

SCIENTIFIC REPORTS



OPEN

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

Gustavo Rodriguez-Alonso¹, Marta Matvienko², Mayra L. López-Valle¹, Pedro E. Lázaro-Mixteco¹, Selene Napsucially-Mendivil¹, Joseph G. Dubrovsky¹ & Svetlana Shishkova¹

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 growth¹. 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 suitable². 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^{3,4}, the adventitious roots of *Ficus pumila*⁵; and the lateral roots of *Zea mays*⁶, *Opuntia arenaria*, and *O. tunicata*⁷. 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*⁸, and *Pachycereus pringlei*⁹ 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, Cacteae, Pachycereae, Cereae, Trichocereae, Notocactae, Rhipsalideae¹⁰, and Hylocereae^{10,11}. Seedlings of all analysed species from the first six tribes always exhibit determinate growth

¹Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico. ²Tecan 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)

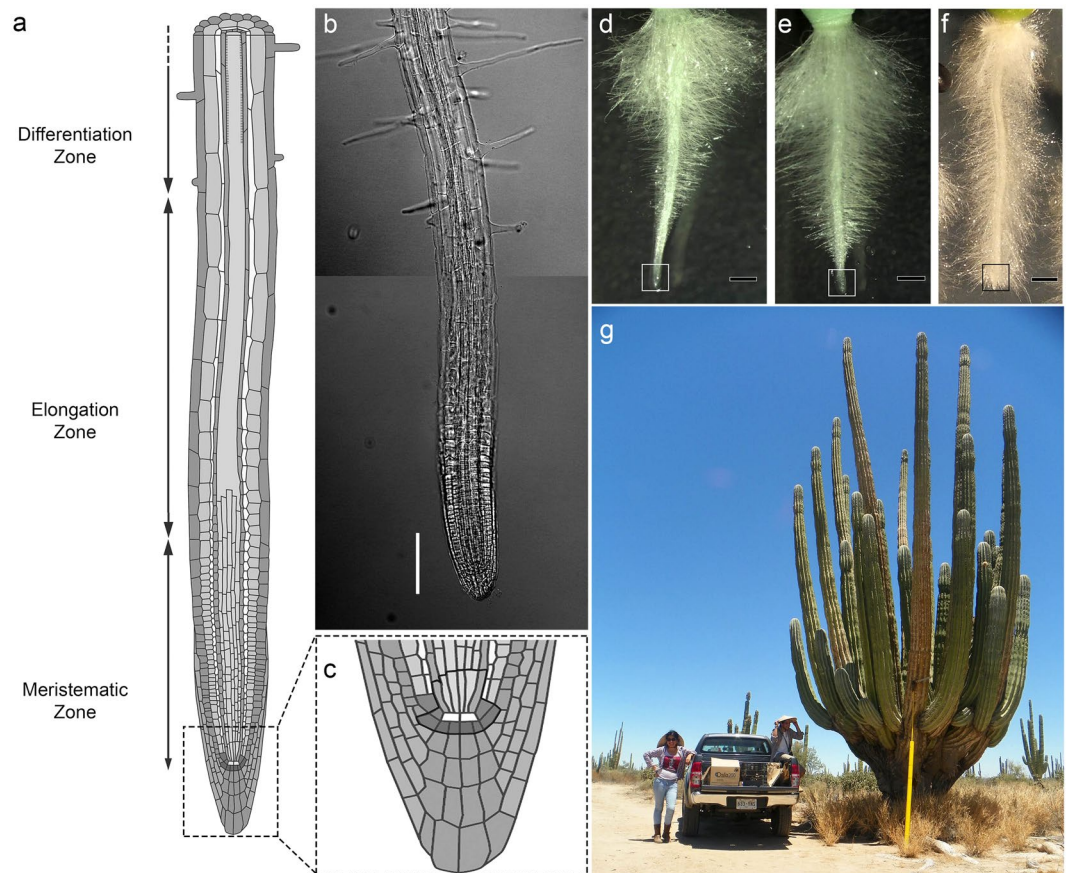


Figure 1. Indeterminate and determinate root growth. (a) Growing roots can be divided in three developmental zones along the longitudinal axis as illustrated here for the *Arabidopsis thaliana* primary root; the root apical meristem (RAM) is located in the meristematic zone and it is present and active 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 growth¹¹, while other seedlings from the same species exhibit indeterminate growth¹⁰ 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 development⁸. 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 condition⁸. These data, together with the incidence of RAM exhaustion in lateral roots and the determinate growth of roots regenerated from calli¹², 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 RAM¹⁰.

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 ⁶	Total contigs	49,045	Average length (nt)	1,080
Av. reads per sample	46.5 × 10 ⁶	N25	1,332	Max. length (nt)	14,583
Paired-end reads	79 × 10 ⁶	N50	728	Min. length (nt)	400
Merged reads	88 × 10 ⁶	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 *Pachycereus pringlei* primary root apex.

PLETHORA (PLT) transcription factors, which belong to the AP2/ERF superfamily¹³. PLT proteins are partially redundant, their activities are dose-dependent, and they share a high number of target genes¹⁴. 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 zone¹⁵. 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 growth¹⁵.

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^{16,17}. As a result, the RAM becomes disorganized and is lost, causing determinate growth^{18–20}. Another transcription factor specifically expressed in the QC and involved in stem cell homeostasis in the root is WUSCHEL-RELATED HOMEODOMAIN 5 (WOX5)^{21,22}. 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^{21,23}. When combined with *scr*, *shr*, or *plt1plt2* mutations, the *wox5* mutation redundantly accelerates RAM exhaustion²¹.

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 reported²⁴. 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 genome²⁵. 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* transcriptome²⁶. 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 species²⁷. 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 KAAS²⁸ 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 genome²⁹ at the time of the analysis. Shortly after this study, the genomes of the Caryophyllales *Amaranthus hypocondriacus*³⁰ and *Chenopodium quinoa*³¹ 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

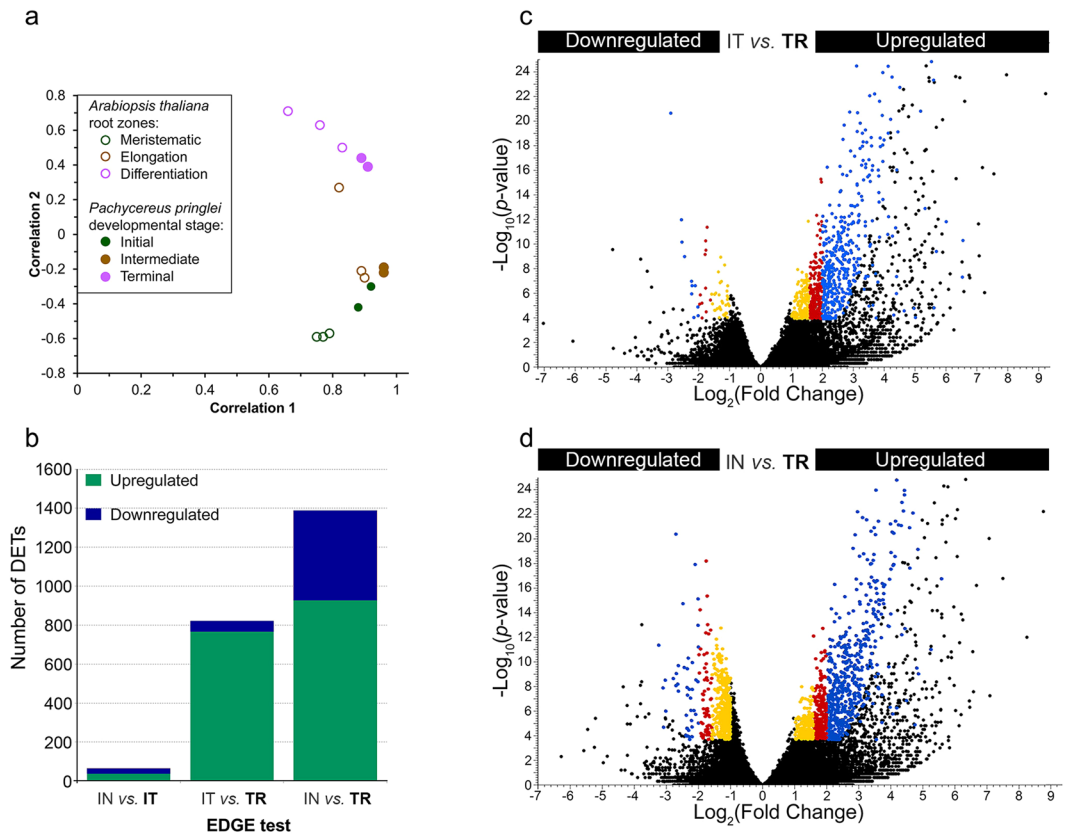


Figure 3. Characterization of the *Pachycereus pringlei* transcriptome. **(a)** Principal component analysis of *P. pringlei* developmental stages and *Arabidopsis* root zones and **(b–d)** the empirical detection of gene expression shows that the transcriptional state of the *P. pringlei* root tip changes 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 ≤ 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 stages³⁷. Transcripts were considered to be DETs when they met these criteria: fold change ≥ 2 , FDR p -value ≤ 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 Schiefelbein³⁵ (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 test³⁸ 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 method³⁹. 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

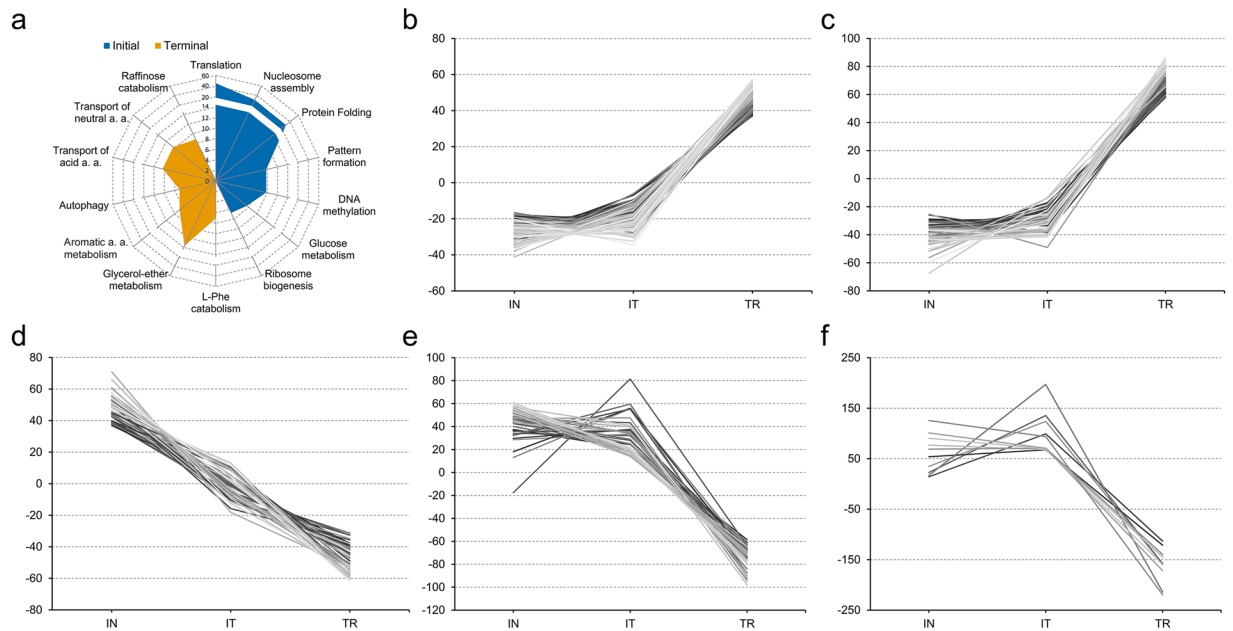


Figure 4. Expression profiles of the differentially expressed transcripts in the *Pachycereus pringlei* root apex. (a) BLAST2GO biological processes overrepresented in the Initial and Terminal developmental stages of the primary root apex of *P. pringlei*. Axis: $-\log_{10}(p\text{-val})$ of the enrichment 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*²⁶. 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 2.25×10^{-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

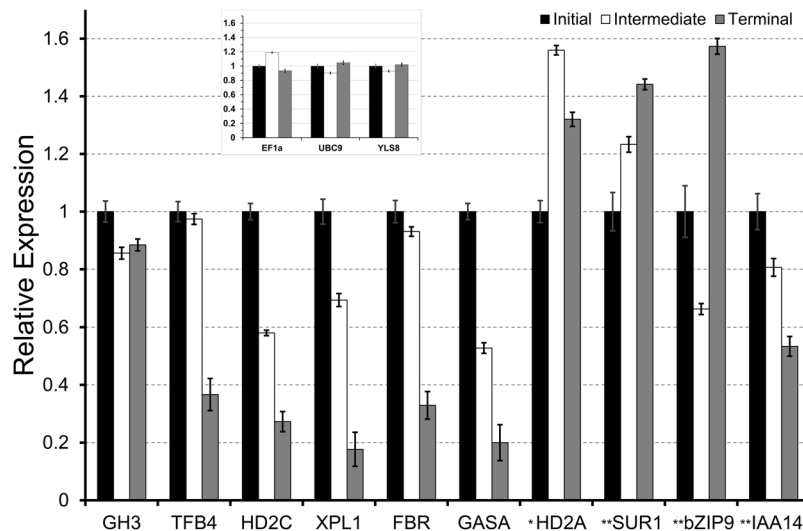


Figure 5. The relative expression levels of selected *Pachycereus pringlei* transcripts quantified using RT-qPCR assays for the primary root apex at the three stages of primary root development. The expression levels were normalized relative to *PpEF1 α* , *PpUBC9*, and *PpYLS8*, respectively (inset). *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 hybridization¹⁰ 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 genes⁴⁰. 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). *PpEF1 α* , *PpUBC9*, and *PpYLS8*, were selected as internal references. The expression patterns of *PpTFB4*, *PpHD2C*, *PpXPL*, *PpFIBR*, *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 Pachycereeae 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; *PpbZIP9*, 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 genes⁴¹. We therefore generated a transcriptional regulatory network (TRN) by retrieving a set of genes that were reported as important regulators of root development^{35,42}. These genes and their first neighbours, together with their interactions, were extracted from the *Arabidopsis* transcriptional regulatory map⁴³. 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

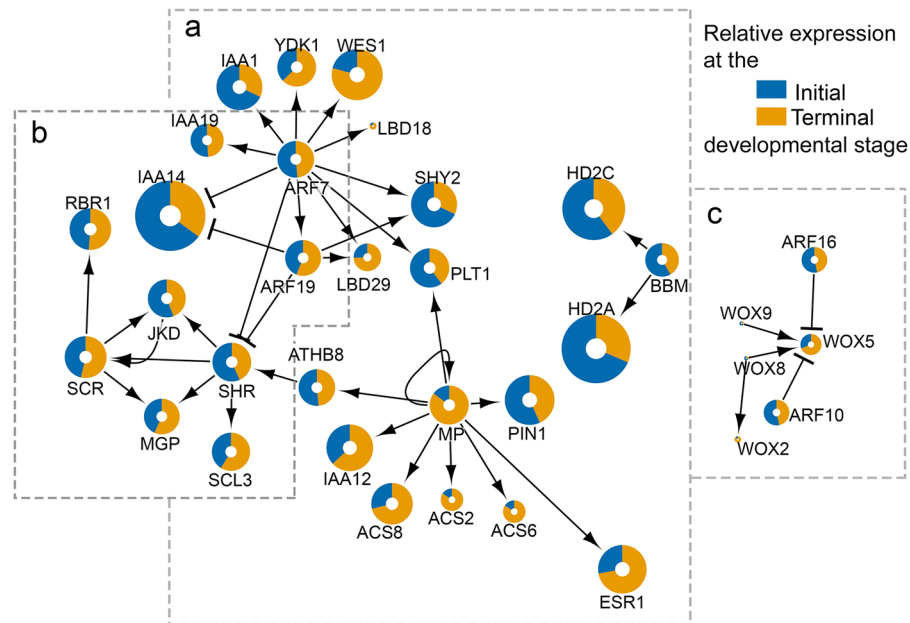


Figure 6. Root development modules of the *Pachycereus pringlei* inferred transcriptional regulatory network. The complete network (Fig. S9) was separated into three distinct modules: (a) PpPLT, which includes PpPLT1 and PpBBM as separate clusters; (b) PpSHR-PpSCR; and (c) PpWOX. Node size represents 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 PLTs¹⁴ 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^{44,45}; 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 formation¹⁶, vascular differentiation⁴⁶, and transition from cell division to cell elongation⁴⁷, 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^{48,49}. 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 cells²¹, 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 Pachycereae tribe species, with a high-coverage genome of *Carnegiea gigantea* and low-coverage genome of three more columnar cacti, including *P. pringlei*²⁴.

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 described³⁶ 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 (KAAS²⁸). 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) method⁵⁰. 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* root³⁵. Differential expression was assessed with EDGE test³⁷ as implemented in the CLC GW. The differentially expressed transcripts (DETs) were extracted according to the following criteria: fold change ≥ 2 , FDR *p*-corrected value ≤ 0.005 , RPKM ≥ 3 in all samples. DETs were subjected to a hypergeometric test³⁸ and *k*-mean clustered³⁹ 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^{51,52}. 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 map⁴³ (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 SuperFamily⁵³, were present. The resulting transcriptional regulatory network (TRN) was visualized in Cytoscape⁵⁴ (v3.2.1).

References

- Sebastian, J. & Lee, J. Root Apical Meristems. In: *eLS*. John Wiley & Sons, Ltd: Chichester, <https://doi.org/10.1002/9780470015902.a0020121.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, B. 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).
- Boke, N. H. Root glochids and root spurs of *Opuntia arenaria* (Cactaceae). *Am. J. Bot.* **69**, 1085–1092 (1979).
- Dubrovsky, J. G. Determinate primary-root growth in seedlings of Sonoran Desert Cactaceae; its organization, cellular basis, and ecological significance. *Planta* **203**, 85–92 (1997).
- Dubrovsky, J. 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, L. 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., Nawy, 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., Reisewitz, 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., Ritland, K. & Hamrick, J. L. Mating system of *Pachycereus pringlei*: an autotetraploid cactus. *Heredity* **72**, 86–94 (1994).
28. Moriya, Y., Itoh, M., Okuda, 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. Azpeitia, E., Weinstein, N., Benítez, 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., Blilou, 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-Ramírez, 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 Paep, A., Speleman, 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**, 2498–24504 (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