	13.6	4.89
AT1G79580	SMB	0015551	no apical meristem family protein	1.04E-80	Y	5.84	6.09	3.13
AT1G26870	FEZ	NSH	NSH	NSH	NSH	0.1	0.1	0.1
AT1G56010	NAC1	0016854	nac domain-containing protein 21 22-like	2.92E-70	Y	26.76	33.91	47.81
AT2G04160	AIR3	0002155	subtilisin-like protease-like	0	N	6.59	10.36	37.91



AT3G25710	TMO5	0015249	transcription factor bhlh30-like	1.25E-64	N	7.38	6.47	4.98
AT5G42800	DFR	0031589	dihydroflavonol reductase	4.90E-104	Y	0.38	0.36	1.38
AT4G30080	ARF16	0001279	auxin response factor 18- like	0	Y	5.09	5.04	4.42
AT4G25470	CBF2	0025501	dehydration-responsive element-binding protein	1.90E-45	N	1.11	1.73	1.02
AT1G04240	SHY2 (IAA3)	0014302	auxin-induced protein 22b	1.21E-70	N	35.46	36.07	16.64
AT4G37650	SHR	0003271	protein short-root	5.85E-172	Y	15.86	18.57	12.14
AT5G03150	JKD	0003299	zinc finger protein jackdaw-like isoform x1	3.80E-96	Y	15.2	16.09	12.09
AT3G54220	SCR	0003217	protein scarecrow-like	0	Y	16.24	26.26	18.73
AT1G03840	MGP	0005022	zinc finger protein magpie-like	8.02E-127	Y	9.11	14.27	11.97
AT1G50420	SCL3	0003587	scarecrow-like protein 3	0	Y	12.54	13.71	17.92
AT5G11260	HY5	0015206	transcription factor hy5- like	1.04E-46	Y	49.43	33.1	27.33
AT4G32880	ATHB8	0000849	homeobox-leucine zipper family protein	0	Y	12.14	14.87	11.42
AT4G14550	IAA14	0018103	auxin-induced protein aux28	6.06E-99	Y	343.11	351.74	185.48
AT5G10510	AIL6	0003496	ap2-like ethylene- responsive transcription factor ail6-like	5.36E-115	Y	23.18	27.88	14.55
AT1G19850	MP	0000685	transcriptional factor b3 family protein auxin- responsive factor aux iaa- related	0	Y	5.84	8.35	33.42
AT3G20840	PLT1	0004605	ovule development family protein	0	Y	15.14	23.61	9.93
AT1G04550	IAA12	0009934	auxin-responsive protein iaa11-like	6.86E-58	N	22.94	32.12	39.07
AT1G12980	ESR1	0012720	ethylene-responsive	1.54E-16	N	21.46	33.81	55.6

			transcription factor 4-like						
AT5G61420	MYB28	0035249	transcription factor myb35	3.00E-68	N	0.41	0.73	1.62	
AT2G20610	SUR1	0007195	tyrosine aminotransferase-like	7.98E- 165	N	27.15	29.93	35.17	
AT3G12280	RBR1	0001201	retinoblastoma-related protein	0	Y	18.44	21.99	19.38	
AT2G33880	WOX9	0045416	wuschel-related homeobox 11-like	1.43E-22	N	1.15	0.48	0.54	
AT1G53230	TCP3	0003653	transcription factor tcp4- like	4.30E-70	N	2.99	3.01	3.47	
AT1G19220	ARF19	0000777	auxin response factor 19- like	0	Y	9.5	13.97	12.13	
AT2G42430	LBD16	0023003	lob domain-containing protein 16	4.33E-65	Y	1.6	1.25	9.77	
AT3G58190	LBD29	0010011	lob domain-containing protein 29-like	9.93E-63	N	3.23	4.69	9.43	
AT2G20180	PIL5	0021486	transcription factor pif3- like	1.29E-26	N	0.85	1.33	1.16	
AT1G03430	AHP5	0018863	histidine-containing phosphotransfer	2.51E-56	N	47.56	43.54	74.89	
AT5G54510	DFL1	0006889	indole-3-acetic acid-amido synthetase	0	Y	22.05	32.66	83.56	
AT5G20730	ARF7	0000777	auxin response factor 19- like	0	Y	12.14	14.87	11.42	
AT3G15540	IAA19	0015499	auxin-induced protein aux22	3.88E-50	Y	8.92	10.69	8.52	
AT3G50340		0010064	PREDICTED: uncharacterized protein LOC100262755	3.12E- 165	Y	3.36	4.24	2.85	
AT2G45420	LBD18	0044113	lob domain-containing protein 18-like	3.58E-67	N	0.36	0.68	1.73	
AT4G14560	IAA1	0014302	auxin-induced protein 22b	6.76E-53	N	35.46	36.07	16.64	
AT4G37390	YDK1	0004749	probable indole-3-acetic acid-amido synthetase	0	Y	10.5	18.51	17.66	

AT4G27260	WES1	0006889	indole-3-acetic acid-amido synthetase	0	Y	22.05	32.66	83.56
AT5G17430	BBM	0002613	ap2-like ethylene-responsive transcription factor bbm-like	1.34E-122	Y	9.55	13.02	6.82
AT5G03740	HD2C	0009674	histone deacetylase hdt1-like	1.02E-24	N	139.84	161.42	92.97
AT5G22650	HD2B	0027114	histone deacetylase isoform 3	8.21E-26	N	286.77	270.64	131.72
AT5G18560	PUCHI	0012720	ethylene-responsive transcription factor 4-like	3.71E-23	N	21.46	33.81	55.6
AT5G37020	ARF8	0001138	auxin response factor 8-like	0	Y	9.71	9.76	5.8
AT1G28130	GH3	0004939	indole-3-acetic acid-amido synthetase -like	0	Y	7.5	7.25	14.31
AT5G60890	MYB34	0035249	transcription factor myb34	1.30E-67	N	0.41	0.73	1.62
AT4G00220	LBD30	0044113	lob domain-containing protein 18-like	5.91E-59	Y	0.36	0.68	1.73
AT2G01420	PIN4	0002917	auxin efflux carrier component 3-like	0	Y	1.66	2.17	5.65
AT1G70940	PIN3	0002917	auxin efflux carrier component 3-like	0	Y	1.66	2.17	5.65
AT1G23080	PIN7	0002917	auxin efflux carrier component 3-like	0	Y	1.66	2.17	5.65

## 7. PUBLICACIONES DERIVADAS DE ESTA TESIS

a) **Gustavo Rodriguez-Alonso**, Marta Matvienko, Mayra L. López-Valle, Pedro E. Lázaro-Mixteco, Selene Napsucialy-Mendivil, Joseph G. Dubrovsky & Svetlana Shishkova (2018) Transcriptomic insights into the genetic regulation of root apical meristem exhaustion and determinate primary root growth in *Pachycereus pringlei* (Cactaceae). Scientific Reports 8, 8529. DOI: 10.1038/s41598-018-26897-1

www.nature.com/scientificreports

# SCIENTIFIC REPORTS

OPEN

## Transcriptomics insights into the genetic regulation of root apical meristem exhaustion and determinate primary root growth in *Pachycereus pringlei* (Cactaceae)

Received: 11 January 2018

Accepted: 17 May 2018

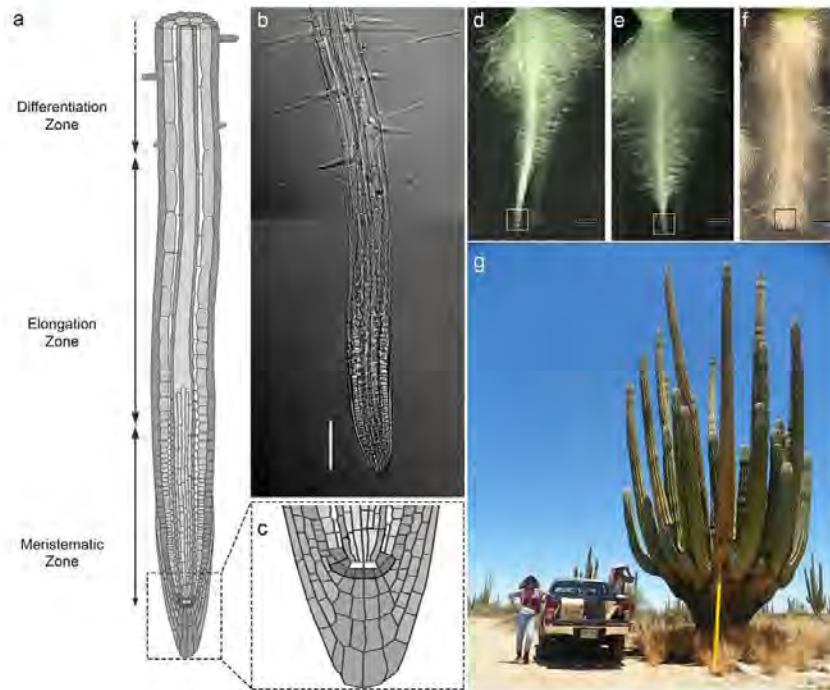
Published online: 04 June 2018

Gustavo Rodriguez-Alonso<sup>1</sup>, Marta Matvienko<sup>2</sup>, Mayra L. López-Valle<sup>1</sup>, Pedro E. Lázaro-Mixteco<sup>1</sup>, Selene Napsucialy-Mendivil<sup>1</sup>, Joseph G. Dubrovsky<sup>1</sup> & Svetlana Shishkova<sup>1</sup>

Many Cactaceae species exhibit determinate growth of the primary root as a consequence of root apical meristem (RAM) exhaustion. The genetic regulation of this growth pattern is unknown. Here, we *de novo* assembled and annotated the root apex transcriptome of the *Pachycereus pringlei* primary root at three developmental stages, with active or exhausted RAM. The assembled transcriptome is robust and comprehensive, and was used to infer a transcriptional regulatory network of the primary root apex. Putative orthologues of *Arabidopsis* regulators of RAM maintenance, as well as putative lineage-specific transcripts were identified. The transcriptome revealed putative orthologues of most proteins involved in housekeeping processes, hormone signalling, and metabolic pathways. Our results suggest that specific transcriptional programs operate in the root apex at specific developmental time points. Moreover, the transcriptional state of the *P. pringlei* root apex as the RAM becomes exhausted is comparable to the transcriptional state of cells from the meristematic, elongation, and differentiation zones of *Arabidopsis* roots along the root axis. We suggest that the transcriptional program underlying the drought stress response is induced during Cactaceae root development, and that lineage-specific transcripts could contribute to RAM exhaustion in Cactaceae.

Plant growth and organogenesis are postembryonic processes sustained by the presence and activity of meristems, which act as reservoirs of pluripotent cells. The root apical meristem (RAM) contains cells with a high mitotic rate, which supply cells for organ growth<sup>1</sup>. After few cell divisions in the RAM, also known as meristematic zone, the cells are displaced from the root apex and enter the elongation zone, where they undergo a longitudinal expansion. The generation and elongation of cells at the meristematic and elongation zone, respectively, further displace the cells into the differentiation zone, where the root cells acquire distinctive features of particular root tissues (Fig. 1a–c). The RAM is present and active in most roots throughout much of the plant's lifecycle in the majority of angiosperms, and it is therefore generally assumed that root growth can continue indefinitely as long as the environmental conditions are suitable<sup>2</sup>. A less common growth pattern for roots is the determinate growth, in which all cells at the root apex cease to divide and become differentiated as a consequence of RAM exhaustion (Fig. 1d–f). Determinate growth has been documented in a variety of angiosperm species, most remarkably in the proteoid lateral roots of Proteaceae species<sup>3,4</sup>; the adventitious roots of *Ficus pumila*<sup>5</sup>; and the lateral roots of *Zea mays*<sup>6</sup>, *Opuntia arenaria*, and *O. tunicata*<sup>7</sup>. Determinate growth of the primary root has only been reported for Cactaceae species from the Cactoideae and Opuntioideae subfamilies, and *Ferocactus peninsulae*, *Stenocereus thurberi*, *S. gummosus*<sup>8</sup>, and *Pachycereus pringlei*<sup>9</sup> were the first species in which determinate growth of the primary root was studied. This growth pattern has subsequently been observed in species from the seven tribes of the Cactoideae subfamily, namely, Cacteeae, Pachycereae, Cereae, Trichocereae, Notocactae, Rhipsalideae<sup>10</sup>, and Hylocereae<sup>10,11</sup>. Seedlings of all analysed species from the first six tribes always exhibit determinate growth

<sup>1</sup>Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico. <sup>2</sup>Tecan Systems, 2450 Zanker Rd, San Jose, CA, 95131, United States. Correspondence and requests for materials should be addressed to S.S. (email: sveta@ibt.unam.mx)



**Figure 1.** Indeterminate and determinate root growth. (a) Growing roots can be divided in three developmental zones along the longitudinal axis as illustrated here for the *Arabidopsis thaliana* primary root; the root apical meristem (RAM) is located in the meristematic zone and it is present and active in most angiosperm roots for long periods. (b) The developmental zones in an *A. thaliana* lateral root can be distinguished by the cell features. Scale bar: 100 µm. (c) The RAM contains a quiescent centre composed of cells with very low mitotic rate (white cells in the scheme). The cells adjacent to quiescent centre, delimited with a thick line, are called stem (initial) cells, and are a source of the dividing cells for the meristem. (d–e): The *Pachycereus pringlei* primary root exhibits determinate growth. In this work, 1 mm of the primary root apex, delimited with white or black box, was collected at three developmental stages: initial, when the RAM is present and fully active (d); intermediate, when the RAM is smaller and the differentiation zone is closer to the root apex (e); and terminal, when the RAM is exhausted and all the cells in the root, including those at the root apex, are differentiated (f). A mature *P. pringlei* is shown in (g). (a) and (c) were taken and modified from Peret, Benjamin (2017): doi:10.6084/m9.figshare.5143987.v4, originally deposited on FigShare as open access content under a CC BY 4.0 license.

of the primary root. Interestingly, some individuals of *Epiphyllum phyllanthus*, an epiphytic species from the Hylocereeae tribe, which inhabits mesic environments, exhibit determinate growth<sup>11</sup>, while other seedlings from the same species exhibit indeterminate growth<sup>10</sup> of the primary root. Cactoideae species either thrive in arid or semi-arid environments, or experience water deficit by exposure to air, such as the epiphytic species of the Rhipsalideae and Hylocereeae tribes. Determinate growth of Cactaceae primary root could therefore represent an evolutionary adaptation to severe drought.

Root hair development as a consequence of early termination of root growth enables faster water uptake and favours seedling establishment; moreover, in *Stenocereus* species, RAM exhaustion induces lateral root initiation and development<sup>8</sup>. In all desert Cactaceae species studied so far, determinate growth can also be observed in at least the first- and second-order lateral roots. Cactaceae species exhibit determinate growth of the primary root independent of the growth condition<sup>8</sup>. These data, together with the incidence of RAM exhaustion in lateral roots and the determinate growth of roots regenerated from calli<sup>12</sup>, indicate that RAM exhaustion in Cactaceae species is a genetically regulated developmental program, rather than a response to environmental conditions. The molecular mechanisms underlying RAM exhaustion in the Cactaceae primary root are unknown, although suppression subtractive hybridization has been used to identify some of the differentially expressed genes between the *S. gummosus* primary root apex with functional and exhausted RAM<sup>10</sup>.

Our knowledge of the genetic regulation of RAM establishment and maintenance is mainly derived from studies of *Arabidopsis thaliana* (*Arabidopsis*) mutants with short root phenotypes. These studies led to the identification of three major genetic regulatory pathways. The most important pathway for RAM maintenance involves the



Sequencing and preprocessing		Assembly		Contig features	
Total reads	279.23 × 10 <sup>6</sup>	Total contigs	49,045	Average length (nt)	1,080
Av. reads per sample	46.5 × 10 <sup>6</sup>	N25	1,332	Max. length (nt)	14,583
Paired-end reads	79 × 10 <sup>6</sup>	N50	726	Min. length (nt)	400
Merged reads	88 × 10 <sup>6</sup>	N75	611	400–2,000 nt	87.63%
				2,000–4,000 nt	10.94%
				>4,000 nt	1.43%

**Table 1.** *De novo* assembled transcriptome of the *Pachycreus pringlei* primary root apex.

PLETHORA (PLT) transcription factors, which belong to the AP2/ERF superfamily<sup>13</sup>. PLT proteins are partially redundant, their activities are dose-dependent, and they share a high number of target genes<sup>13</sup>. Four of the six *Arabidopsis* PLT proteins, PLT1, PLT2, PLT3, and PLT4 (also known as BBM), have similar expression patterns in the root apex, with expression maxima in the quiescent centre (QC), very high expression levels in the stem (initial) cells, strong expression in the rest of the RAM, lower levels of expression in the elongation zone, and low to no expression in the differentiation zone<sup>13</sup>. The primary root of the *plt1* mutant of *Arabidopsis* exhibits subtle alterations of the cell division pattern in the QC and root cap cells, while *plt1 plt2* double mutants develop a much shorter primary root that stops growing soon after germination due to RAM exhaustion, therefore exhibiting determinate growth<sup>13</sup>.

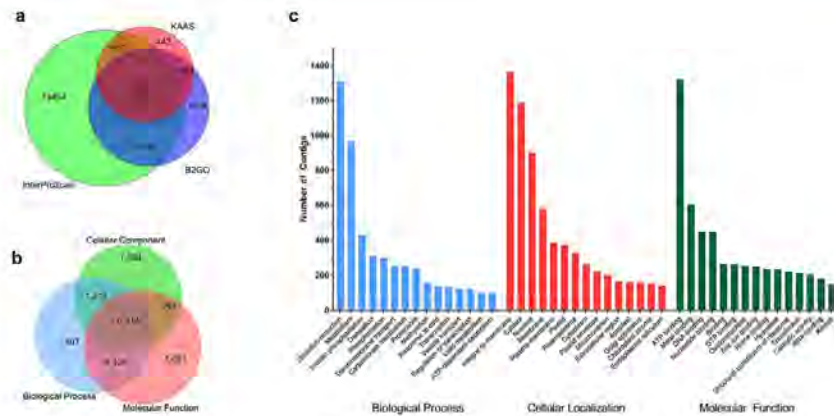
In addition to their well-characterized roles in the radial organization of the root, the GRAS transcription cofactors SCARECROW (SCR) and SHORT-ROOT (SHR) are also involved in RAM maintenance. *Arabidopsis scr* and *shr* loss-of-function mutants exhibit a shorter primary root compared with the wild type, as well as changes in QC identity<sup>16,17</sup>. As a result, the RAM becomes disorganized and is lost, causing determinate growth<sup>16–20</sup>. Another transcription factor specifically expressed in the QC and involved in stem cell homeostasis in the root is WUSCHEL-RELATED HOMEODOMAIN 5 (WOX5)<sup>21,22</sup>. In *Arabidopsis wox5* loss-of-function mutants the columella stem cells differentiate and the QC cells are larger than those of the wild type, while the inducible ectopic expression of WOX5 prevents differentiation of columella and lateral root cap cells<sup>20,21</sup>. When combined with *scr*, *shr*, or *plt1plt2* mutations, the *wox5* mutation redundantly accelerates RAM exhaustion<sup>21</sup>.

The exploration of RAM exhaustion and other developmental processes in Cactaceae has been limited by the lack of reference genomes for these species, although, after completion of this work, early draft genomes of four columnar cacti were reported<sup>24</sup>. Nevertheless, the rapid improvement of RNA-seq as a high-throughput sequencing technology enables the *de novo* assembly of transcriptomes for non-model species, circumventing the need for a reference genome<sup>25</sup>. In this study, we used RNA-seq to generate a *de novo* assembled transcriptome of the *P. pringlei* primary root apex at three developmental stages. To explore gene conservation in Cactaceae species, this transcriptome was compared with recently published *Lophophora williamsii* transcriptome<sup>26</sup>. The *P. pringlei* root apex transcriptome was used to assess differential gene expression and to infer a transcriptional regulatory network, with the aim of expanding our knowledge of RAM exhaustion in Cactaceae species and, consequently, enhancing our understanding of RAM maintenance in angiosperms.

## Results and Discussion

**Transcriptome *de novo* assembly and annotation.** To examine the molecular mechanisms underpinning RAM exhaustion in Cactaceae, we *de novo* assembled the root apex transcriptome of the *P. pringlei* primary root at three developmental stages: i) initial, when the RAM was present, at around 1–4 days after germination (d.a.g.); ii) intermediate, when a smaller RAM was still present and the root hairs were closer to the root apex, at approximately 5–6 d.a.g.; and iii) terminal, when the RAM was exhausted and all cells of the root apex were differentiated, at 9–13 d.a.g. (Fig. 1d–f). RNA-seq of the six samples, two biological replicates per developmental stage, yielded 279.23 million 2 × 100-bp paired-end reads with an average of 46.5 million reads per sample. The processed reads (see Methods) from all samples were used to *de novo* assemble the *P. pringlei* root apex transcriptome. The transcriptome was assembled in two consecutive rounds, first with all the processed reads and mapping reads back to contigs. The unmapped reads were then used to assemble additional contigs under the same parameters of bubble- and word-size as the previous run. This approach allowed the recovery of highly similar sequences, which were expected for large gene families and allelic variants, as *P. pringlei* is a tetraploid species<sup>27</sup>. The assembled transcriptome included 49,045 contigs, hereafter referred as transcripts, of ≥400 nt, with an average length of 1,080 nt (Table 1). Transcripts ranging from 400 to 2,000 nt in length account for 87.63% of the total assembled sequences, while 10.94% and 1.43% of the total transcripts were 2,000–4,000 nt and >4,000 nt long, respectively.

A total of 31,265 contigs (63.7%) were annotated using the CLC GW BLAST2GO plug-in, InterProScan, or KAAS<sup>28</sup> tools (Fig. 2). The species from which the highest number of significant hits was retrieved was *Beta vulgaris* (see Supplementary Fig. S1). This was expected as both, *B. vulgaris* and *P. pringlei*, belong to the order Caryophyllales and the former species was the closest relative of *P. pringlei* with an available reference genome<sup>29</sup> at the time of the analysis. Shortly after this study, the genomes of the Caryophyllales *Amaranthus hypochondriacus*<sup>30</sup> and *Chenopodium quinoa*<sup>31</sup> were released; however, they were not available from the RefSeq database and therefore could not be included in the annotation analysis. The results of the KAAS annotation showed that the assembled transcriptome contained all the putative orthologue sequences required to reconstruct the central metabolism and CAM carbon fixation pathways (see Supplementary Fig. S2). Furthermore, the root apex transcriptome contains putative orthologues of most genes associated with the following: plant hormone signalling pathways, including auxin, cytokinin, gibberellin, abscisic acid, ethylene, brassinosteroid, jasmonic acid, and



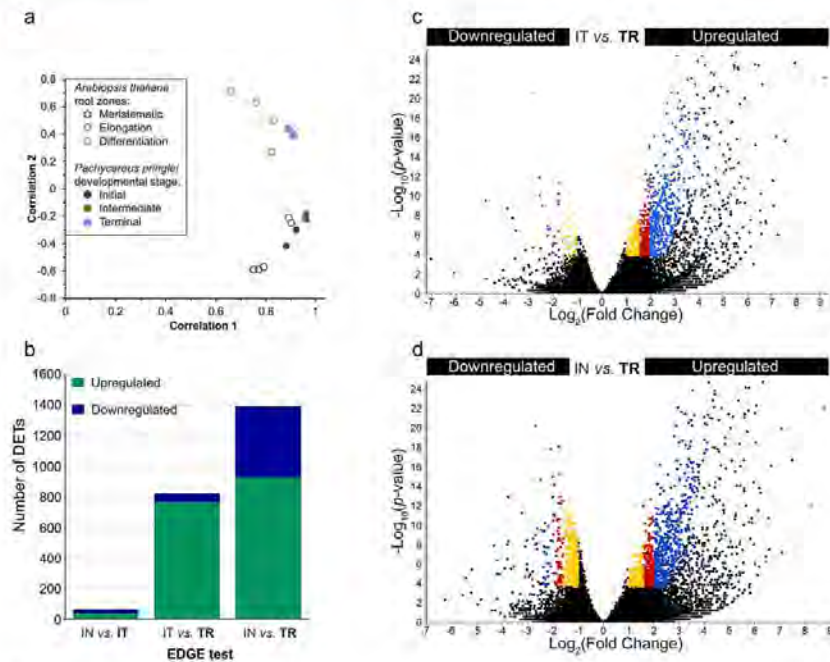
**Figure 2.** Annotation of the *de novo* assembled root tip transcriptome of *Pachycereus pringlei*. (a) 63.7% of the total assembled contigs were annotated using the KEGG Automatic Annotation Server (KAAS), InterProScan, or Blast2GO tools. (b) The number of contigs annotated per B2GO category. (c) The most frequently represented gene ontology subcategories.

salicylic acid; eukaryotic DNA replication; basal eukaryotic transcription factors; synthesis, surveillance, and degradation of mRNA; ribosome biogenesis; and protein processing in the endoplasmic reticulum (Supplementary Fig. S2). Thus, the assembled root apex transcriptome is not only robust but also comprehensive and represents a step forward in the study of the Cactaceae family, for which the available transcriptomes generated so far include expressed sequence tags<sup>32</sup> or specific RNA-seq datasets exploring metabolic processes<sup>26,33</sup>. The transcriptomes previously available for other plant species from the order Caryophyllales were also limited, but were recently expanded by the efforts of the OneKP initiative<sup>34</sup>.

**Transcriptional dynamics of the Cactaceae root development.** To assess transcript abundance, the assembled transcripts were used as reference sequences to separately map the reads from each sample. Then, a principal component analysis (PCA) was performed on the correlation matrix to evaluate sample variability, and to visualize the transcriptional dynamics of the *P. pringlei* root apex at different developmental stages. The PCA analysis showed that the first and second principal components explain 84% and 12% of the observed variability of the data, respectively (see Supplementary Fig. S3). As shown in Fig. 3a, the PCA distinguished the terminal developmental stage transcriptomes from those of the initial and intermediate developmental stages. This was likely because all cells of the primary root apex are differentiated at the terminal developmental stage, while both, meristematic and elongating cells are present at the initial and intermediate stage, although in different proportions. The intermediate stage samples exhibited the least variability, as appreciated by their proximity to each other on the PCA plot, while the samples from the initial stage exhibited the greatest variability. This could be explained because the samples of the initial growth stage include root tips from 1–4 d.a.g. seedlings, resulting in the most heterogeneous biological replicates used in this study. The separation of the root apex samples according to the developmental stage, however, indicates that the observed variability between samples is due to transcriptional state changes during primary root development, and that specific transcriptional programs operate in the *P. pringlei* root apex at particular developmental time points. In contrast, in most angiosperms the RAM maintenance programs remain present in the root apex of the majority of roots.

To compare the transcriptional states of the *P. pringlei* root apex with those of *Arabidopsis*, the raw RNA-seq reads from the meristematic, elongation, and differentiation root zones of *Arabidopsis*<sup>35</sup> were used. A PCA was performed on the *Arabidopsis* data and the results were visualized on the same plot as the *P. pringlei* data (Fig. 3a). The transcriptomes from the *P. pringlei* intermediate and terminal developmental stages were found to group with the transcriptomes of the *Arabidopsis* elongation and differentiation root zones, respectively. The initial stage transcriptome was positioned between those of the meristematic and elongation root zones, probably because the *P. pringlei* root apex in this stage includes cells from both the meristematic and elongation zones. The PCA results also suggest that the differentiated cells of the *P. pringlei* root apex are functional after RAM exhaustion, and that they could perform the typical functions of cells from the differentiation zone of other angiosperm roots. This observation is in congruence with our previous results, which indicated that RAM exhaustion in *S. gummosus* and *P. pringlei* does not involve programmed cell death<sup>16</sup>. The slight shift of the transcriptome from the *P. pringlei* terminal stage towards those of the *P. pringlei* intermediate stage and *Arabidopsis* elongation zone suggests that the *P. pringlei* root apex cells retain some expression programs from their previous, less differentiated state; however, this needs to be further explored. Together, these results indicate that cell differentiation along the primary root axis occurs at two different regions during *P. pringlei* primary root development: distal from the RAM, as in other angiosperms, but also at the very root apex as the RAM becomes exhausted.



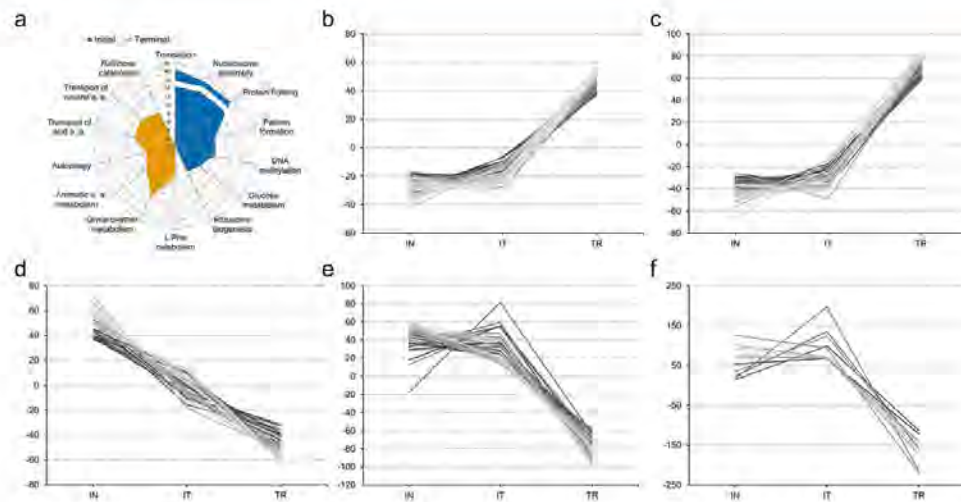


**Figure 3.** Characterization of the *Pachycreus pringlei* transcriptome. (a) Principal component analysis of *P. pringlei* developmental stages and *Arabidopsis* root zones and (b–d) the empirical detection of gene expression shows that the transcriptional state of the *P. pringlei* root tip changes during development. Only 66 differentially expressed transcripts (DETs) were identified in the comparison of the Initial vs. Intermediate developmental stages (b), while 873 and 1,388 DETs were identified between the Intermediate vs. Terminal (b,c), and Initial vs. Terminal (b,d) stages. Colors in c and d correspond to the fold change (FC): blue,  $FC > 4$ ; red,  $3 < FC < 4$ ; yellow,  $2 < FC < 3$ . In all cases, the FDR  $p$ -value was  $\leq 0.005$ . IN: Initial stage; IT: Intermediate stage; TR: Terminal stage. Upregulated and downregulated refer to the expression level at the developmental stage shown in bold.

**Differential gene expression.** An extraction of differential gene expression (EDGE) test was performed to identify differentially expressed transcripts (DETs) in the root apex across developmental stages<sup>37</sup>. Transcripts were considered to be DETs when they met these criteria: fold change  $\geq 2$ , FDR  $p$ -value  $\leq 0.005$ , and RPKM average value  $> 3$  at each of the developmental stages under comparison. As shown in Fig. 3b, with this stringency level, only 66 DETs were identified when comparing the Initial vs. Intermediate stage transcriptomes. By contrast, the numbers of DETs between the Intermediate vs. Terminal and Initial vs. Terminal stage transcriptomes were 873 and 1,388, respectively (Fig. 3b–d). Therefore, the number of DETs increased with the progression of root development. Since very few DETs were detected between the initial and intermediate stages, only the last two comparisons are shown in Fig. 3c,d. For both the Intermediate vs. Terminal and Initial vs. Terminal comparisons, the number of upregulated transcripts at the terminal stage was higher than the number of downregulated transcripts, with approximately 14 times and 2 times more upregulated transcripts, respectively (Fig. 3b). Hence, during the transition from a population of mainly elongating to differentiated cells, many genes are upregulated, while downregulation of expression is observed in far fewer genes. It is generally assumed that in fully differentiated cells, more genes are upregulated in comparison to undifferentiated or partially differentiated cells. This assumption is supported by data retrieved from Huang and Schiefelbein<sup>35</sup> (see Supplementary Table S1). Surprisingly, among the DETs between the initial and terminal developmental stages, the number of upregulated genes was only twice that of downregulated genes.

A hypergeometric test<sup>38</sup> based on the gene ontology (GO) biological process annotation was performed for the DETs. The results, shown in Fig. 4a, suggest that as the root apex cells transition from meristematic to fully differentiated cell states, processes such as protein synthesis and folding, DNA methylation, nucleosome assembly, and ribosome biogenesis are downregulated. By contrast, overrepresented GO subcategories at the terminal developmental stage include amino acid transport; metabolism of aromatic compounds and glycerol ether; autophagy; and stress responses. Furthermore, the DETs were grouped according to their RPKM values using the  $k$ -mean clustering method<sup>39</sup>. A total of 27 clusters were obtained, which can be grouped into three main expression profiles as shown in Fig. 4b–f: transcripts upregulated at the terminal developmental stage (Fig. 4b–c); transcripts





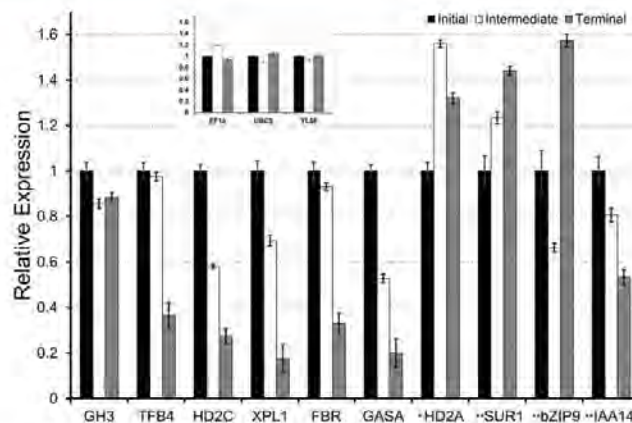
**Figure 4.** Expression profiles of the differentially expressed transcripts in the *Pachycereus pringlei* root apex. (a) BLAST2GO biological processes over-represented in the Initial and Terminal developmental stages of the primary root apex of *P. pringlei*. Axis:  $-\log_{10}(p\text{-val})$  of the enrichment test. (b–f) Representative expression profiles. Some of the functionally related transcripts from distinct GO categories could be grouped within a number of expression profiles. For example, several transcripts from the autophagy or amino acid transport GO categories, among others, are included in (b); the amino acid metabolism, amino acid transport, and glycerol-ether metabolism categories are included in (c); nucleosome assembly, translation, and regulation of DNA methylation, are included in (d); protein folding, protein deubiquitylation, and rRNA processing, are included in (e); some transcripts from the translation, and response to oxidative stress categories are included in (f). (b–f): Numbers on the y-axes correspond to the normalized and centered RPKM values for the Initial (IN), intermediate (IT), and Terminal (TR) growth stages.

downregulated at the terminal stage (Fig. 4d–e); and one cluster of transcripts upregulated at the intermediate developmental stage (Fig. 4f; the complete set of clusters is depicted in Supplementary Fig. S4). Clustering analysis is useful to identify sets of genes that could be involved in the same biological process, or influenced by the same transcriptional regulator. Remarkably, several transcripts predicted to encode 60S or 40S ribosomal proteins were grouped in the same cluster and were downregulated as root development progressed and the root apex cells differentiated. Among the transcripts downregulated at the terminal developmental stage, those encoding putative histones, histone deacetylases, some peroxidases, and gibberellin-regulated proteins (for example, *GASA1* and *GASA4/6* putative orthologues) were retrieved. Among the transcripts upregulated at the terminal developmental stage, there was an overrepresentation of abscisic acid induced transcripts, encoding members of the abscisic stress ripening protein family and abscisic acid receptors. This result suggests that the genetic programs to cope with drought stress are induced in the differentiated cells of the Cactaceae root in any growth condition, as the plants used for the transcriptome analyses were cultivated in Petri dishes and did not experience water deficit. Other DETs upregulated in the terminal developmental stage included transcripts encoding glutathione transferases; proline-rich cell wall proteins; ethylene-responsive, NAC-domain, and WRKY transcription factors; and proteins involved in auxin-related processes, such as ARFs and SAURs.

Remarkably, 17.4% (244) of the DETs could not be annotated and may represent lineage-specific transcripts. Moreover, a significant number of the DETs with a higher abundance in the terminal stage were annotated as uncharacterized proteins (7%) or hypothetical proteins (~2%), with significant hits in a variety of species, including *Populus trichocarpa* and *Phaseolus vulgaris*.

**Conservation of Cactaceae genes and identification of lineage specific transcripts.** To evaluate the conservation of genes in the Cactaceae, we used the previously reported transcriptome of *L. williamsii*<sup>20</sup>. The *L. williamsii* shoot + root transcriptome consists of 40,436 unigenes ranging in length from 200 to 4,170 nt, which were compared with the *P. pringlei* root tip transcriptome (Supplementary Table S2, Supplementary Fig. S5). A high degree of conservation was revealed among the nucleotide sequences of both *de novo* assembled transcriptomes (Supplementary Table S3), with a median E-value of  $2.25 \times 10^{-118}$ . Multiple hits from *L. williamsii* were retrieved for most *P. pringlei* query sequences, as the *P. pringlei* contigs were significantly longer (Supplementary Fig. S5; Supplementary Tables S2 and S3).

Next, to explore whether some of the 244 unannotated *P. pringlei* DETs in the Initial vs Terminal comparison could represent Cactaceae-specific transcripts, we performed a BLASTn search of these DETs against the *L. williamsii* unigenes, revealing significant hits for 53% (130) of them. Furthermore, significant hits for some of



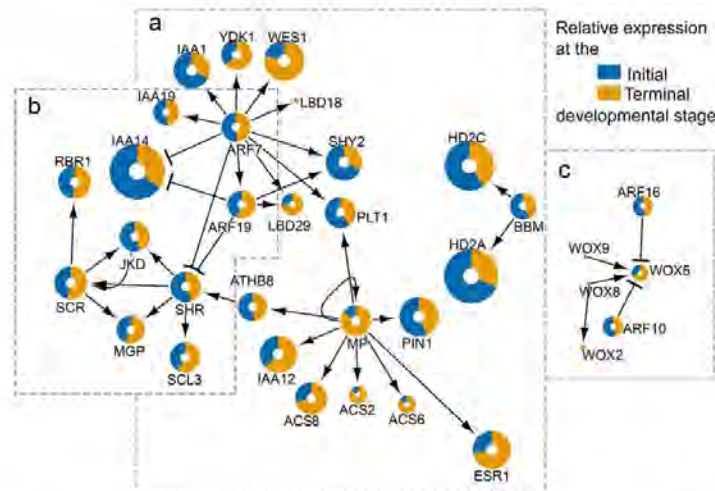
**Figure 5.** The relative expression levels of selected *Pachycereus pringlei* transcripts quantified using RT-qPCR assays for the primary root apex at the three stages of primary root development. The expression levels were normalized relative to *PpEF1α*, *PpUBC9*, and *PpYLS8*, respectively (inset). The trend for RPKM values and transcript abundance evaluated by qRT-PCR between initial and terminal development stages was different for this transcript. \*\*Two biological replicates were used for these three genes, and three biological replicates were used for another ten genes.

the unannotated DETs were also identified in the shoot cambial zone transcriptomes of other Cactaceae species, namely, *Ariocarpus retusus*, *Echinocactus platyacanthus*, *Ferocactus pilosus*, and *Pereskia lychnidiflora* (E. Petrone and T. Terrazas, pers. comm.). Of the DETs that could not be annotated, 163 (66%) contain ORFs encoding putative peptides longer than 50 aa. Of these, 10, 45, and 108 putative peptides were predicted to be >200 aa, 100–200 aa, and 50–100 aa long, respectively (Supplementary Fig. S6). These results strongly suggest the existence of lineage-specific transcripts and their possible involvement in the regulation of RAM exhaustion and determinate root growth in Cactaceae. This assumption requires further exploration as it could lead to the identification of novel root development regulatory molecules, opening up new opportunities for evo-devo research.

**RT-qPCR validation of expression profiles.** In our previous work, several DETs downregulated at the terminal growth stage in *S. gummosus* primary root apices were identified by suppression subtractive hybridization<sup>10</sup> and confirmed using RT-qPCR. We therefore used RT-qPCR to measure the relative expression of putative orthologues of these DETs in *P. pringlei* root tips. First, a set of candidate reference genes was selected by identifying the putative *P. pringlei* orthologues of the *Arabidopsis* superior reference genes<sup>40</sup>. The list was reduced by selecting the *P. pringlei* transcripts with the lowest RPKM variation in the root apex at the initial, intermediate, and terminal stages of primary root development (see Supplementary Table S4, Supplementary Fig. S7). *PpEF1α*, *PpUBC9*, and *PpYLS8*, were selected as internal references. The expression patterns of *PpTFB4*, *PpHD2C*, *PpXPL1*, *PpFBR*, *PpGASA*, were evaluated and the downregulation in the *P. pringlei* root apex at the terminal stage was confirmed by transcriptome and RT-qPCR analyses (Fig. 5, Supplementary Fig. S8). This suggests that the mechanisms of determinate root growth are conserved at least within the Pachycereae tribe of the subfamily Cactoideae. Furthermore, RT-qPCR analysis confirmed the trend in expression level throughout root development of seven more genes, including the three selected reference genes; *PpGH3*, a stably expressed transcript across root development; as well as, *PpIAA14*, downregulated at the terminal developmental stage; *PpZIP9*, and *PpSUR1* upregulated at the terminal developmental stage. With the exception of *PpHD2A*, the expression profiles of 12 of the 13 tested genes were qualitatively reproduced using RT-qPCR, therefore validating the reliability of the *P. pringlei* RNA-seq (Fig. 5, Supplementary Fig. S8).

**Transcriptional regulatory network for the *P. pringlei* RAM.** Cellular processes, either during development or in response to environmental stimuli, are not controlled by single genes, but by an intricate network connecting multiple elements, including but not limited to transcription factors and their target genes<sup>41</sup>. We therefore generated a transcriptional regulatory network (TRN) by retrieving a set of genes that were reported as important regulators of root development<sup>35,42</sup>. These genes and their first neighbours, together with their interactions, were extracted from the *Arabidopsis* transcriptional regulatory map<sup>43</sup>. The putative *P. pringlei* orthologues of these genes were identified by tBLASTn using best bidirectional hit (BBH) analysis. An interaction table for the retrieved orthologous *P. pringlei* nodes, and a table of properties for the *P. pringlei* nodes were generated based on the interactions of the *Arabidopsis* network (Supplementary Table S5, Supplementary Table S6). The resulting TRN inferred for the *P. pringlei* root apex included 81 nodes and 111 edges (see Supplementary Fig. S9). From these, 48 nodes corresponded to *Arabidopsis*-*P. pringlei* BBH, while 30 nodes were retrieved as unidirectional hits. No *P. pringlei* orthologues were identified for three *Arabidopsis* nodes, *YAB1*, *FEZ*, and *LFY*, probably due to their low or null expression levels in the *P. pringlei* root apex. The network was divided into the subnetworks shown in





**Figure 6.** Root development modules of the *Pachycereus pringlei* inferred transcriptional regulatory network. The complete network (Fig. S9) was separated into three distinct modules: (a) PpPLT, which includes PpPLT1 and PpBBM as separate clusters; (b) PpSHR-PpSCR; and (c) PpWOX. Node size represents the average expression level ( $\log_2(\text{RPKM})$ ) of the contig in the *P. pringlei* transcriptome.

Fig. 6, which included orthologues of proteins involved in the three main *Arabidopsis* RAM regulatory pathways. The PpPLT subnetwork (Fig. 6a) was represented by PpPLT1 and PpBBM, which were included in a module that also contained their interaction partners, including ARF proteins and histone deacetylases. At the time of the analysis, only these two PLT members were included in the *Arabidopsis* regulatory map. The recent identification of the target genes of distinct PLTs<sup>14</sup> will provide valuable insight into the intricate regulation of RAM maintenance by these transcription factors, and would allow the inclusion of additional nodes and interactions to the *P. pringlei* TRN. The second subnetwork included the GRAS-domain transcription cofactors PpSHR and PpSCR (Fig. 6b), as well as the transcription factor PpJKD. Recent reports show that SHR, SCR, and JKD form a macromolecular complex that regulates the transcription of different genes according to their cellular context<sup>44,45</sup>; therefore, the apparent expression stability of these nodes in the *P. pringlei* root apex during root development might be explained by their involvement in developmental processes, such as radial pattern formation<sup>46</sup>, vascular differentiation<sup>46</sup>, and transition from cell division to cell elongation<sup>47</sup>, in addition to their roles in RAM maintenance. The PpSHR and PpSCR subnetwork also included other interactors, such as PpMGP, and PpRBR orthologues of *Arabidopsis* genes that have been demonstrated to be important regulators of RAM maintenance<sup>46,49</sup>. The third subnetwork was composed of genes encoding four WOX proteins and their ARF interacting partners (Fig. 6c). However, the expression level of the putative WOX orthologues was very low, with average expression levels  $<4$  RPKM across the developmental stages of the *P. pringlei* root apex. The low RPKM values obtained for PpWOX are expected: for example, the *Arabidopsis* WOX5 is expressed specifically in the QC cells<sup>21</sup>, while the samples for RNA-seq data presented here included a broader cell population.

In conclusion, our results demonstrate that the transcriptional programs operating in the *P. pringlei* root apex are similar to those of the *Arabidopsis* root zones. The described differences in expression level of RAM regulators in the root apex of the cactus across developmental stages suggest that the acquisition of the determinate growth as an adaptive trait might have arose as a result of changes in regulatory sequences. This assumption requires further exploration, which will be possible upon the improvement of the assembly and annotation of the recently published draft nuclear genomes of Pachycereaceae tribe species, with a high-coverage genome of *Carnegiea gigantea* and low-coverage genome of three more columnar cacti, including *P. pringlei*<sup>24</sup>.

## Materials and Methods

**Plant materials.** Fruit of *Pachycereus pringlei* (S. Watson) Britton & Rose were collected from plants (Fig. 1g) growing near Bahía Kino, Sonoran Desert, Mexico. The seeds were sterilized as described<sup>30</sup> and germinated on vertically oriented Petri dishes containing  $0.2 \times$  Linsmaier and Skoog medium (Phyto Technology Laboratories, Lenexa, KS, USA; pH 5.8), 0.8% Bacto Agar (BD Difco, Sparks, MD, USA), with no sucrose added. Plants were grown at 28 °C in a 12/12-h photoperiod with a light intensity of  $80 \mu\text{mol m}^{-2} \text{s}^{-1}$ .

For the primary root apex samples, 1 mm sections of the root tip were detached using a sterilized razor blade and frozen in liquid nitrogen. Root apices at three developmental stages were collected: initial, intermediate, and terminal (Fig. 1d–f). Root hairs, formed by differentiated epidermal cells, served as a hallmark for cell differentiation to distinguish between root developmental stages. Only the apices of roots that grew in contact

with the surface of the medium were used for RNA extraction; roots that grew inside the medium, or away from it, were not sampled. Two biological replicates were used for RNA-seq, and two or three additional biological replicates were used for RT-qPCR analysis. Approximately 350, 450, and 600 root apices were collected for each biological replicate of the initial, intermediate, and terminal developmental stages, respectively. Plant material was stored at -70 °C until RNA extraction.

**RNA extraction, RNA-seq, and *de novo* assembly.** Total RNA was isolated using TRIzol reagent (Thermo-Fisher Scientific, Waltham, MA, USA). All steps of RNA quality analysis, RNA processing, library preparation and 100 cycles of paired-end sequencing were performed at BGI-Tech, Hong Kong. Illumina RNA Library Prep kit was used for cDNA synthesis and Illumina HiSeq 2000 platform was used for sequencing. The quality of the Illumina sequence reads was analysed with CLC Genomics Workbench v 7.5 (CLC GW, CLC bio, Qiagen, Hilden, Germany; <http://www.clcbio.com/>). Read processing was performed using CLC GW as follows: the terminal 13 nt of the 5' end were trimmed, and if both reads from a single RNA fragment overlapped, they were merged. A quality and adaptor trimming step was run on merged and unmerged reads; resulting reads shorter than 40 nt were discarded. The root apex transcriptome assembly was performed several times using the CLC GW *de novo* assembly tool, setting 400 nt as the minimal value for contig length, and varying the word size and bubble size to achieve optimal values, which were 50 and 400 nt, respectively. Unassembled reads were used to perform a second round of assembly using the optimal bubble and word sizes from the previous assembly run. Contigs larger than 3,000 nt were extended using the Genome Finishing Module in CLC GW. Raw reads and assembled contigs were deposited in the NCBI Gene Expression Omnibus under the accession number GSE104832.

**Contig annotation and RNA-seq analysis.** Contig (transcript) sequences were annotated using the CLC GW Blast2GO plugin, the InterPro database, and the KEGG Automated Annotation Service (KAAS<sup>28</sup>). The processed reads obtained for each developmental stage were independently mapped to the contigs generated in the *de novo* assembly, counted, and normalized using the reads per kilobase per million mapped reads (RPKM) method<sup>30</sup>. Principal component analysis (PCA) was performed on the correlation matrix with the RPKM count table for each developmental stage of *P. pringlei* and, separately, for each zone of *Arabidopsis* root<sup>35</sup>. Differential expression was assessed with EDGE test<sup>37</sup> as implemented in the CLC GW. The differentially expressed transcripts (DETs) were extracted according to the following criteria: fold change  $\geq 2$ , FDR *p*-corrected value  $\leq 0.005$ , RPKM  $\geq 3$  in all samples. DETs were subjected to a hypergeometric test<sup>38</sup> and *k*-mean clustered<sup>39</sup> according to their change in expression level across the developmental stages.

**Validation of RNA-seq expression profiles using RT-qPCR.** Total RNA was treated with DNase (Thermo Fischer Scientific) and cleaned using the RNeasy Micro Kit (Qiagen). The first cDNA strand was synthesized using Superscript II Reverse Transcriptase (Thermo Fischer Scientific) and oligo-dT primer, according to the manufacturer's instructions. RT-qPCR was performed on a LightCycler Nano (Roche, Basel, Switzerland) with SYBRGreen (Thermo Fischer Scientific) as fluorescent probe, using 50 ng cDNA for each reaction. Cq values were used to calculate the relative transcript abundance following the described methods<sup>51,52</sup>. The primers used are listed in Supplementary Table S7.

**Comparative transcriptomics and transcriptional regulatory network.** A list of *Arabidopsis* genes previously reported to be important regulators of root development was created by literature mining. The sequences of these genes, together with their interactions and first neighbours, were retrieved from the *Arabidopsis* gene transcriptional regulatory map<sup>13</sup> (downloaded from <http://atrm.cbi.pku.edu.cn/>). The putative *P. pringlei* orthologues of these genes were identified by determining the best bidirectional hit (BBH) for each sequence when possible, or unidirectional hits when the BBH was unsuccessful. The contigs were considered as putative orthologues when the BLAST coverage was  $>40\%$  and the E-value was  $<1 \times 10^{-10}$ , and when domains from the query sequence, verified by SuperFamily<sup>23</sup>, were present. The resulting transcriptional regulatory network (TRN) was visualized in Cytoscape<sup>24</sup> (v3.2.1).

## References

- Sebastian, I. & Lee, I. Root Apical Meristems. In: *eLS*, John Wiley & Sons, Ltd. Chichester, <https://doi.org/10.1002/9780470015902.a020121.pub2> (2013).
- Sablowski, R. Flowering and determinacy in *Arabidopsis*. *J. Exp. Bot.* **58**, 899–907 (2007).
- Dinkelaker, B., Hengeler, C. & Marschner, H. Distribution and function of proteoid roots and other root clusters. *Bot. Acta.* **108**, 183–200 (1995).
- Lamont, R. B. Structure, ecology and physiology of root clusters – a review. *Plant and Soil* **48**, 1–19 (2003).
- Groot, E. P., Sweeney, E. J. & Rost, T. L. Development of the adhesive pad on climbing fir (*Ficus pumila*) stems from clusters of adventitious roots. *Plant and Soil* **248**, 85–96 (2003).
- Varney, G. T. & McCully, M. E. The branch roots of *Zea*. II. Developmental loss of the apical meristem in field-grown roots. *New Phytol.* **118**, 535–546 (1991).
- Bok, N. H. Root glochids and root spurs of *Opuntia arenaria* (Cactaceae). *Am. J. Bot.* **69**, 1085–1092 (1979).
- Dubrovsky, I. G. Determinate primary-root growth in seedlings of Sonoran Desert Cactaceae; its organization, cellular basis, and ecological significance. *Planta* **203**, 85–92 (1997).
- Dubrovsky, I. G. & Gómez-Lomeli, L. F. 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). *Am. J. Bot.* **90**, 823–831 (2003).
- Shishkova, S. *et al.* Determinate primary root growth as an adaptation to aridity in Cactaceae: towards an understanding of the evolution and genetic control of the trait. *Ann. Bot.* **112**, 239–252 (2013).
- Almeida, O. J. G., Paoli, A. A. S., Souza, I. A. & Cota-Sánchez, J. H. Seedling morphology and development in the epiphytic cactus *Epiphyllum phyllanthus* (L.) Haw. (Cactaceae: Hylocereeae). *J. Torrey Bot. Soc.* **140**, 196–217 (2013).



12. Shishkova, S. *et al.* Regeneration of roots from callus reveals stability of the developmental program for determinate root growth in Sonoran Desert Cactaceae. *Plant Cell Rep.* **26**, 547–557 (2007).
13. Aida, M. *et al.* The PLETHORA genes mediate patterning of the *Arabidopsis* root stem cell niche. *Cell* **119**, 109–120 (2004).
14. Santuari, L. *et al.* The PLETHORA gene regulatory network guides growth and cell differentiation in *Arabidopsis* root. *Plant Cell* **28**, 2937–2951 (2016).
15. Galinha, C. *et al.* PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* **449**, 1053–1057 (2007).
16. Helariutta, Y. *et al.* The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* **101**, 555–67 (2000).
17. Nakajima, K., Sena, G., Navy, T. & Benfey, P. N. Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* **413**, 307–311 (2001).
18. Benfey, P. N. *et al.* Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* **119**, 57–70 (1993).
19. Scheres, B. *et al.* Mutations affecting the radial organisation of the *Arabidopsis* root display specific defects throughout the embryonic axis. *Development* **121**, 53–62 (1995).
20. Sabatini, S., Heidstra, R., Wildwater, M. & Scheres, B. SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Gen. Dev.* **17**, 354–358 (2003).
21. Sarkar, A. *et al.* Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* **446**, 811–814 (2007).
22. Nardmann, J., Reiselwitz, P. & Werr, W. Discrete shoot and root stem cell-promoting WUS/WOX5 functions are an evolutionary innovation of angiosperms. *Mol. Biol. Evol.* **26**, 1745–1755 (2009).
23. Stahl, Y., Wink, R. H., Ingram, G. C. & Simon, R. A signaling module controlling the stem cell niche in *Arabidopsis* root meristems. *Curr. Biol.* **19**, 909–914 (2009).
24. Copetti, D. *et al.* Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proc. Natl. Acad. Sci. USA* **114**, 12003–12008 (2017).
25. Wang, Z., Gerstein, M. & Snyder, M. RNA-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
26. Ibarra-Laclette, E. *et al.* De novo sequencing and analysis of *Lophophora williamsii* transcriptome and searching for putative genes involved in mescaline biosynthesis. *BMC Genomics* **16**, 657, <https://doi.org/10.1186/s12864-015-1821-9> (2015).
27. Murawski, D. A., Fleming, T. H., Rutland, K. & Hamrick, J. L. Mating system of *Pachycereus pringlei*: an autotetraploid cactus. *Heredity* **72**, 86–94 (1994).
28. Moriya, Y., Itoh, M., Okada, S., Yoshizawa, A. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185, <https://doi.org/10.1093/nar/gkm321> (2007).
29. Dohm, J. C. *et al.* The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* **505**, 546–549 (2014).
30. Clouse, J. W. *et al.* The Amaranth genome: genome, transcriptome, and physical map assembly. *Plant Genome* **9**, <https://doi.org/10.3835/plantgenome2015.07.0062> (2016).
31. Jarvis, D. E. *et al.* The genome of *Chenopodium quinoa*. *Nature* **542**, 307–312 (2017).
32. Mallona, I., Egea-Cortines, M. & Weiss, J. Conserved and divergent rhythms of crassulacean acid metabolism-related and core clock gene expression in the cactus *Opuntia ficus-indica*. *Plant Physiol.* **156**, 1978–89 (2011).
33. Qingzhu, H. *et al.* Transcriptomic analysis reveals key genes related to betalain biosynthesis in pulp coloration of *Hylocereus polyrhizus*. *Front. Plant Sci.* **6**, 1179, <https://doi.org/10.3389/fpls.2015.01179> (2016).
34. Yang, Y. *et al.* Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol. Biol. Evol.* **32**, 2001–2014 (2015).
35. Huang, L. & Schiefelbein, J. Conserved gene expression programs in developing roots from diverse plants. *Plant Cell* **27**, 2119–2132 (2015).
36. Shishkova, S. & Dubrovsky, J. G. Developmental programmed cell death in primary roots of Sonoran Desert Cactaceae. *Am. J. Bot.* **92**, 1590–1594 (2005).
37. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 149–140 (2010).
38. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
39. Lloyd, S. Least squares quantization in PCM. *Information Theory, IEEE Transactions* **28**, 129–137 (1982).
40. Czechowsky, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139**, 5–17 (2005).
41. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006).
42. Arpetita, F., Weinstein, N., Benitez, M., Mendoza, L. & Alvarez-Buylla, E. R. Finding missing interactions of the *Arabidopsis thaliana* root stem cell niche gene regulatory network. *Front. Plant Sci.* **4**, 110, <https://doi.org/10.3389/fpls.2013.00110> (2013).
43. Jin, J. *et al.* An *Arabidopsis* transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. *Mol. Biol. Evol.* **32**, 1767–1773 (2015).
44. Hirano, Y. *et al.* Structure of the SHR-SCR heterodimer bound to the BIRD/IDD transcriptional factor JKD. *Nat. Plants* **3**, <https://doi.org/10.1038/nplants.2017.10> (2017).
45. Long, Y. *et al.* In vivo FRET-FLIM reveals cell-type-specific protein interactions in *Arabidopsis* roots. *Nature* **548**, 97–102 (2017).
46. Carlsbecker, A. *et al.* Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* **465**, 316–321 (2010).
47. Moubayidin, L. *et al.* Spatial coordination between stem cell activity and cell differentiation in the root meristem. *Dev. Cell.* **26**, 405–415 (2013).
48. Welch, D., Hassan, H., Bilou, I., Immink, R., Heidstra, R. & Scheres, B. *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes Dev.* **21**, 2196–2204 (2007).
49. Cruz-Ramirez, A. *et al.* A bistable circuit involving SCARECROW-RETINOBLASTOMA integrates cues to inform asymmetric stem cell division. *Cell* **150**, 1002–1015 (2012).
50. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
51. Hellemans, J., Mortier, G., De Paepe, A., Spelmann, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19, <https://doi.org/10.1186/gb-2007-8-2-r19> (2007).
52. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**(0034), 1–0034.11 (2002).
53. Wilson, D. *et al.* SUPERFAMILY — sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* **37**, <https://doi.org/10.1093/nar/gkn762> (2009).
54. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

### Acknowledgements

We thank José F. Martínez-Rodríguez and Francisco Molina-Freaner for seed donation; Unidad de Síntesis de ADN, IBT-UNAM, for qPCR-primer synthesis; Marcela Ramírez-Yarza, several bachelor students, Unidad de Biblioteca and Unidad de Cómputo, IBT-UNAM, for excellent technical help. This work was supported by the PAPIIT-UNAM (IN207115, IN201318, IN200818) and CONACyT (CB240055, FOINS219522, CB237430, PhD Fellowships 290654, 288069), Mexico.

### Author Contributions

G.R.-A., M.M., J.G.D. and S.S. designed the research; G.R.-A., M.M., S.N.-M., M.L.L.-V. and P.L.-M. performed the experiments; G.R.-A., M.M. and S.S. analysed the data, J.G.D., M.M. and S.S. provided resources, G.R.-A. and S.S. wrote the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-26897-1>.

**Competing Interests:** The authors declare no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

b. Jorge Reyes-Rivera, **Gustavo Rodríguez-Alonso**, Emilio Petrone, Alejandra Vasco, Francisco Vergara-Silva, Svetlana Shishkova & Teresa Terrazas (2017). Expression of the *KNOTTED HOMEBOX* genes in the Cactaceae cambial zone suggests their involvement in wood development. *Front. Plant Sci.* **8**:218. DOI: 10.3389/fpls.2017.00218



## Expression of the *KNOTTED HOMEBOX* Genes in the Cactaceae Cambial Zone Suggests Their Involvement in Wood Development

Jorge Reyes-Rivera<sup>1†</sup>, Gustavo Rodríguez-Alonso<sup>2†</sup>, Emilio Petrone<sup>1</sup>, Alejandra Vasco<sup>1</sup>, Francisco Vergara-Silva<sup>3</sup>, Svetlana Shishkova<sup>2\*</sup> and Teresa Terrazas<sup>1\*</sup>

<sup>1</sup> Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico.

<sup>2</sup> Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico, <sup>3</sup> Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico

### OPEN ACCESS

#### Edited by:

Federica Valverde,  
Spanish National Research Council,  
Spain

#### Reviewed by:

David Smyth,  
Monash University, Australia  
Ykä Helenatla,  
University of Helsinki, Finland

#### \*Correspondence:

Teresa Terrazas  
tterrzas@ib.unam.mx  
Svetlana Shishkova  
svetlana@ib.unam.mx

<sup>†</sup>These authors have contributed  
equally to this work

#### Specialty section:

This article was submitted to  
Plant Evolution and Development,  
a section of the journal  
*Frontiers in Plant Science*

Received: 23 September 2016

Accepted: 06 February 2017

Published: 03 March 2017

#### Citation:

Reyes-Rivera J, Rodríguez-Alonso G,  
Petrone E, Vasco A, Vergara-Silva F,  
Shishkova S and Terrazas T (2017)  
Expression of the *KNOTTED*  
*HOMEBOX* Genes in the Cactaceae  
Cambial Zone Suggests Their  
Involvement in Wood Development.  
*Front. Plant Sci.* **8**:218.  
doi: 10.3389/fpls.2017.00218

The vascular cambium is a lateral meristem that produces secondary xylem (i.e., wood) and phloem. Different Cactaceae species develop different types of secondary xylem; however, little is known about the mechanisms underlying wood formation in the Cactaceae. The *KNOTTED HOMEBOX* (*KNOX*) gene family encodes transcription factors that regulate plant development. The role of class I *KNOX* genes in the regulation of the shoot apical meristem, inflorescence architecture, and secondary growth is established in a few model species, while the functions of class II *KNOX* genes are less well understood, although the *Arabidopsis thaliana* class II *KNOX* protein *KNAT7* is known to regulate secondary cell wall biosynthesis. To explore the involvement of the *KNOX* genes in the enormous variability of wood in Cactaceae, we identified orthologous genes expressed in species with fibrous (*Pereskia lychnidiflora* and *Pilosocereus alensis*), non-fibrous (*Ariocarpus retusus*), and dimorphic (*Ferocactus pilosus*) wood. Both class I and class II *KNOX* genes were expressed in the cactus cambial zone, including one or two class I paralogs of *KNAT1*, as well as one or two class II paralogs of *KNAT3-KNAT4-KNAT5*. While the *KNOX* gene *SHOOTMERISTEMLESS* (*STM*) and its ortholog *ARK1* are expressed during secondary growth in the *Arabidopsis* and *Populus* stem, respectively, we did not find *STM* orthologs in the Cactaceae cambial zone, which suggests possible differences in the vascular cambium genetic regulatory network in these species. Importantly, while two class II *KNOX* paralogs from the *KNAT7* clade were expressed in the cambial zone of *A. retusus* and *F. pilosus*, we did not detect *KNAT7* ortholog expression in the cambial zone of *P. lychnidiflora*. Differences in the transcriptional repressor activity of secondary cell wall biosynthesis by the *KNAT7* orthologs could therefore explain the differences in wood development in the cactus species.

**Keywords:** Cactaceae, dimorphic wood, fibrous wood, *KNAT7* ortholog, *KNOX* transcription factors, non-fibrous wood, vascular cambium, wood lignification



## INTRODUCTION

As reservoirs of pluripotent cells, meristems have played a leading role in the diversification of angiosperm growth forms. The shoot and root apical meristems maintain primary growth in plants, while the lateral meristems, comprised of the vascular cambium and cork cambium in Eudicotyledons and gymnosperms, are involved in secondary growth (Evert, 2006). The vascular cambium maintains a population of initial (stem) cells, which divide asymmetrically to generate two daughter cells; one maintains the cambial initial identity, while the other divides again and the daughters differentiate to generate secondary phloem or xylem (Groover et al., 2010; Miyashima et al., 2013). Vascular cambium derivatives are thought to have influenced speciation and diversification events (Spicer and Groover, 2010; Lucas et al., 2013). In the Cactaceae, the traits of the secondary xylem (i.e., wood) suggest that it has evolved by heterochronic processes where a change in the timing of developmental processes leads to morphological differences between species (Altesor et al., 1994). The larger species ( $\geq 1.5$  m in height) in this family of succulent plants have fibrous wood with vessel elements in the xylem similar to those typically derived from vascular cambium, with a similar wood chemical composition to that of other woody dicot species (Reyes-Rivera et al., 2015). On the other hand, the wood found in smaller species is generally scarce and non-fibrous, with abundant wide-band tracheids and vessel elements similar to those typical of proto- and metaxylem. In these species, the level of wood lignification is insignificant and the lignin has a heterogeneous chemical composition (Reyes-Rivera et al., 2015). Many species of Cactaceae have dimorphic wood, where one type of wood is produced in the juvenile stages and a different structure is formed in the adult stages of development. To the best of our knowledge, this phenomenon is unique among Eudicotyledons and is related to the globose and globose-depressed growth forms of some Cactaceae species (Mauseth and Plemons, 1995). In species with dimorphic wood that produce wide-band tracheids, these always develop first, before the fibrous or parenchymatous wood is produced (Mauseth and Plemons, 1995; Loza-Cornejo and Terrazas, 2011). The mechanisms shaping the development of the vascular cambium and its derivatives in the Cactaceae are mostly unknown; nevertheless, it was suggested that the wide variation in wood anatomy of different cacti species might be attributed to a variation of gene expression patterns and gene expression level (Mauseth and Plemons, 1995; Mauseth, 2006; Landrum, 2008; Vázquez-Sánchez and Terrazas, 2011).

The interplay of diverse factors in the regulation of vascular cambium activity has been reported previously (Liu et al., 2014; Nieminen et al., 2015), and the roles of growth regulators, including auxin, cytokinin, and ethylene, in this process are well established. Later in development, genes regulating cell expansion, secondary xylem differentiation, lignification, and secondary wall deposition contribute to wood formation (reviewed in Růžicka et al., 2015; Ye and Zhong, 2015; Zhong and Ye, 2015). Transcription profiling of the vascular cambium in aspen (*Populus tremula*) uncovered similarities between the gene regulatory networks operating in the shoot apical meristem

and the vascular cambium. In particular, it was reported that four members of the *KNOTTED1-LIKE HOMEBOX (KNOX)* gene family, namely *PttKNOX1*, *PttSHOOT MERISTEMLESS (PttSTM)*, *PttKNOX2*, and *PttKNOX6*, are highly expressed in both tissues (Schrader et al., 2004). Functional analysis suggests that the *Populus* orthologs of *Arabidopsis thaliana (Arabidopsis)* class I *KNOX* genes *STM* and *KNOTTED-LIKE HOMEBOX OF ARABIDOPSIS THALIANA 1 (KNAT1)/BREVIPEDICELLUS (BP)*, *PttSTM/ARBORKNOX 1 (ARK1)/ARK1a* and *PttKNOX1/ARK2*, respectively, regulate secondary growth (Groover et al., 2006; Du et al., 2009; Liu et al., 2015). While the roles of class I *KNOX* genes in the regulation of shoot apical meristem, inflorescence architecture, and compound leaf development are well established, the functions of class II *KNOX* genes are less well understood. In general, class I transcripts are less abundant than class II *KNOX* transcripts, and are expressed in specific regions of meristems, particularly in the shoot apical meristem and leaf meristems. By contrast, class II transcripts are found in differentiating cells and all mature plant organs (Serikawa et al., 1997; reviewed by Hay and Tsiantis, 2010; Arnaud and Pautot, 2014), but not in the shoot apical meristem (Furumizu et al., 2015). Moreover, the dark-green serrated leaf phenotype of the class II *knat3 knat4 knat5* triple loss-of-function mutant in *Arabidopsis* is similar to that of the class I gain-of-function mutants (Furumizu et al., 2015), suggesting opposing functions for genes of the two classes. Here, we report that class I and class II *KNOX* genes are expressed in the cambial zone, consisting of vascular cambium and derived cells, of Cactaceae species with fibrous (*Pereskia lychnidiflora* and *Pilosocereus alensis*), non-fibrous (*Ariocarpus retusus*), and dimorphic (*Ferocactus pilosus*) wood. We also present the phylogeny of class I and class II *KNOX* proteins encoded in the sequenced plant genomes retrieved from the Phytozome database, confirming monophyly of class I and class II *KNOX* proteins, assigning the Cactaceae *KNOX* proteins into clades, and exploring the number of paralogs of the plant species in every *KNOX* clade.

## MATERIALS AND METHODS

### Plant Material and Tissue Sampling

Four Cactaceae species with different wood types were used in this study: *Pereskia lychnidiflora* DC. (subfamily: Pereskioideae); *Pilosocereus alensis* (F. A. C. Weber) Byles & G. D. Rowley (subfamily: Cactoideae, tribe: Cereeae); *Ariocarpus retusus* Scheidw (Cactoideae, tribe: Cacteeae); and *Ferocactus pilosus* (Galeotti ex Salm-Dyck) Werderm. (Cactoideae, Cacteeae). Samples were collected from the cambial zone of adult individuals growing in Mexico. For the tall species (*F. pilosus*, *P. lychnidiflora*), the cambial zone was harvested from one individual of both species in the field (*F. pilosus* at the Mexican Plateau, San Luis Potosí state, location 23°19'31"N; 100°33'30"W, voucher TT890 MEXU; and *P. lychnidiflora* at the Isthmus of Tehuantepec, Oaxaca state, location 16°22'52"N; 95°19'03"W, voucher TT966 MEXU), with samples collected during the rainy season to ensure that the vascular cambium



was active. All surrounding tissues including the cortex were removed as described by Liu et al. (2014). The cambial zone was peeled off with a disposable microtome knife and immediately frozen in liquid nitrogen. One whole individual of both *A. retusus* and *P. alensis* were placed in pots containing their native soil and transported from their natural habitat to the laboratory (*A. retusus* from the Mexican Plateau, Nuevo León state, location 23°23'39"N; 100°21'57"W, voucher SA1976 MEXU; and *P. alensis* from the Pacific coast, Michoacán state, location 18°46'41"N; 103°08'05"W, voucher JRR3 MEXU). Species identification was confirmed by S. Arias [Instituto de Biología, Universidad Nacional Autónoma de México (IB, UNAM)], the leading taxonomy specialist of the Cactaceae family in Mexico. The cambial zone was harvested as described above. The cambial zone samples were stored at -80°C until RNA extraction. For expression analysis by RT-PCR, one adult plant of *A. retusus*, *F. pilosus*, and *P. lychnidiflora*, one 2-year-old juvenile plant of the first two species, and two 5-year-old young individuals of *A. retusus*, previously collected from the same locations, were donated by the Botanical Garden IB, UNAM. For the root tip transcriptome of *Pachycereus pringlei* (S. Watson) Britton & Rose, plants were germinated from seeds collected near Bahía Kino, Sonora state, and donated by F. Molina-Freaner and J. F. Martínez-Rodríguez (Instituto de Ecología, campus Hermosillo, UNAM).

For the wood maceration, fine wood chips (2 mm thickness) were obtained every 5 mm, from the region closest to the pith (young wood) to the section closest to the vascular cambium (mature wood), and each region was processed separately. The non-fibrous *A. retusus* samples were placed in 2-mL microcentrifuge tubes filled with Franklin solution (5:1:4 acetic acid:hydrogen peroxide:water; Ruzin, 1999), while the dimorphic *F. pilosus* samples and the fibrous *P. alensis* and *P. lychnidiflora* were placed into tubes containing Jeffrey solution (equal volumes of 10% chromic acid and 10% nitric acid; Berlyn and Miksche, 1976). For each species, 0.2 g of the tissue was used. The samples were then incubated at 56°C for 4 h (non-fibrous wood) and 24 h (dimorphic and fibrous wood). Additionally, all samples were sonicated in a Branson 200 ultrasonic cleaner (Branson Ultrasonics, Danbury, CT, USA) until completely macerated, washed with water, and dehydrated using a series of ethanol washes at 50, 70, 90%, and absolute concentrations. The *A. retusus* macerations were stained with a 0.1% aqueous solution of toluidine blue for 12 h (Ruzin, 1999) and mounted onto slides. The macerations of the wood samples from another three species were stained with Safranin for 2 h and mounted onto slides using synthetic resin (Entellan, Merck Millipore, Darmstadt, Germany). All wood elements were photographed using a BX51 optical microscope (Olympus Corporation, Tokyo, Japan) and the images were analyzed using Image-Pro Plus v. 6.1 software (Media Cybernetics, Inc., Bethesda, MD, USA).

### RNA Extraction, cDNA Synthesis, and PCR Amplification

Total RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's protocol, including the optional step of centrifugation before the

separation of phases, or using the Spectrum Plant Total RNA Isolation Kit (Sigma-Aldrich, St. Louis, MO, USA). The cDNA was synthesized using SuperScript II Reverse Transcriptase (Invitrogen), according to the manufacturer's instructions. Degenerate PCR primers were used to amplify the *KNOX* genes. Primers for amplifying putative cacti orthologs of *KNAT3* were designed (KNAT3\_Cact-F: 5'-GAGAGRAATAATGGCWTATCATC-3' and KNAT3\_Cact-R: 5'-CCTTCTGGTTCTACTTCCCTC-3') based on the alignments of nucleotide sequences encoding the class II KNOX proteins *KNAT3* (*Arabidopsis*) and its closest homologs from *Beta vulgaris* and *Pachycereus pringlei*. For the Cactaceae orthologs of *KNAT1*, primers designed by Du et al. (2009) were used. PCR products were purified with Sephadex Centri-sep columns (Thermo Fisher Scientific, Waltham, MA, USA) as instructed by the manufacturer. The amplified and purified products were sequenced in a 3500xL Genetic Analyzer sequencer (Applied Biosystems, Foster City, CA, USA) using the PCR primers. Platinum Taq polymerase (Thermo Fischer Scientific, Waltham, MA, USA) was used for PCR reactions. Primers used for RT-PCR are listed in **Supplementary Table S1**. RNA-seq was performed at the Beijing Genome Institute, Hong Kong; the vascular cambium and root tip transcriptomes were *de novo* assembled using Trinity v. 2.2<sup>1</sup> and CLC Genomic Workbench v. 7.5 (Qiagen<sup>2</sup>), respectively.

### Sequence Alignment and Phylogenetic Analysis

KNOX-like protein sequences were retrieved from the Phytozome database<sup>3</sup> (v. 11; last accessed on May 9, 2016) using the PhytoMine tools. All proteins with KNOX (IPR005540 [InterProscan definition]/PF03790 [Pfam definition] and IPR005541/PF03791), ELK (IPR005539/PF03789), and HD domains (IPR009057/PF05920) were retrieved. *B. vulgaris* sequences were retrieved using tBLASTn from the RefBee v. 1.2 genome assembly (The *Beta vulgaris* resource<sup>4</sup>, Dohm et al., 2013) using a BLOSUM80 substitution matrix (Henikoff and Henikoff, 1992; at the time, *B. vulgaris* was the only species from the order Caryophyllales with a sequenced genome). Chimeric sequences and those from the early release genomes were discarded. After that, KNOX protein sequences from the tree species *Betula luminifera* and *Juglans nigra* were added. For Cactaceae species, the KNOX protein sequences used were deduced from the amplified and sequenced PCR fragments (see previous section), from our RNA-seq data and the *de novo* assembly of the cambial zone transcriptome of the four species reported in this work (the same RNA samples were used as starting material for the cambial zone transcriptome assembly), from the root tip transcriptome of the Cactoideae subfamily species *Pachycereus pringlei* (the analysis of the transcriptomes will be reported elsewhere), and from the recently published *Lophophora williamsii* transcriptome (Ibarra-Laclette et al., 2015). The resulting sequences were

<sup>1</sup><https://github.com/trinityrnaseq/trinityrnaseq/releases>

<sup>2</sup><https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/>

<sup>3</sup><http://phytozome.jgi.doe.gov>

<sup>4</sup><http://bvseq.molgen.mpg.de/index.shtml>

aligned with Clustal Omega<sup>3</sup> and the alignment file was manually edited. Alignment positions with more than 30% gaps were not included in the analysis. The identity and similarity values (%) were obtained from a pairwise alignment (Needle-EMBOSS<sup>6</sup>), with a complete gap deletion for each pair.

The phylogeny was reconstructed with MEGA7 (Kumar et al., 2016). A maximum likelihood algorithm based on the JTT substitution model (Felsenstein, 1981) was used to resolve the phylogenetic relationship of the KNOX proteins derived from the Cactaceae species and the plant species with sequenced genomes. The resulting topology was statistically tested with a 1000 replicate bootstrap for both the complete and the selected datasets. The NWK files were visualized and edited on FigTree<sup>7</sup> (v. 1.4.2). The BELL 1 protein of *Arabidopsis* (AT5G41410.1) was used as the outgroup. Nucleotide sequences reported in this work were deposited in GenBank under the accession numbers KX891335–KX891338 for *F. pilosus*; KX891339–KX891343 and KX891349 for *A. retusus*; KX891344–KX891346 for *P. lychnidiflora*; KX891347 and KX891348 for *P. alensis*; and KX870027–KX870031 for *P. pringlei*.

## RESULTS

### Morphological Analysis of Xylem Cells

To examine the detailed cellular features of the secondary xylem in the four species studied, we performed a morphological analysis of xylem cells using wood macerations. The young xylem of *A. retusus* with non-fibrous wood and *F. pilosus* with dimorphic wood forms wide-band tracheids and vessel elements with annular and helical secondary wall patterns formed during early development. The mature fibers and vessel elements with pseudoscalariform and alternate pitting were formed in adult plants during late development. No differences were observed between young and mature wood of *P. alensis* and *P. lychnidiflora*, with both having fibrous wood. The tracheary elements of *A. retusus* and young *F. pilosus* had a low proportion of lignified cell wall, similar to those of proto- and metaxylem (i.e., primary xylem), whereas vessel elements and fibers of mature *F. pilosus*, *P. alensis*, and *P. lychnidiflora*, had a higher proportion of lignified cell walls. This analysis allowed the unequivocal identification of each secondary xylem cell type and its cell wall features (Figure 1) and thus confirmed the wood types we previously identified using an anatomical analysis of wood sections (Reyes-Rivera et al., 2015). The variability in size and pitting of vessel elements associated with the presence of fibers was found in species with dimorphic and fibrous wood.

### KNOX Family Genes Are Expressed in the Cactaceae Cambial Zone

To explore the involvement of the KNOX family in the enormous variability of wood morphologies in the Cactaceae, we looked for orthologs of these genes expressed in the Cactaceae cambial

zone, which comprises the vascular cambium and recently derived cells. RT-PCR using degenerate primers for the class I KNOX genes resulted in a major amplification product of the predicted molecular weight in the four Cactaceae species studied. The amino acid sequences inferred from the amplified and sequenced fragments of these genes (*ArKNOX1a* from *Ariocarpus retusus*, *FpKNOX1a* from *Ferocactus pilosus*, *PaKNOX1a* from *Pilosocereus alensis*, and *PIKNOX1a* from *Pereskia lychnidiflora*) correspond to nearly 70% of the ARK2 and KNAT1 protein sequences of *Populus* and *Arabidopsis*, respectively, including part of the KNOX1 domain and the entirety of the KNOX2, ELK, and HD domains [Figure 2 (The last letters in the gene names refer to: “a,” amplification as a method of gene isolation; “e,” the inferred coding sequence extended after alignment with sequences resulting from the RNA-seq; and “r,” RNA-seq as a method of gene identification)].

Fragments of class II KNOX genes were successfully amplified by RT-PCR with degenerate primers from the three species from the Cactoideae subfamily, but were not isolated from *P. lychnidiflora*. The amino acid sequences inferred from the amplified and sequenced fragments of these three genes, *ArKNOX3a*, *FpKNOX3a*, and *PaKNOX3a*, cover nearly 95% of the KNAT3 protein sequence, including the KNOX1, KNOX2, and ELK domains, as well as most of the HD domain (Figure 3).

In addition to the seven transcripts mentioned above, eight more KNOX transcripts were *de novo* assembled from our RNA-seq data of the cambial zone samples of three Cactaceae species, *A. retusus*, *F. pilosus*, and *P. lychnidiflora*. The amino acid sequences deduced from the amplified transcripts of *F. pilosus* were later extended by alignment with the assembled transcripts, and were therefore renamed *FpKNOX1ae* and *FpKNOX3ae* (Figures 2–5). Expression of all KNOX transcripts in the cambial zone of *A. retusus*, *F. pilosus*, and *P. lychnidiflora* was confirmed by RT-PCR (Supplementary Figure S1). Remarkably, in the cambial zone of 2-year-old juvenile plants of both *A. retusus* and *F. pilosus*, one of the two orthologs of KNOX3, namely, *ArKNOX3a* and *FpKNOX3ae*, and of KNOX7, namely, *ArKNOX7r1* and *FpKNOX7r2*, were expressed. No expression of KNOX1 orthologs was detected in young plants. At this stage, only very scarce accumulation of wood, represented by wide-band tracheids and vessel elements, was detected in these species. From the six KNOX genes expressed in the vascular cambium of adult *A. retusus* plant, expression of only *ArKNOX7r1* was detected in the tubercle of 5-year-old plant (data not shown).

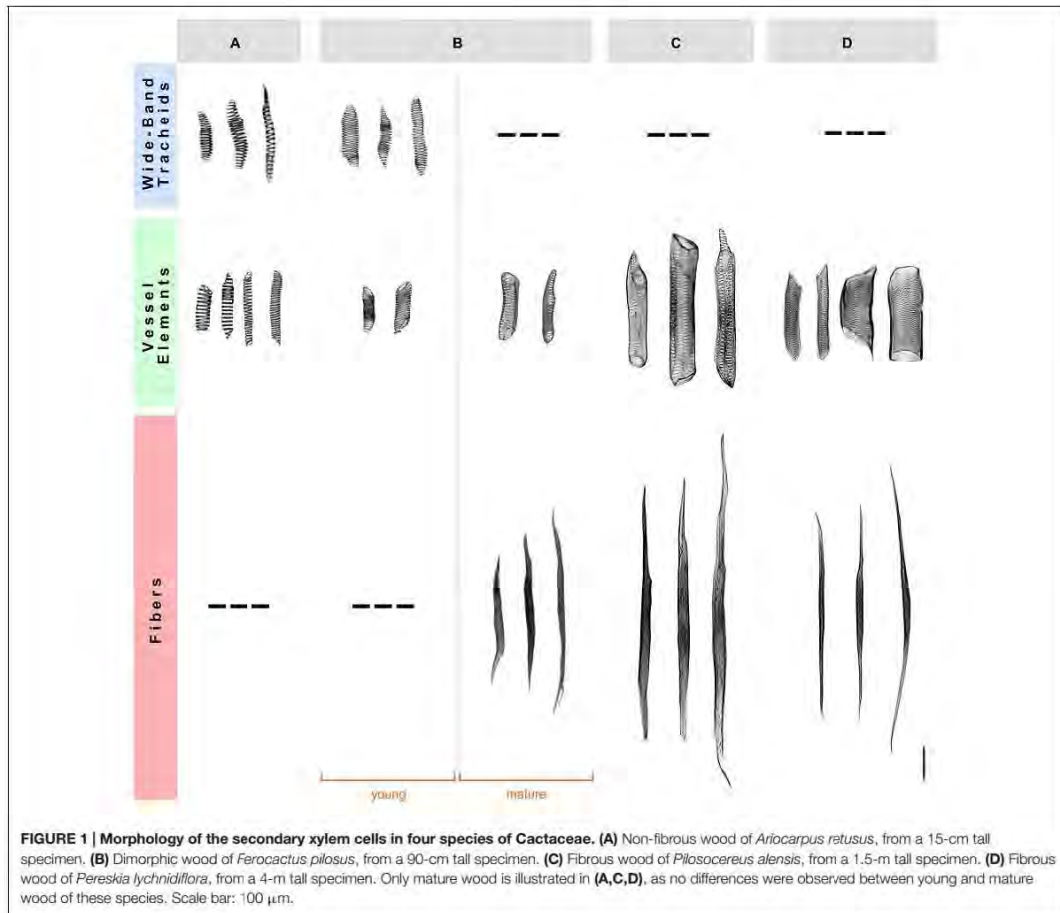
### Phylogenetic Analysis of the KNOX Family

Initially, 524 KNOX-like protein sequences were retrieved from the Phytozome database. Chimeric sequences and those from draft genome releases were filtered out, and the inferred protein sequences of Cactaceae, the trees *B. luminifera* and *J. nigra*, as well as those retrieved from the *B. vulgaris* genome, were included. The resulting matrix contained 478 aligned protein sequences belonging to 45 species with sequenced genomes, six Cactaceae species, and two species of tree. The identifiers of the sequences retrieved from the Phytozome database are listed in Supplementary Table S2. The maximum likelihood phylogeny

<sup>3</sup><http://www.ebi.ac.uk/Tools/msa/clustalo/>

<sup>6</sup>[http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](http://www.ebi.ac.uk/Tools/psa/emboss_needle/)

<sup>7</sup><http://tree.bio.ed.ac.uk/>



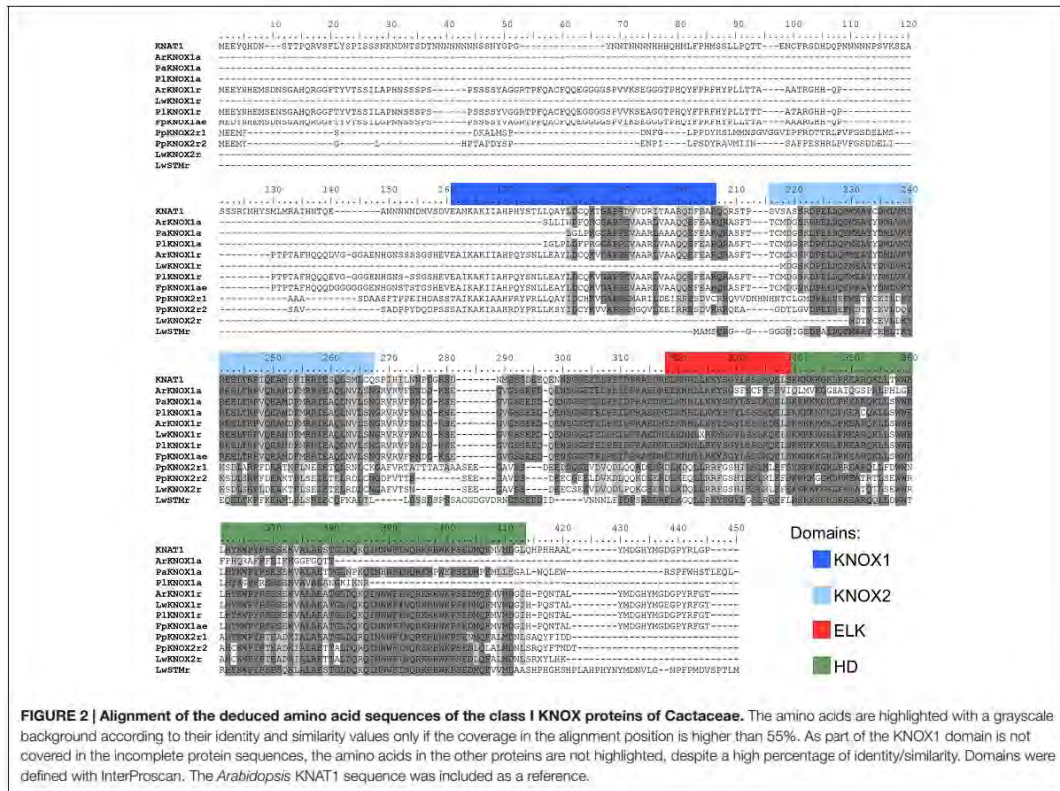
grouped the KNOX proteins into five main clades (Figures 4, 5), which were named according to their putative *Arabidopsis* ortholog: i.e., the KNAT1 clade (gray), KNAT2-KNAT6 clade (cyan), STM clade (pink), KNAT3-KNAT4-KNAT5 clade (blue), and KNAT7 clade (green). The first three clades belong to the class I KNOX and the last two belong to the class II KNOX proteins. Figure 5 depicts the phylogenetic relationships of the subset of KNOX proteins from *Arabidopsis*, the Cactaceae, and *B. vulgaris*. *Arabidopsis* was used as the reference species, with *B. vulgaris* from order Caryophyllales, to which the Cactaceae family belongs, included as the closest sister taxon with a sequenced genome.

The molecular phylogenetic analysis confirmed that the four Cactaceae class I KNOX genes amplified from the cambial zone by RT-PCR with degenerate primers (*ArKNOX1a*, *FpKNOX1ae*, *PaKNOX1a*, and *PIKNOX1a*) are part of the KNAT1 clade (Figures 4, 5); moreover, two of the *de novo* assembled

transcripts, *ArKNOX1r* and *PIKNOX1r*, also belonged to the KNAT1 clade. The class II *ArKNOX3a*, *FpKNOX3ae*, and *PaKNOX3a* genes expressed in the vascular zone of the three species fall into the KNAT3-KNAT4-KNAT5 clade, along with two other class II KNOX genes, *ArKNOX3r* and *PIKNOX3r* (the only *Pereskia* class II gene identified in this study). Four *de novo* assembled transcripts, *ArKNOX7r1*, *ArKNOX7r2*, *FpKNOX7r1*, and *FpKNOX7r2*, belong to the KNAT7 clade (Figures 4, 5 and Table 1). The identity and similarity percentages for the proteins encoded by the identified Cactaceae genes are shown in Supplementary Figure S2.

Of the five KNOX transcripts detected in the *P. pringlei* root tip, both of the class I KNOX genes were found in the KNAT2-KNAT6 clade, one class II sequence was attributed to the KNAT3-KNAT4-KNAT5 clade, and two more belonged to the KNAT7 clade. Among the five KNOX genes identified in the recently published *L. williamsii* transcriptome



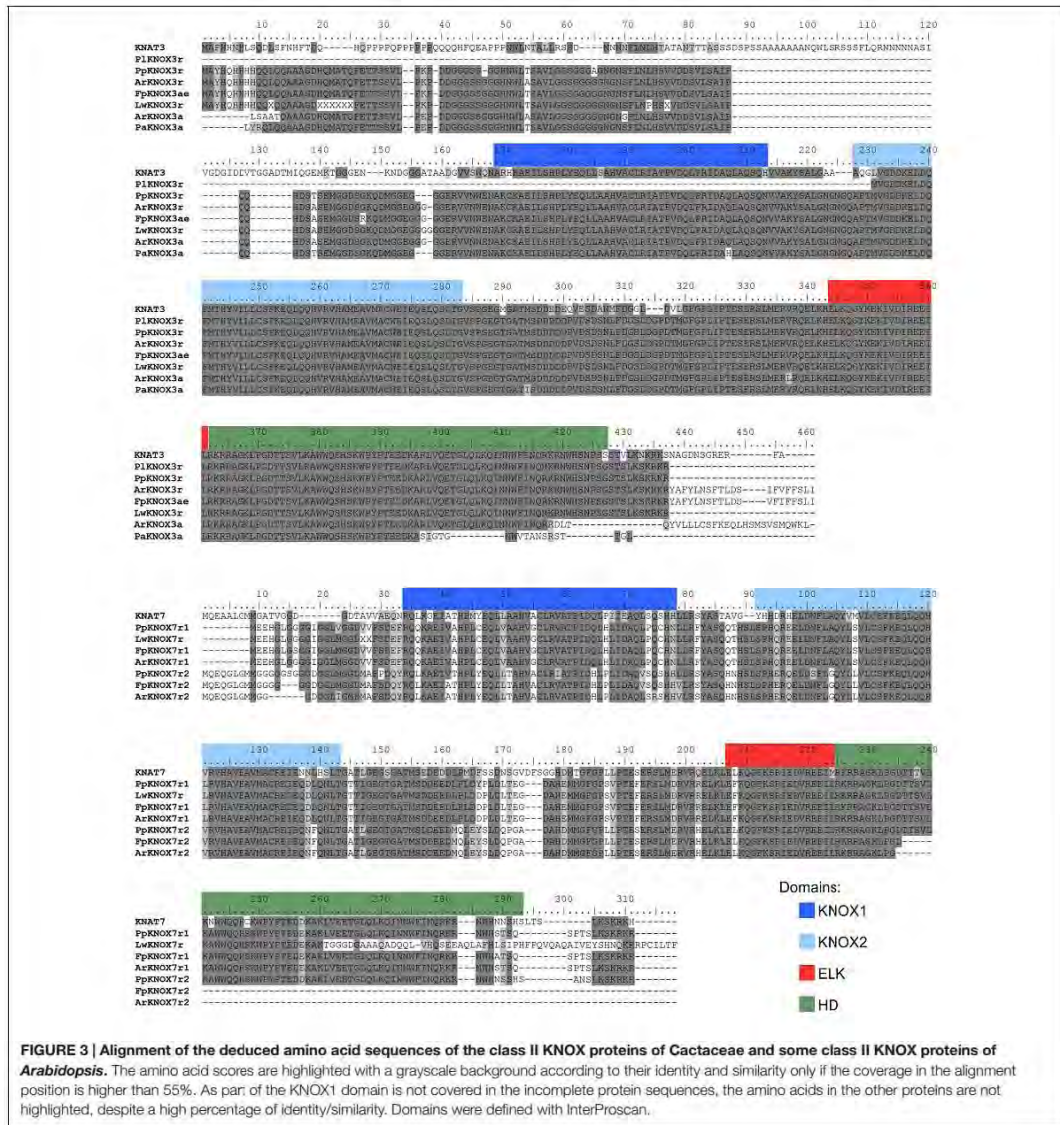


(Ibarra-Laclette et al., 2015), one was found in each of the five clades (Figures 4, 5 and Table 1). In the global KNOX phylogeny (Figure 4), *B. vulgaris* sequences were always resolved as sister to the Cactaceae sequences. In the phylogeny for the subset of KNOX proteins (Figure 5), *B. vulgaris* class I KNOX proteins were resolved as sister to the Cactaceae sequences in all three clades, while this was not the case for two clades of class II KNOX. The only *B. vulgaris* paralog of KNAT3, KNAT4, and KNAT5 fell in the subclade of the *A. thaliana* sequences. Within the KNAT7 clade, the *B. vulgaris* sequence represented the sister sequence for the Cactaceae KNOX7r2 subclade, whereas the KNOX7r1 sequences grouped as a separate subclade. The possible reasons of this small inconsistency could be the incomplete sequences of many Cactaceae proteins, as well as the very restricted number of sequences in this phylogeny (Figure 5).

**DISCUSSION**

Although, it is well established that some KNOX transcription factors are important for the maintenance of the shoot apical meristem, their role in vascular cambium activity and secondary

growth is less well understood. Four class I KNOX genes were reported to be expressed in the vascular cambium of *P. tremula* (Schrader et al., 2004; Du et al., 2009), while the expression of a class II KNOX gene was detected in the vascular cambium of *J. nigra* (Huang et al., 2009). It was therefore of particular interest to determine whether genes of both classes are expressed in the vascular zones of cacti, and whether different paralogs are expressed in species producing different types of wood. By performing RT-PCR with degenerate primers in the four Cactaceae species (*A. retusus*, *F. pilosus*, *P. lychnidiflora*, and *P. alensis*) and RNA-seq with subsequent *de novo* transcriptome assembly for three of these species, we have found, and consequently confirmed by RT-PCR (Supplementary Figure S1), that the transcripts of both class I and class II KNOX genes are present in the cambial zone of the adult individuals in all four species (Table 1). The possibility exists that other KNOX genes could be expressed at very low levels in the Cactaceae vascular zone, impeding successful *de novo* assembly. Genome sequencing of species from the Cactaceae family will facilitate a more detailed analysis of the KNOX expression patterns. To annotate the Cactaceae KNOX sequences, a phylogenetic tree was constructed

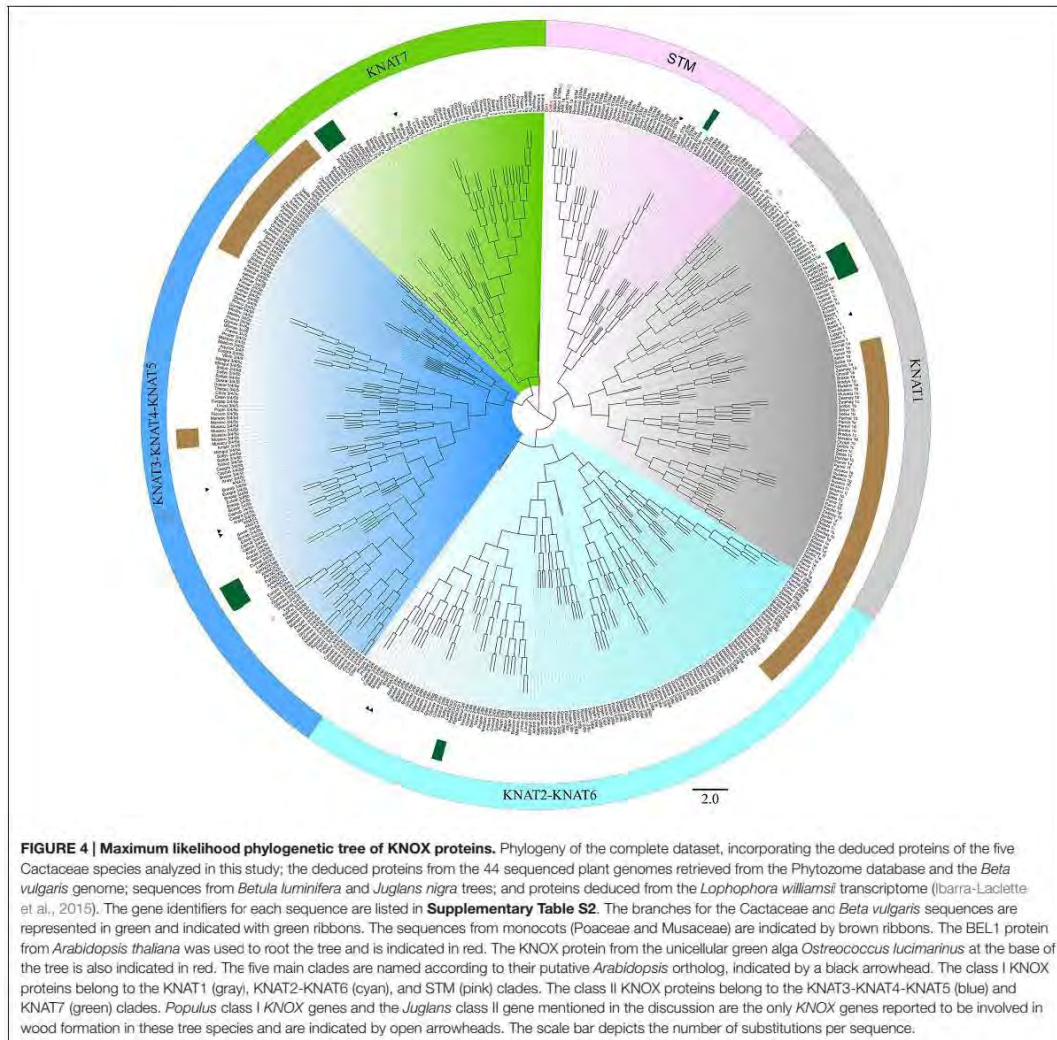


**FIGURE 3 |** Alignment of the deduced amino acid sequences of the class II KNOX proteins of Cactaceae and some class II KNOX proteins of *Arabidopsis*. The amino acid scores are highlighted with a grayscale background according to their identity and similarity only if the coverage in the alignment position is higher than 55%. As part of the KNOX1 domain is not covered in the incomplete protein sequences, the amino acids in the other proteins are not highlighted, despite a high percentage of identity/similarity. Domains were defined with InterProScan.

that included KNOX sequences from sequenced angiosperm genomes from the Phytozome database and from the *B. vulgaris* genome as a species from the order Caryophyllales, to which the Cactaceae family belongs. Expression of all KNOX genes in the cambial zone of adult *A. retusus*, *F. pilosus*, and *P. lychnidiflora* plants was confirmed by RT-PCR (Supplementary Figure S1). Meanwhile, in the tubercle of 5-year-old *A. retusus* plant, we detected expression of only one of the six *ArKNOX*

genes reported in this work, *ArKNOX7r1*. Expression of two KNOX genes, the orthologous *ArKNOX7r1/FpKNOX7r1* and *ArKNOX3a/FpKNOX3ae*, was detected in the vascular cambium of juvenile *A. retusus* and *F. pilosus* plants, when they were just starting to accumulate wood. This finding further suggests that the KNOX genes, reported in this work, are involved in the formation of mature wood, and that the *KNOX7r1* paralog has broader functions in Cactaceae development.

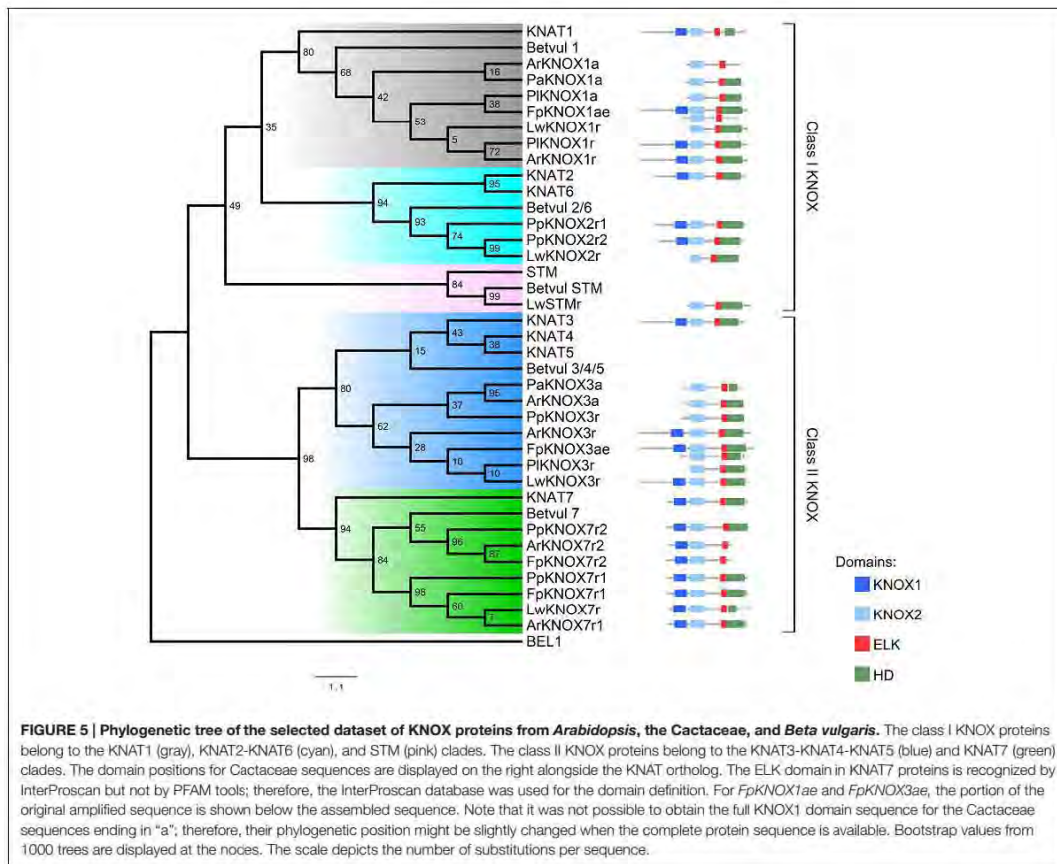




The six class I KNOX genes found to be expressed in the cambial zone of the four cactus species are putative orthologs of the *Arabidopsis* *BP/KNAT1* gene, as they belong to the KNAT1 clade. The *Populus* gene *ARK2/PttKNOX1*, which is expressed in the vascular cambium and developing xylem (Schrader et al., 2004; Du et al., 2009), also belongs to this clade (white arrowhead in **Figure 4**). Previous work has suggested that *ARK2* is involved in vascular cambium activity and secondary growth; *Populus* plants constitutively overexpressing *ARK2* had a wider cambial zone and decreased differentiation of the lignified secondary xylem, while the downregulation of *ARK2* resulted in the early differentiation of lignified secondary xylem cells (Du et al., 2009).

Importantly, we did not find putative orthologs of the *Arabidopsis* *STM* and *KNAT2-KNAT6* genes expressed in the Cactaceae cambial zone. By contrast, both the *ARK1a* and *ARK1b* paralogs in *Populus* (orthologous to *STM*) are shown to be expressed in the vascular cambium (Popgenie database<sup>8</sup>). *ARK1* overexpression in a hybrid *Populus* resulted in pleiotropic phenotypes, including the slower differentiation of cambium-derived cells (Groover et al., 2006). While the *KNAT1* ortholog *ARK2* is expressed in both the vascular cambium and the developing xylem of *Populus*, *ARK1a* was shown to be

<sup>8</sup><http://popgenie.org/>



downregulated in non-meristematic secondary vascular tissues (Groover et al., 2006), thus showing more similarity to the *STM* expression pattern in the shoot apex. *STM* expression was previously also detected in the *Arabidopsis* inflorescence stems during induced secondary growth (Ko and Han, 2004). As expected, we did not find a *STM* ortholog in the *P. pringlei* root tip transcriptome, while two *P. pringlei* paralogs orthologous to KNAT2-KNAT6 were expressed in the root tip (Figures 4, 5 and Table 1). The *STM* putative ortholog was present in the *L. williamsii* shoot and root transcriptome, suggesting that *STM* might be expressed in some Cactaceae tissues, most probably in the shoot apical meristem. The absence of putative *STM* orthologs in the cambial zone transcriptome suggests possible differences in the vascular cambium genetic regulatory network in the Cactaceae versus *Arabidopsis* and *Populus*. Interestingly, monocots do not have orthologs of *Arabidopsis STM*; the grass genes involved in shoot apical meristem maintenance (maize: *KNOTTED1* and *ROUGH SHEATH1*; rice: *ORYZA SATIVA HOMEBOX1 (OSH1)* and *OSH15*; Tsuda et al., 2011;

Bolduc et al., 2014) belong to the KNAT1 clade (Figure 4). Moreover, monocot *KNOX* genes clustered as a single subclade within the KNAT1, KNAT2-KNAT6, and KNAT7 clades, while within the KNAT3-KNAT4-KNAT5 clade two subclades, one from Poaceae and another from Musaceae, were present (brown ribbons on Figure 4).

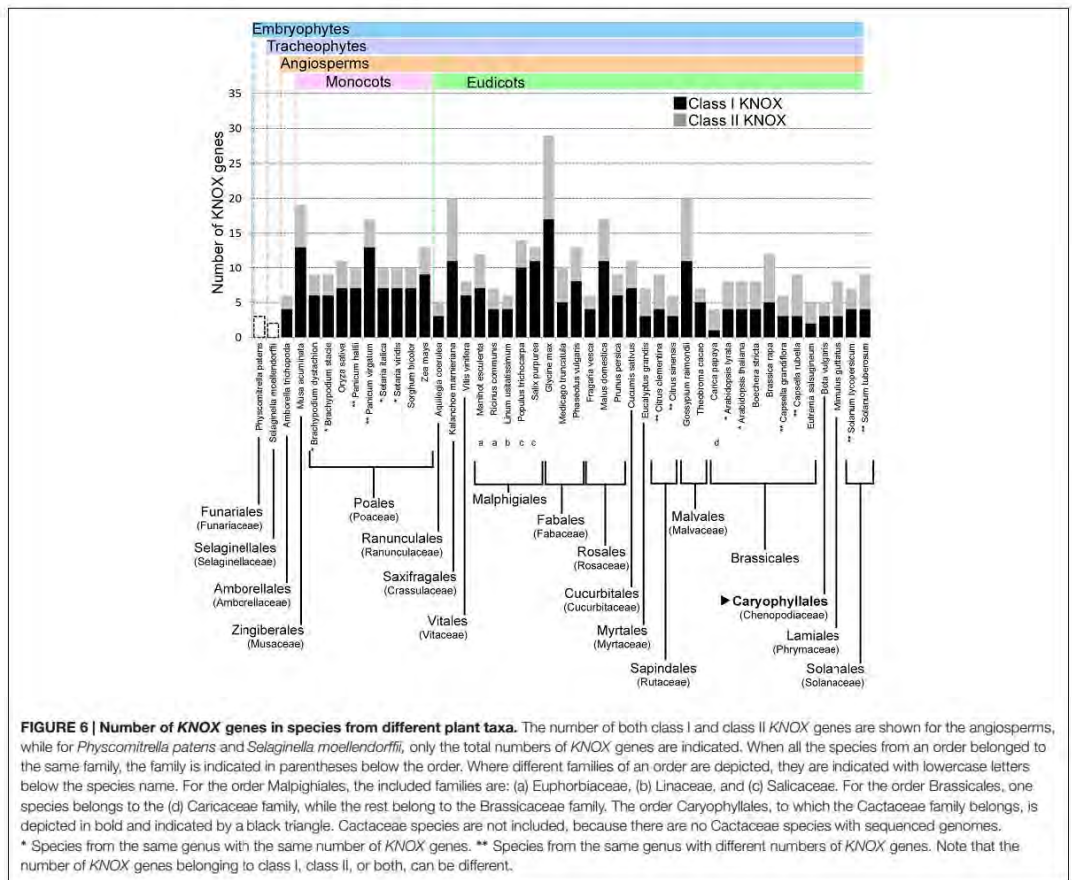
For the three species used for RNA-seq and the *de novo* transcriptome assembly, two class I paralogs were identified in the vascular zone of *A. retusus* and *P. lychnidiflora*, while only one was identified in *F. pilosus* (Figures 2, 4, 5 and Table 1). The missing class I *KNOX* paralog could have been lost from *F. pilosus* as an evolutionary developmental process, enabling speciation after a gene duplication in the Cactaceae, a phenomenon that has been well documented in other families. Using data from sequenced genomes (Supplementary Table S2), we showed that the number of class I and class II *KNOX* genes in a moss, a lycophyte, and 43 angiosperm species varies significantly, particularly between Eudicotyledonous species (Figure 6). As there is still no Cactaceae species with a sequenced



**TABLE 1 |** Number of *KNOX* genes from this study expressed in the cambial zone, root tip, and shoot/root of Cactaceae species.

Species	Genes expressed in	Number of <i>KNOX</i> orthologs				
		Class I clades			Class II clades	
		<i>STM</i>	<i>KNAT1</i>	<i>KNAT2-KNAT6</i>	<i>KNAT3-KNAT4-KNAT5</i>	<i>KNAT7</i>
<i>Pereskia lychnidiflora</i>	Cambial zone	0	2	0	1	0
<i>Ariocarpus retusus</i>	Cambial zone	0	2	0	2	2
<i>Ferocactus pilosus</i>	Cambial zone	0	1	0	1	2
<i>Pilosocereus alensis</i> <sup>a</sup>	Cambial zone	0	1	0	1	0
<i>Pachycereus pringlei</i>	Root tip	0	0	2	1	2
<i>Lophophora williamsii</i> <sup>b</sup>	Shoot and root	1	1	1	1	1

<sup>a</sup> No RNA-seq data available. <sup>b</sup> For comparison, the data from Ibarre-Laclette et al. (2015) were used.



genome, Cactaceae were not included in this analysis. All of the analyzed angiosperm species have several class II genes, and all but *Carica papaya* have more than one class I gene. There are 29 *KNOX* genes in *Glycine max* (17 class I and

12 class II genes) but only four in *Carica* (one class I and three class II genes). Furthermore, different species can have a different number of paralogs in every clade (Supplementary Table S2 and Figure 5). The number of *KNOX* genes can



also vary within species of the same family; for instance, within diploid species of the Brassicaceae family (which includes the largest number of species with sequenced genomes), the number of *KNOX* genes varies from five in *Eutrema salsugineum* to 12 in *Brassica rapa*. Even species within a single genus can have variable numbers of *KNOX* genes; while the *Arabidopsis* species *A. thaliana* and *A. lyrata* both possess eight *KNOX* genes, there are six and nine genes in the *Capsella grandiflora* and *C. rubella* genomes, respectively (Figure 6 and Supplementary Table S2).

We found that two putative orthologs of the *Arabidopsis* class II *KNOX* paralogs *KNAT3*, *KNAT4*, and *KNAT5* were expressed in the cambial zone of *A. retusus*, but just one was expressed in the cambial zone of *F. pilosus* and *P. lychnidiflora* (Figures 3–5 and Table 1). Only one putative paralog expressed in the root tip of *P. pringlei* belonged to this class. One paralog from the published shoot and root transcriptome of *L. williamsii* (Ibarra-Laclette et al., 2015) also belongs to this class. Remarkably, among the three species used for cambial zone transcriptome assembly, two paralogs in *A. retusus* and *F. pilosus* belonged to the *KNAT7* clade, whereas no putative *KNAT7* ortholog was found in the cambium of *P. lychnidiflora* (*P. alensis* was not considered as the transcriptome assembly was not performed for this species). *KNAT7* is a transcriptional repressor of secondary cell wall biosynthesis (Liu et al., 2014). *P. lychnidiflora* has fibrous wood with significant cell wall lignification, *A. retusus* has non-fibrous wood with abundant wide-band tracheids and less lignification, while in the dimorphic wood of *F. pilosus*, the wide-band tracheids develop before more lignified cells. Our findings therefore suggest that the repressor activity of the *KNAT7* orthologs during secondary cell wall biosynthesis could explain the differences in wood type among these species, including the lower lignification level observed in *A. retusus* and *F. pilosus*.

We found that the transcripts of both *KNOX* gene classes are expressed in the cambial zone of the four Cactaceae species studied, and therefore could be expressed either in different cell types or in the same cell type within the vascular zone. Several *KNOX* proteins have been shown to selectively move from one cell layer to another (reviewed by Han et al., 2014), suggesting that even if the class I and II genes are transcribed in different cell types, the encoded proteins could coexist in the same cell as a result of protein trafficking. In any case, as no evidence of mutual repression was found when class I and class II *KNOX* loss-of-function mutations of *Arabidopsis* were combined (Furumizu et al., 2015), *KNOX* proteins belonging to different classes could exert their functions within the same cell. *KNOX* proteins are known to interact with a sister group of proteins from the BEL1-LIKE-HOMEODOMAIN (BHL, or BELL from the founding *BELL 1* gene) family; both *KNOX* and BELL proteins belong to the TALE (three amino acid loop extension) superclass of homeobox proteins. The *KNOX*-BELL heterodimer formation affects its cellular localization and the *KNOX* target selection (reviewed by Di Giacomo et al., 2013; Arnaud and Pautot, 2014). From heterologous expression experiments it was proposed that particular *KNOX* proteins could interact with different BELL partners leading to numerous combinations with distinct activities, and thus

regulating different sets of targets, including transcription factors and hormonal pathways, and ultimately influence multiple plant developmental processes. Recently, however, a more specific selectivity was suggested for the *in vivo* *KNOX*-BELL interactions (Furumizu et al., 2015), which could also enhance the properties of each heterodimerization partner; thus, the interaction between the transcriptional repressor *KNAT7* and its partner BHL6 could enhance their activity in repressing genes involved in secondary cell wall biosynthesis (Liu et al., 2014).

Knowledge of the processes of duplication and subfunctionalization of the regulatory genes helps us understand the evolution of the different aspects of plant development (Pires and Dolan, 2012), including the evolution of vascular development. Our work contributes to the elucidation of the mechanisms by which different wood types are formed in the Cactaceae, and provides insight into the evolutionary history of *KNOX* genes in different angiosperm species and their role in the speciation of land plants.

## AUTHOR CONTRIBUTIONS

TT, JR-R, AV, FV-S, GR-A, and SS designed the work; TT, JR-R, EP, AV, GR-A, and SS performed the experiments and analyzed the data; GR-A and JR-R prepared the figures; and TT and SS wrote the manuscript. All the authors have read and approved the manuscript.

## ACKNOWLEDGMENTS

Funding was provided by the DGAPA-PAPIIT, UNAM (grants IN209012 and 210115 to TT and IN207115 to SS) and by Consejo Nacional de Ciencia y Tecnología (CONACYT) [CB2014-240055 grant to SS; Ph.D. scholarships 220343 to JR-R; 367139 to GR-A and 700638 to EP, and a postdoc grant (220343) to JR-R]. We thank Jorge Nieto Sotelo for the valuable comments, Paul Gaytan, Eugenio Lopez and personnel of the DNA synthesis and sequencing unit of the IBt-UNAM for oligonucleotide synthesis, and Laura Márquez Valdelamar for DNA sequencing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00218/full#supplementary-material>

**FIGURE S1 | Class I and II *KNOX* transcript expression detected in cambial zone samples by end point RT-PCR.**

**FIGURE S2 | The identity and similarity matrix of the deduced *KNOX* proteins identified in this study. (A) Class I proteins. (B) Concatenated *KNOX1*, *KNOX2*, ELK, and HD domains of class I proteins. (C) Class II proteins. (D) Concatenated *KNOX1*, *KNOX2*, ELK, and HD domains of class II proteins.**

**TABLE S1 | PCR primers used in this study.**

**TABLE S2 | The identifiers of the *KNOX* sequences used for molecular phylogenetic analysis.**

## REFERENCES

- Altesor, A., Silva, C., and Ezcurrea, E. (1994). Allometric neoteny and the evolution of succulence in cacti. *Bot. J. Linn. Soc.* 114, 283–292. doi: 10.1111/j.1095-8339.1994.tb01937.x
- Arnaud, N., and Pautot, V. (2014). Ring the BELL and tie the KNOX: roles for TALEs in gynoecium development. *Front. Plant Sci.* 5:93. doi: 10.3389/fpls.2014.00093
- Berlyn, G. P., and Miksche, J. P. (1976). *Botanical Microtechnique and Cytochemistry*. Ames, IA: Iowa State University Press.
- Bolduc, N., Tyers, R. G., Freeling, M., and Hake, S. (2014). Unequal redundancy in maize *knotted1* homeobox genes. *Plant Physiol.* 164, 229–238. doi: 10.1104/pp.113.228791
- Di Giacomo, E., Iannelli, M. D., and Frugis, G. (2013). TALE and shape: how to make a leaf different. *Plants* 2, 317–342. doi: 10.3390/plants2020317
- Dohm, J. C., Minoche, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., et al. (2013). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505, 546–549. doi: 10.1038/nature12817
- Du, J., Mansfield, S. D., and Groover, A. T. (2009). The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. *Plant J.* 60, 1000–1014. doi: 10.1111/j.1365-3113.2009.04017.x
- Evert, R. F. (2006). *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development*, 3rd Edn. Hoboken, NJ: John Wiley & Sons.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. doi: 10.1007/BF01734359
- Furumizu, C., Alvarez, J. P., Sakakibara, K., and Bowman, J. L. (2015). Antagonistic roles of *KNOX1* and *KNOX2* genes in patterning the land plant body plan following an ancient gene duplication. *PLoS Genet.* 11:e1004980. doi: 10.1371/journal.pgen.1004980
- Groover, A. T., Mansfield, S. D., DiFazio, S. P., Dupper, G., Fontana, J. R., Millar, R., et al. (2006). The *Populus* homeobox gene *ARBORKNOX1* reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Mol. Biol.* 61, 917–932. doi: 10.1007/s11103-006-0059-y
- Groover, A. T., Nieminen, K., Helariutta, Y., and Mansfield, S. D. (2010). "Wood formation in *Populus*," in *Genetics and Genomics of Populus, Plant Genetics and Genomics: Crops and Models 8*, eds S. Jansson, R. Bhaleao, and A. Groover (Berlin: Springer Science+Business Media), doi: 10.1007/978-1-4419-1541-2\_10
- Han, X., Kumar, D., Chen, H., Wu, S., and Kim, J. Y. (2014). Transcription factor-mediated cell-to-cell signalling in plants. *J. Exp. Bot.* 65, 1737–1749. doi: 10.1093/jxb/ert422
- Hay, A., and Tsiantis, M. (2010). KNOX genes: versatile regulators of plant development and diversity. *Development* 137, 3153–3165. doi: 10.1242/dev.030049
- Henikoff, S., and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* 89, 10915–10919.
- Huang, Z., Meilan, R., and Woeste, K. (2009). A *KNAT3*-like homeobox gene from *Juglans nigra* L., *JnKNAT3*-like, highly expressed during heartwood formation. *Plant Cell Rep.* 28, 1717–1724. doi: 10.1007/s00299-009-0771-6
- Ibarra-Laclette, E., Zamudio-Hernández, F., Pérez-Torres, C. A., Albert, V. A., Ramírez-Chávez, E., Molina-Tortés, J., et al. (2015). *De novo* sequencing and analysis of *Lophophora williamsii* transcriptome, and searching for putative genes involved in mescaline biosynthesis. *BMC Genomics* 16:657. doi: 10.1186/s12864-015-1821-9
- Ko, J., and Han, K. (2004). *Arabidopsis* whole-transcriptome profiling defines the features of coordinated regulations that occur during secondary growth. *Plant Mol. Biol.* 55, 433–453. doi: 10.1007/s11103-004-1051-z
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Landrum, J. V. (2008). Wide-band tracheids from a southern African succulent and their responses to varying light intensities: a preadaptation for future water stress? *Int. Bot.* 4, 99–103. doi: 10.3923/ijb.2008.99.103
- Liu, L., Filkov, V., and Groover, A. (2014). Modeling transcriptional networks regulating secondary growth and wood formation in forest trees. *Physiol. Plant.* 151, 156–163. doi: 10.1111/pp.12113
- Liu, L., Zinkgraf, M., Petzold, H. E., Beers, E. P., Filkov, V., and Groover, A. (2015). The *Populus* ARBORKNOX1 homeodomain transcription factor regulates woody growth through binding to evolutionarily conserved target genes of diverse function. *New Phytol.* 205, 682–694. doi: 10.1111/nph.13151
- Loza-Cornejo, S., and Terrazas, T. (2011). Morfo-anatomía de plántulas en especies de Pachycereae: ¿hasta cuándo son plántulas? *Bol. Soc. Bot. Méx.* 88, 1–13.
- Lucas, W. J., Groover, A., Lichtenberger, R., Furuta, K., Yadav, S. R., Helariutta, Y., et al. (2013). The plant vascular system: evolution, development and functions. *J. Integr. Plant Biol.* 55, 294–388. doi: 10.1111/jipb.12041
- Mauseth, J. D. (2006). Structure-function relationships in highly modified shoots of Cactaceae. *Ann. Bot.* 98, 901–926.
- Mauseth, J. D., and Plemons, B. J. (1995). Developmentally variable, polymorphic woods in cacti. *Am. J. Bot.* 82, 1199–1205.
- Miyashima, S., Sebastian, J., Lee, J.-Y., and Helariutta, Y. (2013). Stem cell function during plant vascular development. *EMBO J.* 32, 178–193. doi: 10.1038/emboj.2012.301
- Nieminen, K., Blomster, T., Helariutta, Y., and Mähönen, A. P. (2015). Vascular cambium development. *Arabidopsis Book* 13:e0177. doi: 10.1199/tab.0177
- Pires, N. D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philos. Trans. R. Soc. B.* 367, 508–518. doi: 10.1098/rstb.2011.0252
- Reyes-Rivera, J., Canché-Escamilla, G., Soto-Hernández, M., and Terrazas, T. (2015). Wood chemical composition in species of Cactaceae: the relationship between lignification and stem morphology. *PLoS ONE* 10:e0123919. doi: 10.1371/journal.pone.0123919
- Růžička, K., Ursache, R., Hejätko, J., and Helariutta, Y. (2015). Xylem development—from the cradle to the grave. *New Phytol.* 207, 519–535. doi: 10.1111/nph.13383
- Ruzin, E. S. (1999). *Plant Microtechnique and Microscopy*. Oxford: Oxford University Press.
- Schrader, J., Nilsson, J., Mellerowicz, E., Berglund, A., Nilsson, P., Hertzberg, M., et al. (2004). A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell* 16, 2278–2292. doi: 10.1105/tpc.104.024190
- Serikawa, K. A., Martinez-Laborda, A., Kim, H.-S., and Zambryski, P. C. (1997). Localization of expression of *KNAT3*, a class 2 knotted1-like gene. *Plant J.* 11, 853–861.
- Spicer, R., and Groover, A. (2010). Evolution of development of vascular cambia and secondary growth. *New Phytol.* 186, 577–592. doi: 10.1111/j.1469-8137.2010.03236.x
- Tsuda, K., Ito, Y., Sato, Y., and Kurata, N. (2011). Positive autoregulation of a *KNOX* gene is essential for shoot apical meristem maintenance in rice. *Plant Cell* 23, 4368–4381. doi: 10.1105/tpc.111.090050
- Vázquez-Sánchez, M., and Terrazas, T. (2011). Stem and wood allometric relationships in Cactaceae (Cactaceae). *Trees* 25, 755–767. doi: 10.1007/s00468-011-0553-y
- Ye, Z.-H., and Zhong, R. (2015). Molecular control of wood formation in trees. *J. Exp. Bot.* 66, 4119–4131. doi: 10.1093/jxb/erv081
- Zhong, R., and Ye, Z.-H. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* 56, 195–214. doi: 10.1093/pccp/pcu140

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Reyes-Rivera, Rodríguez-Alonso, Petrone, Vasco, Vergara-Silva, Shishkova and Terrazas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.