Title: Characterization of the *KNOTTED1-LIKE HOMEOBOX (KNOX)* gene family in *Pinus pinaster* Ait.

Authors: Natalia Bueno[1], José Manuel Alvarez[1*], Ricardo J. Ordás

Address: Instituto Universitario de Biotecnología de Asturias (IUBA), Departamento de Biología de Organismos y Sistemas, Universidad de Oviedo, Spain

[1] Natalia Bueno and José M. Alvarez contributed equally to this work.

* Corresponding author:

Telephone: +34 985 45 81 06

Email: alvarezmanuel@uniovi.es

ABSTRACT

KNOTTED1-LIKE HOMEOBOX (*KNOX*) genes are a family of plant-specific homeobox transcription factors with important roles in plant development that have been classified into two subfamilies with differential expression domains and functions. Studies in angiosperms have shown that class I members are related to the maintenance of meristem homeostasis and leaf development, whereas class II members promote differentiation of tissues and organs. However, little is known about its diversification and function in gymnosperms. By combining PCR-based detection and transcriptome data analysis, we identified four class I and two class II *KNOX* genes in *Pinus pinaster*. Expression analyses showed that class I members were mainly expressed in meristematic regions and differentiating tissues, with practically no expression in lateral organs, whereas expression of class II members was restricted to lateral organs. Furthermore, overexpression of *P. pinaster KNOX* genes in *Arabidopsis thaliana* caused similar phenotypic effects to those described for their angiosperms counterparts. This is the first time to our knowledge that functional analyses of class II members are reported in a conifer species. These results suggest a high conservation of the *KNOX* gene family throughout seed plants, as the functional differentiation of both subfamilies observed in angiosperms might be partially conserved in gymnosperms.

KEYWORDS: *KNOTTED1-LIKE HOMEOBOX* (*KNOX*), meristem maintenance, *Pinus pinaster*, tissue differentiation.

1. INTRODUCTION

KNOTTED1-LIKE HOMEOBOX (KNOX) genes are a group of plant-specific transcription factors that belong to the TALE superclass of homeobox genes, which are characterized by the presence of a three amino acid loop extension (TALE) motif between helices 1 and 2 of the homeodomain (HD) [1,2,3]. The name of this family comes from its founding member, designated *Knotted1* (*Kn1*), which in turn was the first gene encoding a homeobox protein isolated in plants. This gene was isolated in maize gain-of-function mutants that showed altered leaf development [4]. Since then, *KNOX* genes have been identified in all land plant groups and in specific Phyla of green algae, but not in red algae [5]. All KNOX proteins are characterized by the presence of several conserved domains [6,7,8]. The TALE-type HD, located in the C-terminal region of the protein, is responsible for the recognition of promoter sequences in downstream target genes; the ELK domain, situated upstream the HD, is thought to act as a nuclear localization signal to participate in transcriptional repression, and could facilitate protein-protein interactions; and the MEINOX domain, situated in the N-terminal half of the protein, includes KNOX1 and KNOX2 subdomains separated by a poorly conserved region, which are thought to participate in suppressing target gene expression and homo-dimerization, respectively.

The *KNOX* gene family has a monophyletic origin, as shown by phylogenetic analyses based on the homeodomain, the MEINOX domain or the full sequence [9]. *KNOX* genes have been traditionally classified in two subfamilies, class I and class II, based on sequence similarity, intron position, phylogenetic relationships and expression criteria [10]. Recently, a new subfamily of *KNOX* genes that lacks the ELK-HD region was described in some eudicot species, which was designated class M [5,11].

Among seed plants, the *KNOX* gene family has been extensively studied in the model species *Arabidopsis thaliana*, which contains four class I genes (*SHOOT MERISTEMLESS* or *STM*, *BREVIPEDICELLUS/KNOTTED IN ARABIDOPSIS THALIANA 1* or *BP/KNAT1*, *KNAT2* and *KNAT6*), four class II genes (*KNAT3*, *KNAT4*, *KNAT5* and *KNAT7*), and one class M gene (*KNATM*).

Class I *KNOX* genes have received special attention, since they were found to play important roles in plant growth and development. These genes are expressed mainly in meristematic regions and less differentiated tissues, but not in lateral (mature) organs, and have been associated to the maintenance of meristematic potentials [9,12,13,14,15]. *STM* participates in stem cell pool maintenance by inducing *ISOPENTYL TRANSFERASE* (*IPT*) gene expression and therefore cytokinin biosynthesis in the shoot apical meristem (SAM), and its expression is also induced by cytokinins [16,17,18]. *STM* has also an essential role in the floral meristem and carpel formation [19]. This gene positively regulates the expression of class I members *BP/KNAT1* and *KNAT2*, which participates redundantly with *STM* in stem cell maintenance and carpel development, respectively [20,21]. *KNAT6* also was shown to have redundant activities with *STM* in SAM maintenance and the establishment of the boundaries between SAM and cotyledons during embryogenesis [22].

Class II *KNOX* genes, in contrast, have a broader expression pattern. They are expressed both in differentiating tissues and mature organs in flowering plants, but not in meristematic zones [23]. In *A. thaliana*, it has been proposed that class II *KNOX* genes have opposed roles to those described for class I *KNOX* genes, as they promote differentiation of aerial organs and suppress meristematic capability [23]. It remains unclear how both classes of genes carry out their function. It has been proposed that this could be due to mutual repression, opposing modes of transcriptional regulation,

regulation of different of downstream gene targets, or modification of their activity by other proteins.

In addition, studies on leaf development in angiosperms [24,25,26,27,28,29] showed that class I *KNOX* overexpression causes overproliferation, while the loss of class I *KNOX* function can cause premature differentiation, leading to more complex and simplified leaves, respectively. On the other hand, class II *KNOX* genes are required for promoting tissue maturation. Therefore, loss of class II *KNOX* function leads to overproliferation and leaf complexity and the gain of class II *KNOX* function causes leaf simplification.

Despite the extensive research done in *A. thaliana*, little is known about the *KNOX* gene family in gymnosperms. Four class I *KNOX* genes, designated *KN1* to *KN4*, have been isolated in several spruce and pine species to date [30,31,32], and some of them have been involved in SAM formation and maintenance [33,34]. Recently, identification of members of the class II subfamily was reported in some gymnosperm species such as *Gnetum gnemon* and *Picea abies* among others [23]. No class M genes have been described in gymnosperms to date, and it is unclear whether this is due to the lack of class M members in the gymnosperm lineage or to an incomplete sampling. Although gymnosperm and angiosperm species share morphological and physiological features, the patterning during embryogenesis differs significantly, and there are key differences that may alter the underlying genetic programs. At the moment, it is not known whether the model of genic expression during angiosperm development may be applicable to conifers. Thus, analysis of the tissue specific expression of *KNOX* genes using other model species outside the angiosperms are needed to elucidate similarities and differences in the regulatory mechanisms of plant development.

Our starting work hypothesis is that the established functional differentiation of class I and II subfamilies of *KNOX* genes in angiosperms may be partially conserved in gymnosperms and therefore the model proposed for *A. thaliana* can be extrapolated to

gymnosperms. Here, we present the characterization of the *KNOX* gene family in *Pinus pinaster*, a pine species native to the Mediterranean region that has been established as the model conifer species from Southwest Europe, which includes four class I and two class II members. The analysis of their expression pattern in plant and ectopic overexpression of these genes in *A. thaliana* showed that both *KNOX* subfamilies maintained, at least partially, the functional differentiation observed in angiosperms.

2. MATERIAL AND METHODS

2.1 Identification and isolation of *KNOX* genes in *Pinus pinaster*

Four class I *KNOX* genes from *Picea abies*, *Picea glauca*, *Picea mariana*, *Pinus taeda* and *Pinus strobus* were found using the public database GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Based on the high degree of similarity between conifer sequences, specific primers were designed for the isolation of their respective orthologues in *P. pinaster* using the Primer3 software [35]. Specifically, using cDNA obtained from *P. pinaster* embryos as template, two sets of primers were used in nested PCRs for the isolation of each gene in order to improve the specificity of the amplification. PCR products were cloned using CloneJET PCR Cloning Kit (Thermo Scientific, Waltham, MA, USA) and sequenced (at least three clones per band) at the University of Oviedo DNA Analysis Facility (Spain). The complete cDNA and partial genomic sequences were obtained by Rapid Amplification of cDNA Ends (RACE) using the FirstChoice RLM-RACE kit (Ambion, Applied Biosystems Inc., Foster City, CA, USA) and by genome walking through the GenomeWalker Universal Kit (Clontech Laboratories, Mountain View, CA, USA), respectively. All primers used for the isolation of class I *KNOX* genes in *P. pinaster* are included in Table S1 (available as Supplementary Material).

In order to identify new *KNOX* members in *P. pinaster*, we carried out a screening of *P. pinaster* transcriptome and proteome data obtained in the frame of the European projects ProCoGen [36] and SustainPine (http://www.scbi.uma.es/sustainpinedb/home_page) [37], and in the PLAZA Gymnosperms (https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/) database. Sequences containing the characteristic KNOX conserved domains were screened using the BLASTP and/or TBLASTN algorithms [38], and HMM profile via HMMER (http://hmmer.org/) with default settings. *P. pinaster* isolated sequences were used as queries for the identification of new class I members. As scarce class II and no class M *KNOX* genes had been described in any conifer species, we also conducted a screening using the conserved domains from *A. thaliana* class II and class M KNOX proteins as queries. Only sequences with a significant similarity were considered in this study. When possible, isolated genes in *P. pinaster* were designated as *PpKN* followed by a number based on their homology with *KNOX* genes previously described in other conifer species. Exon-intron patterns were also determined through the comparison with genome data. Similarly, a search for class II members in other conifer species such as *P. abies* and *P. taeda* in the ConGenIE (http://congenie.org/) and PLAZA Gymnosperms databases was performed. All sequences obtained were uploaded to GenBank and their accession numbers are included in Table S2 (available as Supplementary Material). In order to obtain the percentage of pairwise identity between PpKN paralogs, alignments of the protein sequences were carried out through Geneious software using the cost matrix Blosum62.

2.2 Phylogenetic analysis

KNOX sequences identified in this work from *P. pinaster*, *P. abies* and *P. taeda*, along with other KNOX protein sequences, found in public databases, from the green algae *Ostreococcus tauri*, *Chlamydomonas reinhardtii*, *Micromonas* sp. and *Acetabularia*

acetabulum; the moss *Physcomitrella patens*; the spikemoss *Selaginella moellendorffii*; the fern *Ceratopteris richardii*; the conifers *P. abies*, *P. mariana*, *P. glauca*, *P. pinaster*, *P. taeda* and *P. strobus*; and the flowering plants *Amborella trichocarpa* (basal angiosperm), *A. thaliana* (eudicot), *Zea mays* and *Oryza sativa* (monocots) were used for the phylogenetic analysis. Accession numbers for all sequences are listed in Table S2 (available as Supplementary Material). Protein sequences were aligned using the MAFFT plug-in in Geneious software (Biomatters Ltd., New Zealand). Alignments were edited manually in order to eliminate gaps and limit the analysis to the conserved regions. The unrooted consensus trees were generated through Geneious software by the Neighbour-Joining method and the Jukes-Cantor genetic distance model from 100 bootstrap replicates, using the green alga OtKNOX sequence as outgroup for the trees.

We also calculated non-synonymous (Ka) and synonymous (Ks) nucleotide substitution rates for the *P. pinaster KNOX* gene family through the Computational Biology Unit (CBU) Ka/Ks Calculation tool (http://services.cbu.uib.no/tools/kaks), using the parsimony method to generate the resulting phylogenetic tree.

2.3 Expression analysis of *KNOX* genes in *Pinus pinaster*

2.3.1 Plant material

2.3.1.1 *Somatic embryogenesis*

Somatic embryogenesis in *P. pinaster* was used to study the expression of *KNOX* genes and their role in the physiological and molecular mechanisms of embryogenesis in conifers. Four different developmental stages from P5LV4.1 embryogenic line [39,40] were collected along the maturation process: proembryogenic masses (PEM) proliferating in the presence of the plant growth regulators auxins and cytokinins; early embryos (EE) with a translucent embryo proper and a long suspensor; late embryos (LE) with a prominent and opaque embryo proper; and mature embryos (ME) with well-defined apical

meristem and cotyledons [41,42]. Samples were snap frozen in liquid nitrogen and stored at -80 °C until analysis.

2.3.1.2 *Germinating embryos, seedlings and adult material*

Mature seeds from open pollinated *P. pinaster* trees from ES08 Meseta Castellana provenance (Spain) were provided by "Servicio de Material Genético del Ministerio de Medio Ambiente" (Spain). After imbibition in water with aeration for 48 hours, seeds were transferred to wet vermiculite and maintained at 23 °C under a 16-hour photoperiod. Germinated embryos were collected when the radicle length was inferior to 1 cm (G1), between 1-2 cm (G2) and 2-3 cm (G3). Furthermore, different tissues were excised from three-week-old seedlings in order to assess *PpKN* spatial pattern expression: root tip (5 mm of the apical part of the root), shoot apex (3 mm of the emerging epicotyl including the shoot apical meristem and needle primordia), the 5-mm portion of the hypocotyl situated right under the shoot apex, and cotyledons. Young needles from two-month-old seedlings and mature needles from adult trees growing in natural stands in Northern Spain were also collected. All tissues were snap frozen in liquid nitrogen and stored at -80 °C until use.

2.3.2 RNA extraction, cDNA synthesis and quantitative real time PCR (RT-qPCR)

RNA was isolated using the GeneMATRIX Universal RNA Purification Kit (EURx, Gdańsk, Poland), quantified by spectrophotometry and checked by agarose gel electrophoresis. For each sample, 1 µg of total RNA (0.5 µg for somatic embryogenesis samples) was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems Inc., Foster City, CA, USA) following the manufacturer's instructions.

Gene expression analysis was performed by RT-qPCR with a Bio-Rad CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Primers were designed with Primer3 software [35] following the recommended parameters [43] and are included in Table S3 (available as Supplementary Material). *P. pinaster ubiquitin* gene (Acc. AF461687) was used as endogenous reference gene [44,45,46]. Individual reactions were prepared in triplicate with 5 µl of iQ SYBR Green Supermix (Bio-Rad), oligonucleotide primers (0.20 µM each) and 100 ng of cDNA to a final volume of 10 µl. Amplification was carried out as follows: 95 °C 3 min; 45 cycles of 95 °C 10 s and 60 °C 30 s, with a final melting curve to discard the presence of non-specific products. Negative controls (no template) and RT- controls (non-retrotranscribed RNA) were also included.

Analysis of RT-qPCR data was performed using the qpcR package for R software (http://www.dr-spiess.de/qpcR.html) [47]. Relative abundance of each transcript was calculated as the mean of the three technical replicates and normalized to the mean expression value of the reference gene in each sample. Results were expressed as mean normalized expression values ± standard error of two biological replicates. Each biological replicate consists of a pool of several embryos or tissues excised from different plants. Statistical analysis was performed with t-test analysis or ANOVA using the Student-Newman-Keuls test for post hoc comparisons (SIGMA-PLOT v11 software, Chicago, IL, USA).

2.3.3 Fluorescent *in situ* hybridization (FISH)

Fluorescent *in situ* hybridization (FISH) was performed in order to determine *PpKN2*, *PpKN4* and *PpKN5* mRNA localization in *P. pinaster* shoot apexes excised from three-week-old seedlings obtained as mentioned in section 2.3.1.2. Tissues (less than 5 mm long) were immediately fixed with freshly prepared FAA solution (3.7% formaldehyde, 5% glacial acetic acid, 50% ethanol), incubated under vacuum overnight at 4 °C,

dehydrated in an ascendant ethanol series (50, 75, 90, and 100%) and embedded in Technovit® 8100 (Heraeus kulzer GmbH, Wehrheim, Germany) according to manufacturer's instructions.

Ten-µm longitudinal sections of embedded shoot tips were obtained using a microtome (Nikon, Tokyo, Japan) and mounted on Menzel-Gläser Superfrost Ultra Plus slides (Thermo-Scientific, Waltham, Massachusetts, USA). Air-dried sections were directly used for FISH [48]. Antisense probes (33-nucleotide single-stranded) labelled with Cyanine 5 (Cy5) in their 3' end were designed to hybridize in a specific region of each gene (see Table S4, available as Supplementary Material). Sections were observed and photographed under a Leica DMRXA fluorescence microscope (Leica Microsystems, Wetzlar, Germany) and images were processed using the ConfocalUniovi ImageJ software (http://spi03.sct.uniovi.es/confocaluniovi/).

2.4 Overexpression of *Pinus pinaster KNOX* genes in *A. thaliana*

In order to confirm functional conservation, we studied if the opposite effects of class I and class II *KNOX* genes overexpression on leaf morphology [24,25,26,27,28,29] is conserved in *P. pinaster* by ectopic expression of these genes under the control of the cauliflower mosaic virus (CaMV) 35S promoter in *A. thaliana.*

2.4.1 Vector construction

Overexpression vectors for all *KNOX* genes identified in *P. pinaster* (*PpKN1-6*) were constructed using the Gateway Technology (Invitrogen, Carlsbad, California, USA). Specific primers were designed to amplify the full-length coding sequence of each *P. pinaster KNOX* gene flanked by attB1 and attB2 sites by PCR. cDNA from *P. pinaster* shoot apexes was used as template. Primers used in this study are listed in Table S5 (available as Supplementary Material). attB-PCR products were purified using the NucleoSpin Extract II Kit (Macherey-Nagel, Germany). BP recombination reaction was

267 performed with PCR products containing attB sites, the donor Gateway pDONR221
268 vector and Clonase II (Invitrogen) using the Gateway Technology according to
269 manufacturer's instructions. One microliter of the cloning reaction mixture was used to
270 transform competent DH5α *Escherichia coli* cells using an Electro Cell Manipulator
271 600/630 following the electroporation protocol supplied by the manufacturer. After
272 selecting positive colonies, plasmids were isolated using the NucleoSpin Plasmid DNA
273 Purification Kit (Macherey-Nagel) and analyzed by restriction digestion, PCR and DNA
274 sequencing in order to confirm the presence of the corresponding *PpKN* gene in the so-
275 called entry clone. Subsequently, LR recombination reaction was performed to transfer
276 the corresponding *PpKN* gene into the destination vector pK7WG2 from the MultiSite
277 Gateway series (http://gateway.psb.ugent.be) [49], in which the gene of interest is under
278 the control of the CaMV 35S promoter. These constructions were used to study if the
279 effects of the overexpression of *P. pinaster KNOX* genes are similar to those described
280 for *KNOX* genes from *A. thaliana* and other angiosperm species. *E. coli* transformation,
281 plasmid purification and expression vector checking were carried out as described above.
282
283 2.4.2 Genetic transformation of *A. thaliana* plants mediated by *Agrobacterium*
284 *tumefaciens*
285 The expression vectors, designated p35S::*KN1* to p35S::*KN6*, were introduced by
286 electroporation into *Agrobacterium tumefaciens* strain AGL1 cells. *A. tumefaciens*-
287 mediated transformation of *A. thaliana* ecotype Columbia (Col-0) was performed using
288 the floral dip method [50]. T1 seeds were collected and cultured in Murashige and Skoog
289 medium (MS) containing 50 μg ml^{-1} kanamycin and 250 μg ml^{-1} cefotaxime to select
290 transgenic T1 plants, which were transferred into soil and maintained at 25 °C under a 16-
291 hour photoperiod at a photon flux of 150 ± 5 μmol m^{-2} s^{-1}, watering twice a week with tap
292 water, in order to analyze their phenotype. Overexpression of *PpKN* genes in T1 plants

was confirmed by PCR and/or RT-qPCR. RNA extraction, cDNA synthesis and RT-qPCR were performed as described in section 2.3. cDNA from non-transgenic *A. thaliana* Col-0 plants and cDNA from *P. pinaster* shoot apexes were used as negative and positive controls, respectively.

3. RESULTS

3.1 Identification of *KNOX* genes in *Pinus pinaster*

In order to characterize the complete *KNOX* gene family in *P. pinaster*, we designed specific primers for the isolation of the four class I *KNOX* genes previously described in conifers [30,31,32]. Based on the high degree of sequence conservation between these species, we obtained the orthologue mRNA sequences in *P. pinaster* for each gene, which were designated *PpKN1*, *PpKN2*, *PpKN3* and *PpKN4* according to the criteria followed in the above cited works. We also conducted a screening in the transcriptome and proteome databases ProCoGen, SustainPine and PLAZA Gymnosperms for the presence of new members of the *KNOX* gene family in this species. We did not find any additional class I members in *P. pinaster*, although two class II *KNOX* genes were identified and designated *PpKN5* and *PpKN6*. We also identified and described novel class II *KNOX* genes in the genome of other conifers, including two members in *P. taeda* (*PtKN5* and *PtKN6*), and one member in *P. abies* (*PaKN5*). No class M members were found in any of the analyzed conifer species.

The comparison between cDNA sequences and genomic data showed that class I *KNOX* genes in conifers contain five exons (Figure 1A). It was not possible to obtain the full genomic sequence by genome walking probably due to the presence of a very long third intron. The search in the ConGenIE and PLAZA Gymnosperms databases showed that *P. abies KN1* and *KN3* third introns are larger than 68 and 15 Kb, respectively. Class II

KNOX genes contain six exons in *P. pinaster* (Figure 1A). These gene structures are similar to those reported for their *A. thaliana* counterparts in The Arabidopsis Information Resource (https://www.arabidopsis.org/).

The analysis of the PpKN deduced protein sequences through InterProScan software showed that all of them contain the characteristic domains and motifs of KNOX proteins (Figure 1B). Both classes have differences outside the precisely conserved third helix of the homeodomain, as previously described in angiosperms [10]. It is remarkable the high degree of similarity between conserved domains in PpKN paralogs, especially in the ELK-HD region, although the degree of conservation is lower outside those regions. In particular, the similarity between class II PpKN proteins was 58.3% for the complete sequence and 92.3% for the HD, whereas class I proteins had an overall identity of 59.7% and a similarity of 85.3% for the HD. PpKN1 and PpKN2 are the class I KNOX proteins that showed the highest pairwise percentage of identity, being 79.5% for the complete amino acid sequence and 88.0% for the HD. However, it is remarkable that the highest value of similarity for the HD corresponded to the pair PpKN3 and PpKN4 (89.0%), although the conservation degree for the complete sequence was lower (57.7%). The similarity for the rest class I KNOX proteins ranged between 53.5% (for PpKN1 and PpKN4 alignment) and 61.1% (for PpKN2 and PpKN3 alignment) for the complete sequence, and 81.3% (for PpKN1 and PpKN3 alignment) and 86.3% (for PpKN2 and PpKN4 alignment) for the HD. PpKN proteins also shared a high degree of similarity with their orthologues from other conifer species (between 84.2% and 98.5% for KN6 and KN5 orthologues from *P. pinaster*, *P. abies* and *P. taeda*, respectively; Figure S1, available as Supplementary Material).

3.2 Phylogenetic analysis of *Pinus pinaster KNOX* gene family

To determine the phylogenetic position of the isolated *P. pinaster KNOX* genes, we constructed a phylogenetic tree with KNOX protein sequences from green algae, bryophytes, lycophytes, ferns, conifers and angiosperms, monocots and dicots (Figure 2). KNOX proteins from green algae were used as outgroup. The resultant consensus tree showed that *P. pinaster* class I and II members group together with their counterparts from other land plant groups, supporting the division of the *KNOX* gene family into two clades with a monophyletic origin, whereas green algae *KNOX* genes constituted an independent group. Among class I members, results showed that KN1 and KN2 orthologues from different conifer species grouped together, independently from KN3 and KN4 orthologues. Interestingly, AtSTM constituted the sister group of conifer KN3 orthologues. It is also noticeable that the evolution of the *KNOX* gene family in *P. pinaster* was under negative or purifying selection, as the Ka/Ks ratio obtained for each node of the phylogenetic tree was inferior to 1 in all cases (Figure S2, available as Supplementary Material).

3.3 Analysis of *Pinus pinaster KNOX* gene expression

To shed light on the specific roles of *P. pinaster KNOX* gene family members, gene expression was analyzed by RT-qPCR during somatic embryogenesis and embryo germination (Figure 3), and in different tissues from seedlings and adult trees (Figure 4). During somatic embryogenesis, *PpKN1* showed a reduction of its expression levels in early and late embryos compared to PEMs, although they increased again to reach its maximum in mature embryos. Its expression decreased considerably with the beginning of the germination, remaining relatively constant along the process. In seedlings, this gene was mainly expressed in hypocotyl, root apex and shoot apex, with low levels of expression in cotyledons and young needles, and it was undetectable in needles excised from adult trees (Figure 4).

PpKN2 showed the highest expression levels among the six *P. pinaster KNOX* genes during somatic embryogenesis and germination. Its expression increased progressively during embryogenesis and reached its peak in mature embryos. During germination, *PpKN2* expression profile was similar to that described for *PpKN1*. In seedlings, this gene was mainly expressed in hypocotyl, although it was also expressed in shoot, root tips and cotyledons at very low levels. No expression was detected in young and mature needles.

Conversely, the expression levels of *PpKN3* were the lowest of the *P. pinaster KNOX* genes in most analyzed stages and tissues. During embryogenesis, the highest transcript abundance was found in PEMs. Then, its levels decreased dramatically and were maintained at low levels throughout the transition from early to mature embryos, increasing again during germination. In seedlings, *PpKN3* was mainly expressed in hypocotyl and shoot apex, being slightly expressed in the root apex, cotyledons and mature needles. No expression was detected in young needles (Figure 4).

During embryogenesis and germination *PpKN4* expression pattern was very similar to that shown by *PpKN1*, except it was slightly downregulated in PEMs compared to LEs. *PpKN4* was predominantly expressed in the shoot apex of seedlings. It was also expressed in hypocotyl, and very little expression was detected in the rest of analyzed tissues.

Class II member *PpKN5* was expressed at low levels during the first stages of embryogenesis, increasing in mature embryos and germinating embryos. This gene was mainly expressed in young needles, with remarkably high expression in cotyledons, mature needles, shoot apex and root apex, and lower expression in hypocotyl. *PpKN6* transcript levels, however, remained relatively low and constant during embryogenesis and germination. Of note, *PpKN6* expression levels showed a significant increase in early embryos. In seedlings, this gene expression was mainly observed in mature needles, young needles, shoot apex and cotyledons. It also showed high levels of expression in hypocotyl and root apex.

The expression domain of several class I (*PpKN2* and *PpKN4*) and class II (*PpKN5*) *KNOX* genes was determined by FISH in shoot apexes excised from *P. pinaster* seedlings. Class I *KNOX* genes *PpKN2* and *PpKN4* expression was detected both in the central and peripheral zone of the SAM, and also in incipient needles (Figure 5A, B). By contrast, class II member *PpKN5* gene expression was restricted to needle primordia (Figure 5C).

3.4 Overexpression of *P. pinaster KNOX* genes in *A. thaliana*

To gain insight into their function, we obtained transgenic *A. thaliana* lines expressing *P. pinaster* class I and class II *KNOX* genes under the control of the CaMV 35S promoter (Figure 6). Two to six independent lines per each gene were analyzed and overexpression confirmed by PCR or RT-qPCR. Overexpression of class I *KNOX* genes *PpKN1*, *PpKN2*, *PpKN3* and *PpKN4* caused an altered phenotype consisting on various degrees of serration and lobing of *A. thaliana* rosette and cauline leaves in all lines analyzed, although no ectopic meristem formation was observed in any case. Some transformants overexpressing *PpKN1*, *PpKN2* or *PpKN4* showed a stronger altered phenotype, which, in addition, showed very small lobed leaves, short shoots and infertility. These strong phenotypes usually were correlated with a high overexpression (Figure S3). As expected, overexpression of class II *KNOX* genes did not alter leaf morphology and no other phenotypic defects were observed in transgenic plants (Figure 6).

4. DISCUSSION

4.1 The *P. pinaster KNOX* gene family: diversity and phylogenetic analysis

Previous studies reported the existence of four class I *KNOX* genes in various conifer species [30,31,32]. Consistent with these reports, we also identified four class I *KNOX* genes in *P. pinaster*. The fact that the number of *KNOX* class I genes remains constant in

several conifer species might indicate that diversification of this subfamily took place before the diversification of the Pinaceae lineage. However, as *P. pinaster* genome has not yet been released and transcriptome databases could be incomplete, we cannot discard the presence of additional class I *KNOX* genes expressed at very low levels, under very specific conditions or restricted to small groups of cells. Two members of the class II subfamily were identified in *P. pinaster* following a similar procedure. A subsequent analysis of *P. abies* and *P. taeda* transcriptome databases showed that class II *KNOX* genes are also present in these conifer species, in concordance with previous studies [23]. In this case, whereas two class II members were identified in *P. pinaster* and *P. taeda*, only one member was found in *P. abies*, which might be due to an independent diversification of this gene subfamily in each genus of the Pinaceae or to the lack of data in the available transcriptome databases. With few exceptions, class II KNOX protein sequences are more conserved compared to class I proteins. It has been suggested that this might indicate that class II *KNOX* genes had been under stronger purifying selection [5]. The fact that gymnosperms contain both classes of *KNOX* genes reinforces the conservation of this gene family throughout land plants, since members from both classes of *KNOX* genes have been described from bryophytes to angiosperms [5]. Furthermore, class I and class II *KNOX* genes showed different exon-intron structure and both classes had differences outside the precisely conserved third helix of the homeodomain, as previously described in angiosperms [10]. No class M members were identified in conifers, possibly indicating that these genes are exclusive to some eudicot species.

As expected based on their sequence similarity, each *P. pinaster* class I and class II member grouped closely with their orthologues from other conifer species, which indicates a high conservation of this gene family across the Pinaceae lineage. Furthermore, the analysis of the substitution rates suggests that the *P. pinaster KNOX* gene family was

under a negative or purifying selection in order to remove those polymorphisms that could lead to a change of the gene function, which might be applicable to other conifer species. Class I genes grouped together with members of this subfamily from different angiosperm species. Previous phylogenetic analyses using both the complete amino acid sequences and highly-conserved regions established that conifer class I genes constitute a monophyletic group, which is the sister group of AtSTM among others, and its members probably originated as the result of three successive duplication events: the first duplication event probably led to the appearance of *KN4* genes, the second one may have caused the emergence of *KN3* genes, and the last one would have given rise to *KN1* and *KN2* genes [32]. Our data, in concordance with those reported by [23], are not consistent with the monophyly of the conifer class I sequences [32], which might be due to the fact that our study includes sequences from both class I and II KNOX proteins from multiple plant groups, offering a more general perspective of gene family evolution.

Our results reinforce the proximity between KN1 and KN2 orthologues, which probably arose after a duplication event [23,32]. KN3 orthologues group together with AtSTM, which might indicate that both originated from an ancestor gene that was already present in the common ancestor of seed plants. No specific orthologues for other *A. thaliana* members were found in conifers. It could be due to the lack of information in the available databases or to their absence in the gymnosperm lineage as the result of an independent evolution and diversification of *KNOX* subfamilies in each seed plant group. Some angiosperm species like *Fragaria vesca* only contain two class I *KNOX* genes, whereas others have gained a great number of new paralogs, as 11 members have been identified in *Glycine max* [5]. In addition, recent studies suggest that gene expansion by duplication is not the only driving force of evolution. The genome analysis of 20 flowering plants showed a set of genes that were restored to single-copy status [51]. Therefore, a loss of *KNOX* genes in the gymnosperm lineage cannot be excluded.

4.2 Functional analysis of *P. pinaster KNOX* genes

Diverse roles have been described for *KNOX* genes in different species. Interestingly, the function of class I and class II members has diverged throughout plant evolution [52]. In the green alga *Chlamydomonas reinhardtii*, the BELL protein Gamete-specific *plus1* (Gsp1) and the KNOX protein Gamete-specific *minus1* (Gsm1) provided by gametes of plus and minus mating types, respectively, form a heterodimer that initiates the zygotic developmental program [53]. In mosses, both class I and II *KNOX* genes are mainly expressed during the diploid phase of their life cycle (sporophyte), but whereas class I *KNOX* genes stimulate cell proliferation during sporophyte development, class II *KNOX* genes regulate the haploid to diploid developmental transition and have been involved in the repression of the diploid program in gametophyte [54,55,56]. In angiosperms, class I and II *KNOX* genes play opposing roles in plant growth and development. While class I members participate in the maintenance of meristematic regions, class II members are involved in tissue differentiation [23].

The high degree of sequence conservation and phylogenetic proximity to *KNOX* genes from angiosperms could suggest a conservation of *KNOX* gene function across seed plants. *PpKN* expression patterns support this hypothesis, despite some differences with those from angiosperms have been described. Similar to *A. thaliana*, *P. pinaster* class I *KNOX* gene expression was very low in lateral organs like young and mature needles. However, whereas its expression is mainly restricted to meristematic regions in angiosperms, *P. pinaster* class I *KNOX* genes showed a broader expression pattern. Transcripts for all *KNOX* genes were found mainly in the hypocotyl and shoot apex, but also in the root apex and at low levels in cotyledons, being even expressed in some cases in young and/or mature needles. Class I *KNOX* genes isolated in *P. abies* were also expressed in other tissues such as stems, roots, and female and male cone buds, but not in

lateral organs [31]. The first *KNOX* gene identified in conifers (*HBK1*, here designated *PaKN2* for convenience) was isolated from a female strobilus library, and was shown to be expressed in the undifferentiated cell population situated in the center of the vegetative meristem, but not in needle primordia, suggesting a role in both vegetative and reproductive meristems [30]. In conifers, the SAM contains a population of stem cells in the central zone (CZ) and a surrounding peripheral zone (PZ) where cell differentiation takes place, as it was described for angiosperms. However, whereas three clonally distinct layers can be differentiated in the SAM of *A. thaliana* (tunica-corpus model), in conifers some cells from the superficial layer of the CZ, which are called apical initials, not only divide anticlinally but also periclinally, contributing to the internal cell lineages [57]. In our case, FISH analyses of class I *KNOX* genes *PpKN2* and *PpKN4* in the shoot apex showed that these genes had a wider expression domain to that described for their *A. thaliana* counterparts, as they were expressed both in the meristem and in incipient needles. The fact that *P. pinaster* class I *KNOX* genes are expressed in the SAM might indicate an important role in the maintaining of meristematic potential. However, their expression in needle primordia might indicate additional roles in needle development, although more studies should be done in order to determine their putative participation in this process.

In angiosperms, each class I *KNOX* gene shows a well-defined expression domain in the SAM that can overlap with the corresponding expression domain from other genes of this subfamily [8,10]. It has also been reported the expression of some class I *KNOX* genes outside the SAM. For example, *AtBP/KNAT1* is also expressed below the SAM and in the cells surrounding vascular elements of the stem, and *AtKNAT6* expression is detected in different tissues such as the shoot apex and root [8]. The fact that *P. pinaster* class I *KNOX* transcripts were found in the hypocotyl may indicate that these genes are expressed in cells associated with the vascular system, similarly to what have been observed for

AtBP/KNAT1, although more studies should be done in order to verify this hypothesis. In fact, RNA sequencing (RNA-Seq) transcriptome analysis of different tissues from *P. pinaster* one-month-old seedlings excised by laser capture microdissection (LCM) showed that *PpKN3* was mainly expressed in the SAM and in vascular tissues from hypocotyl, roots, cotyledons and young needles (exImage Microdisecction Atlas: http://v22.popgenie.org/microdisection/) [36]. Of note, *PpKN* expression was detected in cotyledons and some *PpKN* genes were also expressed in young and/or mature leaves, although at very low levels. This seems contradictory with the absence of class I *KNOX* gene expression from incipient and developing cotyledons during embryogenesis in *A. thaliana* [58], as well as from lateral organs in angiosperms and *P. abies* [31]. This could be explained as the result of *PpKN* gene expression in small groups of cells in cotyledons and needles, similarly to what has been observed for *ZmRS1* and *ZmKNOX3* at the base of maize leaves [59].

Several studies in *A. thaliana* showed that class I *KNOX* genes are firstly expressed during embryogenesis, contributing to the formation and establishment of different domains and boundaries in the SAM [60]. *P. pinaster* class I *KNOX* gene expression was analyzed during somatic embryogenesis, a good experimental system for the study of the underlying physiological and molecular mechanisms of embryogenesis in conifers [61,62]. Expression patterns were very similar for all class I *KNOX* genes, except for *PpKN3*. Interestingly, studies in *P. abies* showed that *KN3* and *KN4* orthologues are only expressed in embryogenic cell lines competent to form fully mature embryos, suggesting that these genes might be necessary for the correct SAM establishment and maintenance, whereas *KN1* and *KN2* orthologues are expressed with no significant differences in all the embryogenic lines analyzed, suggesting a more general role in embryo development [31,34]. Furthermore, overexpression of *PaKN1* in embryogenic cultures accelerated the differentiation process and increased the number of mature cotyledonary embryos

obtained, which had bigger SAMs in comparison to non-transgenic somatic embryos, whereas its down-regulation caused an arrest of the maturation process [33].

In *A. thaliana*, it has been reported that class I *KNOX* members have discrete and overlapping functions [20,22]. All class I *KNOX* genes isolated from *P. pinaster* were expressed in all analyzed tissues and their expression domain in the shoot apex was very similar. The overexpression of *P. pinaster* class I members in *A. thaliana* caused similar phenotypic effects to what has been reported for overexpression of class I members from angiosperms and other conifers in *A. thaliana* [13,34,63], altering the normal leaf development and giving rise to lobulated and serrated leaves. No ectopic meristem formation on leaves was observed in any case, as it has been described for some *A. thaliana KNOX* genes [9,13]. Phenotypic effects of *PpKN3* overexpression were less severe than those observed for the rest of *PpKN* class I genes. Results obtained in *P. abies* were very similar, since the ectopic expression of these genes under the control of CaMV 35S promoter in *A. thaliana* plants also caused an alteration of the shape of the rosette leaves and, additionally, overexpression of *PaKN1* and *PaKN2* also altered the flower morphology [30,34]. Altogether, these data might indicate that *P. pinaster* class I *KNOX* genes might have redundant roles in plant development, although more studies are needed to determine whether there is a functional divergence of the class I *KNOX* subfamily members. The high degree of sequence similarity, their phylogenetic proximity and similar expression patterns, together with data available from other species, might indicate little functional differentiation for *PpKN1* and *PpKN2*, whereas *PpKN3* seems to play a differential role than the rest of class I members.

P. pinaster class II *KNOX* genes were mainly expressed in cotyledons and young and mature needles, similar to the results obtained in *A. thaliana* [23]. They were also expressed in all other analyzed tissues. In addition, their expression in shoot apexes was restricted to incipient and developing needles, and no expression was detected in the

meristematic region by FISH. Furthermore, as expected, no apparent phenotypic defects were observed as consequence of the overexpression of *P. pinaster* class II *KNOX* genes in *A. thaliana*. It has been previously reported that the gain of class II *KNOX* function causes leaf simplification in the angiosperm *Cardamine hirsuta*, a species closely related to *A. thaliana* that has dissected leaves, whereas loss-of-function mutations in *A. thaliana* cause an increase of the leaf complexity [23,29]. Altogether, these data suggest that these genes could be involved in organ differentiation, similarly to what has been described in angiosperms.

5. CONCLUSIONS

Our results prove the presence and expression of both class I and class II *KNOX* genes in *P. pinaster* and other conifer species, indicating a high conservation of this gene family throughout land plants. Furthermore, it seems that the established functional differentiation of class I and II subfamilies in angiosperms may be partially conserved in gymnosperms. *P. pinaster* class I *KNOX* genes were expressed in meristematic regions and they may participate in the maintenance of undifferentiated state of those cells similarly to what has been described in angiosperms. However, class I *KNOX* gene expression was not limited to meristematic regions, so these genes could perform additional roles to those described in angiosperms. *P. pinaster* class II *KNOX* genes were mainly expressed in developing and mature organs, and therefore they could be involved in promoting tissue maturation. This study provides new insights into the *KNOX* gene family diversity and evolution in gymnosperms and will facilitate molecular studies to characterize the role of their members in stem cell maintenance and tissue differentiation in this group of seed plants.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest. All authors revised and approved the final manuscript.

FUNDING

This work was supported by "Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria" (INIA) and "Fondo Europeo de Desarrollo Regional" (FEDER) cofunding (RTA2013-00048-C03-02 and RTA2017-00063-C04-04), and University of Oviedo (PAPI-17-PEMERG-3). Natalia Bueno was funded by a predoctoral grant from "Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología" (FICYT) (BP10-098).

ACKNOWLEDGEMENTS

We would like to thank Dr. Rafael Cañas from University of Málaga for the helpful advice provided in gene identification and Dr. Silvia Alvarez-Diaz from The Walter and Eliza Hall Institute of Medical Research (Australia) for the helpful comments and kindly assisting with language editing and proofreading.

AUTHOR CONTRIBUTION

Natalia Bueno: Investigation, Formal analysis, Writing- Original draft preparation. **José M. Alvarez**: Conceptualization, Methodology, Writing- Reviewing and Editing. **Ricardo J. Ordás**: Supervision, Resources, Writing- Reviewing and Editing.

REFERENCES

[1] T. Bürglin, Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals, Nucleic Acids Res. 25 (1997) 4173-4180.

[2] K. Mukherjee, T.R. Bürglin, Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution, J. Mol. Evol. 65(2) (2007) 137-153.

[3] K. Mukherjee, L. Brocchieri, T.R. Bürglin, A comprehensive classification and evolutionary analysis of plant homeobox genes, Mol. Biol. Evol. 26(12) (2009) 2775-2794.

[4] E. Vollbrecht, B. Veit, N. Sinha, S. Hake, The developmental gene *Knotted-1* is a member of a maize homeobox gene family, Nature 350 (1991) 241-243.

[5] J. Gao, X. Yang, W. Zhao, T. Lang, T. Samuelsson, Evolution, diversification, and expression of KNOX proteins in plants, Front. Plant Sci. 6(882) (2015).

[6] T. Sakamoto, A. Nishimura, M. Tamaoki, M. Kuba, H. Tanaka, S. Iwahori, M. Matsuoka, The conserved *KNOX* domain mediates specificity of tobacco KNOTTED1-type homeodomain proteins, Plant Cell 11 (1999) 1419-1432.

[7] H. Nagasaki, T. Sakamoto, Y. Sato, M. Matsuoka, Functional analysis of the conserved domains of a rice KNOX homeodomain protein, OSH15, Plant Cell 13 (2001) 2085-2098.

[8] S. Scofield, J.A.H. Murray, *KNOX* gene function in plant stem cell niches, Plant Mol. Biol. 60 (2006) 929-946.

[9] S. Hake, H. Smith, H. Holtan, E. Magnani, G. Mele, J. Ramirez, The role of *KNOX* genes in plant development, Annu. Rev. Cell Dev. Bi. 20 (2004) 125-151.

[10] R. Kerstetter, E. Vollbrecht, B. Lowe, B. Veit, J. Yamaguchi, S. Hake, Sequence analysis and expression patterns divide the maize *knotted1-like* homeobox genes into two classes, Plant Cell 6 (1994) 1877-1887.

[11] E. Magnani, S. Hake, *KNOX* lost the *OX*: The Arabidopsis *KNATM* gene defines a novel class of KNOX transcriptional regulators missing the homeodomain, Plant Cell 20 (2008) 875-887.

[12] S.E. Clark, S.E. Jacobsen, J.Z. Levin, E.M. Meyerowitz, The *CLAVATA* and *SHOOT MERISTEMLESS* loci competitively regulate meristem activity in *Arabidopsis*, Development 122 (1996) 1567-1575.

[13] G. Chuck, C. Lincoln, S. Hake, *KNAT1* induces lobed leaves with ectopic meristems when overexpressed in Arabidopsis, Plant Cell 8 (1996) 1277-1289.

[14] K. Endrizzi, B. Moussian, A. Haecker, J.Z. Levin, T. Laux, The *shoot meristemless* gene is required for maintenance of undifferentiated cells in Arabidopsis shoot and floral meristems and acts at a different regulatory level than the meristem genes *WUSCHEL* and *ZWILLE*, Plant J. 10 (1996) 967-979.

[15] J.A. Long, E.I. Maon, J.I. Medford, M.K. Barton, A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*, Nature 379 (1996) 66-69.

[16] H.M. Rupp, M. Frank, T. Werner, M. Strnad, T. Schmülling, Increased steady state mRNA levels of the *STM* and *KNAT1* homeobox genes in cytokinin overproducing *Arabidopsis thaliana* indicate a role for cytokinins in the shoot apical meristem, Plant J. 18 (1999) 557-563.

[17] S. Jasinski, P. Piazza, J. Craft, A. Hay, L. Woolley, I. Rieu, A. Phillips, P. Hedden, M. Tsiantis, KNOX action in *Arabidopsis* is mediated by coordinate regulation of cytokinin and gibberellin activities, Curr. Biol. 15 (2005) 1560-1565.

679 [18] O. Yanai, E. Shani, K. Dolezal, P. Tarkowski, R. Sablowski, G. Sandberg, A.
680 Samach, N. Ori, *Arabidopsis* KNOXI proteins activate cytokinin biosynthesis. Curr. Biol.
681 15 (2005) 1566-1571.
682 [19] S. Scofield, W. Dewitte, J.A.H. Murray, The *KNOX* gene *SHOOT MERISTEMLESS*
683 is required for the development of reproductive meristematic tissues in Arabidopsis, Plant
684 J. 50 (2007) 767-781.
685 [20] S. Scofield, W. Dewitte, J.A.H. Murray, A model for Arabidopsis class-1 *KNOX* gene
686 function. Plant Signal. Behav. 3(4) (2008) 257-259.
687 [21] S. Scofield, W. Dewitte, J.A.H. Murray, STM sustains stem cell function in the
688 *Arabidopsis* shoot apical meristem and controls *KNOX* gene expression independently of
689 the transcriptional repressor AS1, Plant Signal. Behav. 9 (2014) e28934.
690 [22] E. Belles-Boix, O. Hamant, S.M. Witiak, H. Morin, J. Traas, V. Pautot, *KNAT6*: an
691 *Arabidopsis* homeobox gene involved in meristem activity and organ separation, Plant
692 Cell 18 (2006) 1900-1907.
693 [23] C. Furumizu, J.P. Alvarez, K. Sakakibara, J.L. Bowman, Antagonistic roles for
694 *KNOX1* and *KNOX2* genes in patterning the land plant body plan following an ancient
695 gene duplication, PLOS Genet. 11 (2015) e1004980.
696 [24] D. Hareven, T. Gutfinger, A. Parnis, Y. Eshed, E. Lifschitz, The making of a
697 compound leaf: genetic manipulation of leaf architecture in tomato, Cell 84 (1996) 735-
698 744.
699 [25] G. Bharathan, T.E. Goliber, C. Moore, S. Kessler, T. Pham, N.R. Sinha, Homologies
700 in leaf form inferred from *KNOXI* gene expression during development, Science 296
701 (2002) 1858-1860.
702 [26] A. Hay, M. Tsiantis, The genetic basis for differences in leaf form between
703 *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*, Nat. Genet. 38(8) (2006)
704 942-947.

[27] L.A. Nikolov, M. Tsiantis, Interspecies gene transfer as a method for understanding the genetic basis for evolutionary change: progress, pitfalls, and prospects, Front. Plant Sci. 6(1135) (2015) 1135.

[28] M.I. Rast-Somssich, S. Broholm, H. Jenkins, C. Canales, D. Vlad, M. Kwantes, G. Bilsborough, R. Dello Ioio, R.M. Ewing, P. Laufs, P. Huijser, C. Ohno, M.G. Heisler, A. Hay, M. Tsiantis, Alternate wiring of a *KNOXI* genetic network underlies differences in leaf development of *A. thaliana* and *C. hirsuta*, Genes Dev. 29(22) (2015) 2391-2404.

[29] M. Das Gupta, M. Tsiantis, Gene networks and the evolution of plant morphology, Curr. Opin. Plant Biol. 45 (2018) 82-87.

[30] A. Sundås-Larsson, M. Svenson, H. Liao, P. Engström, A homeobox gene with potential developmental control function in the meristem of the conifer *Picea abies*, PNAS 95 (1998) 15118-15122.

[31] H.I. Hjortswang, A. Sundås-Larsson, G. Bharathan, P.V. Bozhkov, S. von Arnold, T. Vahala, *KNOTTED1*-like homeobox genes of a gymnosperm, Norway spruce, expressed during somatic embryogenesis, Plant Physiol. Bioch. 40 (2002) 837-843.

[32] C. Guillet-Claude, N. Isabel, B. Pelgas, J. Bousquet, The evolutionary implications of *knox-I* gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence, Mol. Biol. Evol. 21 (2004) 2233-2245.

[33] M. Belmonte, M. Tahir, D. Schroeder, C. Stasolla, Overexpression of *HBK3*, a class I *KNOX* homeobox gene, improves the development of Norway spruce (*Picea abies*) somatic embryos, J. Exp. Bot. 58 (2007) 2851-2861.

[34] E. Larsson, F. Sitbon, S. von Arnold, Differential regulation of *Knotted1-like* genes during establishment of the shoot apical meristem in Norway spruce (*Picea abies*), Plant Cell Rep. 31 (2012) 1053-1060.

[35] S. Rozen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, Methods Mol. Biol. 132(3) (2000) 365-386.

731 [36] R.A. Cañas, Z. Li, M.B. Pascual, V. Castro-Rodríguez, C. Ávila, L. Sterck, Y. Van
732 de Peer, F.M. Cánovas, The gene expression landscape of pine seedling tissues, Plant J.
733 91(6) (2017) 1064-1087.
734 [37] J. Canales, R. Bautista, P. Label, J. Gómez-Maldonado, I. Lesur, N. Fernández-Pozo,
735 M. Rueda-López, D. Guerrero-Fernández, V. Castro-Rodríguez, H. Benzekri, R.A. Cañas,
736 M-A. Guevara, A. Rodrigues, P. Seoane, C. Teyssier, A. Morel, F. Ehrenmann, G. Le
737 Provost, C. Lalanne, C. Noirot, C. Klopp, I. Reymond, A. García-Gutiérrez, J-F. Trontin,
738 M-A. Lelu-Walter, C. Miguel, M.T. Cervera, F.R. Cantón, C. Plomion, L. Harvengt, C.
739 Avila, M.G. Claros, F.M. Cánovas, *De novo* assembly of maritime pine transcriptome:
740 implications for forest breeding and biotechnology, Plant Biotechnol. J. 12(3) (2014) 286-
741 299.
742 [38] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J.
743 Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search
744 programs, Nucleic Acids Res. 25(17) (1997) 3389-3402.
745 [39] J.M. Alvarez, M. Cortizo, R.J. Ordás, Cryopreservation of somatic embryogenic
746 cultures of *Pinus pinaster*: effects on regrowth and embryo maturation, Cryo-Lett. 33(6)
747 (2012) 476-484.
748 [40] J.M. Alvarez, N. Bueno, M. Cortizo, R.J. Ordás, Improving plantlet yield in *Pinus*
749 *pinaster* somatic embryogenesis, Scand. J. Forest Res. 28(7) (2013) 613-620.
750 [41] I. Hakman, S. von Arnold, Somatic embryogenesis and plant regeneration from
751 suspension cultures of *Picea glauca* (white spruce), Physiol. Plantarum 72(3) (1988) 579-
752 587.
753 [42] S. Tereso, K. Zoglauer, A. Milhinhos, C. Miguel, M.M. Oliveira, Zygotic and somatic
754 embryo morphogenesis in *Pinus pinaster*: comparative histological and histochemical
755 study, Tree Physiol. 27 (2007) 661-669.

[43] M.K. Udvardi, T. Czechowski, W.R. Scheible, Eleven golden rules of quantitative RT-PCR, Plant Cell 20(7) (2008) 1736-1737.

[44] P. Alonso, M. Cortizo, F.R. Cantón, B. Fernández, A. Rodríguez, M.L. Centeno, F.M. Cánovas, R.J. Ordás, Identification of genes differentially expressed during adventitious shoot induction in *Pinus pinea* cotyledons by subtractive hybridization and quantitative PCR, Tree Physiol. 27(12) (2007) 1721-1730.

[45] J.M. Alvarez, M. Cortizo, R.J. Ordás, Characterization of a type-A response regulator differentially expressed during adventitious caulogenesis in *Pinus pinaster*, J. Plant Physiol. 169(18) (2012) 1807-1814.

[46] J.M. Alvarez, M. Cortizo, N. Bueno, A. Rodríguez, R.J. Ordás, *CLAVATA1-LIKE*, a leucine-rich-repeat protein receptor kinase gene differentially expressed during adventitious caulogenesis in *Pinus pinaster* and *Pinus pinea*, Plant Cell Tiss. Org. Cult.112(3) (2013) 331-342.

[47] C. Ritz, A-N. Spiess, qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis, Bioinformatics 24(13) (2008) 1549-1551.

[48] J. Alvarez, N. Bueno, R.A. Cañas, C. Avila, F.M. Cánovas, R.J. Ordás, Analysis of the *WUSCHEL-RELATED HOMEOBOX* gene family in *Pinus pinaster*: New insights into the gene family evolution, Plant Physiol. Bioch. 123 (2018) 304-318.

[49] M. Karimi, D. Inzé, A. Depicker, GATEWAY^{TM} vectors for *Agrobacterium*-mediated plant transformation, Trends Plant Sci. 7(5) (2002) 193-195.

[50] X. Zhang, R. Henriques, S-S. Lin, Q-W Niu, N-H. Chua, *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method, Nat. Protoc. 1(2) (2006) 1-6.

780 [51] R. De Smet, K.L. Adams, K. Vandepoele, M. Van Montagua, S. Maere, Y. Van de
781 Peer, Convergent gene loss following gene and genome duplications creates single-copy
782 families in flowering plants, PNAS 110(8) (2013) 2898-2903.
783 [52] K. Tsuda, S. Hake, Diverse functions of KNOX transcription factors in the diploid
784 body plan of plants, Curr. Opin. Plant Biol. 27 (2015) 91-96.
785 [53] J.H. Lee, H. Lin, S. Joo, U. Goodenough, Early sexual origins of homeoprotein
786 heterodimerization and evolution of the plant KNOX/BELL family, Cell 133 (2008) 829-
787 840.
788 [54] K. Sakakibara, T. Nishiyama, H. Deguchi, M. Hasebe, Class 1 *KNOX* genes are not
789 involved in shoot development in the moss *Physcomitrella patens* but do function in
790 sporophyte development, Evol. Dev. 10 (2008) 555-566.
791 [55] K. Sakakibara, S. Ando, H.K. Yip, Y. Tamada, Y. Hiwatashi, T. Murata, H. Deguchi,
792 M. Hasebe, J.L. Bowman, *KNOX2* genes regulate the haploid-to-diploid morphological
793 transition in land plants, Science 339 (2013) 1067-1070.
794 [56] E. Frangedakis, D. Saint-Marcoux, L.A. Moody, E. Rabbinowitsch, J.A. Langdale,
795 Nonreciprocal complementation of *KNOX* gene function in land plants, New Phytol. 216
796 (2017) 591-604.
797 [57] S. Conway, A. Drinnan, Analysis of surface growth in the conifer shoot apical
798 meristem, Int. J. Plant Sci. 178 (2017) 273-287.
799 [58] J.A. Long, M.K. Barton, The development of apical embryonic pattern in
800 *Arabidopsis*, Development 125 (1998) 3027-3035.
801 [59] D. Jackson, B. Veit, S. Hake, Expression of maize *knotted1*-related homeobox genes
802 in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot,
803 Development 129 (1994) 405-413.
804 [60] A. Hay, M. Tsiantis, *KNOX* genes: versatile regulators of plant development and
805 diversity, Development 137 (2010) 3153-3165.

806 [61] L.H. Filonova, P.V. Bozhkov, S. von Arnold, Developmental pathway of somatic
807 embryogenesis in *Picea abies* as revealed by time-lapse tracking, J. Exp. Bot. 51(343)
808 (2000) 249-264.
809 [62] S. von Arnold, I. Sabala, P. Bozhkov, J. Dyachok, L. Filonova, Developmental
810 pathways of somatic embryogenesis, Plant Cell Tiss. Org. Cult. 69(3) (2002) 233-249.
811 [63] A. Hay, D. Jackson, N. Ori, S. Hake, Analysis of the competence to respond to
812 KNOTTED1 activity in Arabidopsis leaves using a steroid induction system, Plant
813 Physiol. 131 (2003) 1671-1680.

SUPPLEMENTARY MATERIAL

Supplementary Material for this article is available online:

Figure S1. Conifer KNOX class II amino acid sequences alignment.

Figure S2. Phylogenetic tree of the *Pinus pinaster KNOX* gene family generated by the parsimony method with the values of non-synonymous (Ka) and synonymous (Ks) substitution rates for each node, which were calculated through the Computational Biology Unit (CBU) Ka/Ks Calculation tool (http://services.cbu.uib.no/tools/kaks).

Figure S3. Phenotypic effects of *PpKN4* overexpression in *Arabidopsis thaliana*. Transformants showing strong altered phenotype comprising very small lobed leaves, short shoots and infertility (#1 and #2) and transformants just showing lobed leaves (#4 and #7). Different phenotype is correlated with *PpKN4* overexpression level.

Table S1. List of primers used for the isolation of class I *KNOX* genes in *Pinus pinaster*.

Table S2. KNOX protein sequences used for the phylogenetic analyses and their GenBank Accession numbers.

Table S3. Primers used for the expression analysis of *KNOX* genes by quantitative real time PCR (RT-qPCR).

Table S4. Probes used in the fluorescent *in situ* hybridization (FISH).

Table S5. Primers used for 35S::*PpKN* vector construction.

834 **LIST OF FIGURES**

835

836 **Figure 1. Identification of *KNOX* gene family in *Pinus pinaster*.** (A) Full-length cDNAs
837 of *P. pinaster KNOX* genes *PpKN2* (class I) and *PpKN6* (class II) showing UTRs (white)
838 and exons (grey). (B) PpKN deduced proteins showing the position of the conserved
839 domains obtained through InterProScan. MEINOX (KNOX1 and KNOX2) domain
840 (blue), ELK domain (purple) and homeodomain (green).

841 **Figure 2. Phylogenetic analysis of the *KNOX* gene family.** Consensus tree of the KNOX
842 gene family constructed from 81 amino acid sequences from green algae and all major
843 groups of land plants. The tree was obtained by the Neighbour-Joining method and the
844 Jukes-Cantor genetic distance model using 100 bootstrap replicates through the Geneious
845 software. Green alga OtKNOX sequence was used as outgroup for the tree. New
846 sequences identified in this work are marked with an asterisk. Green algae KNOX
847 sequences are shown in black, whereas class I and class II KNOX sequences are shown
848 in red and blue, respectively. Aa: *Acetabularia acetabulum*; Amt: *Amborella trichopoda*;
849 At: *Arabidopsis thaliana*; Chr: *Chlamydomonas reinhardti*; Cr: *Ceratopteris richardii*;
850 Mc: *Micromonas* sp.; Os: *Oryza sativa*; Ot: *Ostreococcus tauri*; Pa: *Picea abies*; Pg: *Picea*
851 *glauca*; Ph: *Physcomitrella patens*; Pm: *Picea mariana*; Pp: *Pinus pinaster*; Ps: *Pinus*
852 *strobus*; Pt: *Pinus taeda*; Sm: *Selaginella moellendorffii*; Zm: *Zea mays*.

853 **Figure 3. Expression profiles of *Pinus pinaster KNOX* genes (*PpKN1* to *PpKN6*)**
854 **during different stages of somatic embryogenesis and embryo germination obtained**
855 **by quantitative real-time PCR (RT-qPCR).** Results are expressed as mean values of
856 relative expression ± Standard Error from two biological replicates. PEM: Proembryogenic
857 Masses; EE: Early Embryo; LE: Late Embryo; ME: Mature embryo; G1: germinated
858 embryo with a radicle < 1 cm; G2: germinated embryo with radicle between 1-2 cm; G3:
859 germinated embryo with radicle between 2-3 cm. Different letters indicate significant

differences in the relative mRNA abundance (Student–Newman–Keuls test, $\alpha = 0.05$). Bar, 1 mm (PEM; EE; LE; ME); 1 cm (G1-G3).

Figure 4. Expression levels of *Pinus pinaster KNOX* genes (*PpKN1* to *PpKN6*) in different tissues of seedlings and adult trees by quantitative real-time PCR (RT-qPCR). Results are expressed as mean values of relative expression ± Standard Error from two biological replicates. Rt: Root tip; Hy: Hypocotyl; Sa: Shoot Apex; Co: Cotyledons, YN: Young Needles; MN: Mature Needles. Different letters indicate significant differences in the relative mRNA abundance (Student–Newman–Keuls test, $\alpha = 0.05$).

Figure 5. Localization of *KNOX* mRNAs by fluorescent *in situ* hybridization (FISH) in longitudinal sections of shoot apexes excised from three-week-old *Pinus pinaster* seedlings. Specific anti-mRNA probes labelled with Cyanine 5 (red signal) were used for the detection of *PpKN2* (A), *PpKN4* (B) and *PpKN5* (C) transcripts. Blue signal represents nuclei staining with 4,6-diamidino-2-phenylindole (DAPI). cz: central zone of the meristem; pz: peripheral zone of the meristem; np: needle primordia. Bar, 50 µm.

Figure 6. Phenotypic effects of *Pinus pinaster KNOX* gene overexpression in *Arabidopsis thaliana*. (A) Schematic representation of the constructions used in the experiment, in which *P. pinaster KNOX* genes were under the control of the cauliflower mosaic virus 35S promoter. (B) *A. thaliana* ecotype Columbia-0 (Col-0) wild type. (C-K) Overexpression of class I members in *A. thaliana* caused various degrees of serration and lobing of rosette and cauline leaves, and in some cases altered growth and infertility. (C-D) Transgenic *A. thaliana* T1 lines overexpressing *PpKN1*. (E) Leaf morphology comparison between Col-0 and 35S::*PpKN1* T1 plants. (F-G) Transgenic *A. thaliana* T1 lines overexpressing *PpKN2*. (H) Overexpression of *PpKN3* caused light serration of leaf margins compared to other class I members. (I-K) Transgenic *A. thaliana* T1 lines overexpressing *PpKN4*. In some cases, leaf lobes (arrows) were already visible in four-

886 leaf plantlets (K). (L) Ectopic expression of the class II members, *PpKN5* and *PpKN6*,

887 has no phenotypic effects. Bar, 1 cm (B-F, H-J, L), 1 mm (G and K).

Figure 1

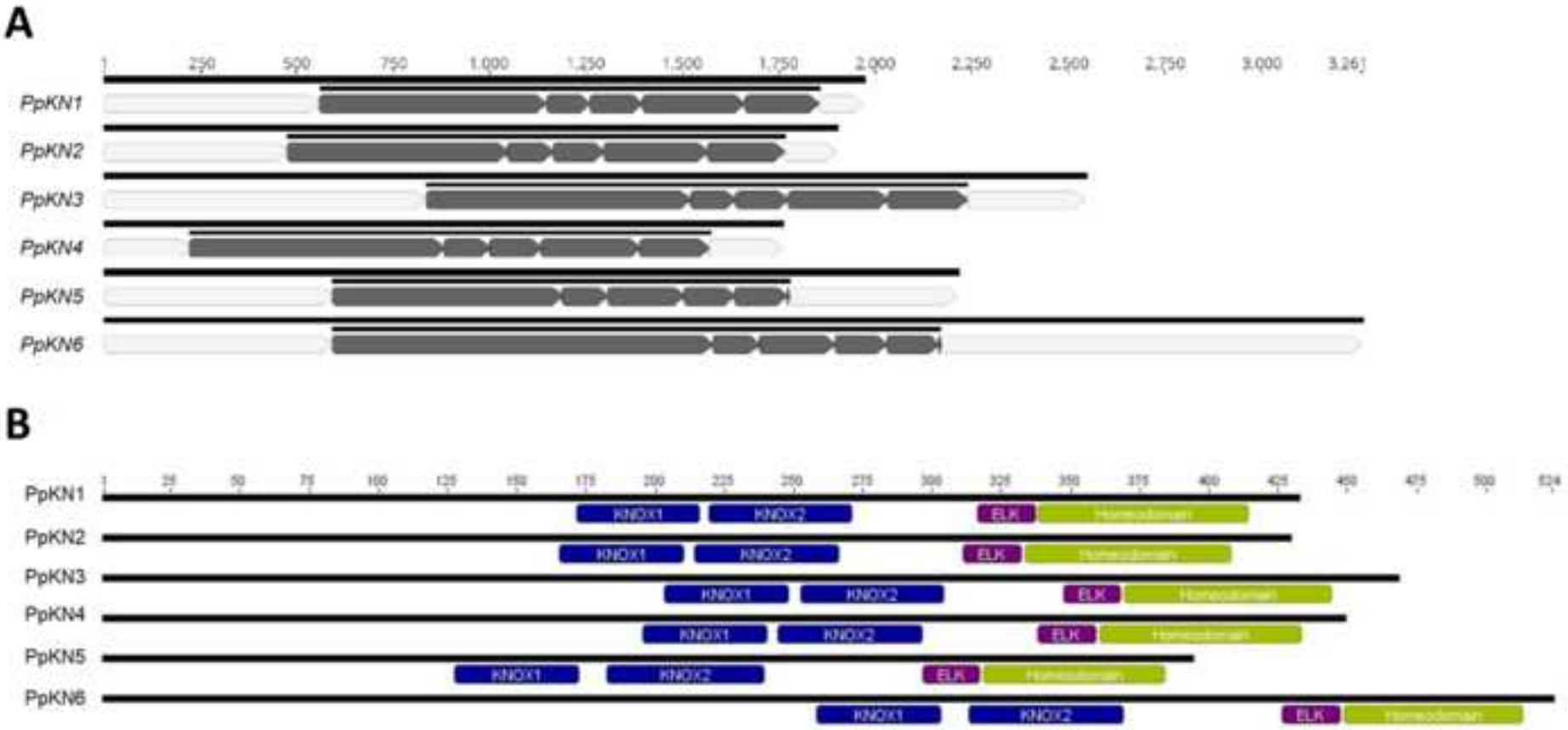

A
1
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,261
PpKN1
PpKN2
PpKN3
PpKN4
PpKN5
PpKN6
B
PpKN1
PpKN2
PpKN3
PpKN4
PpKN5
PpKN6
KNOX1
KNOX2
ELK
Homeodomain

Figure 2

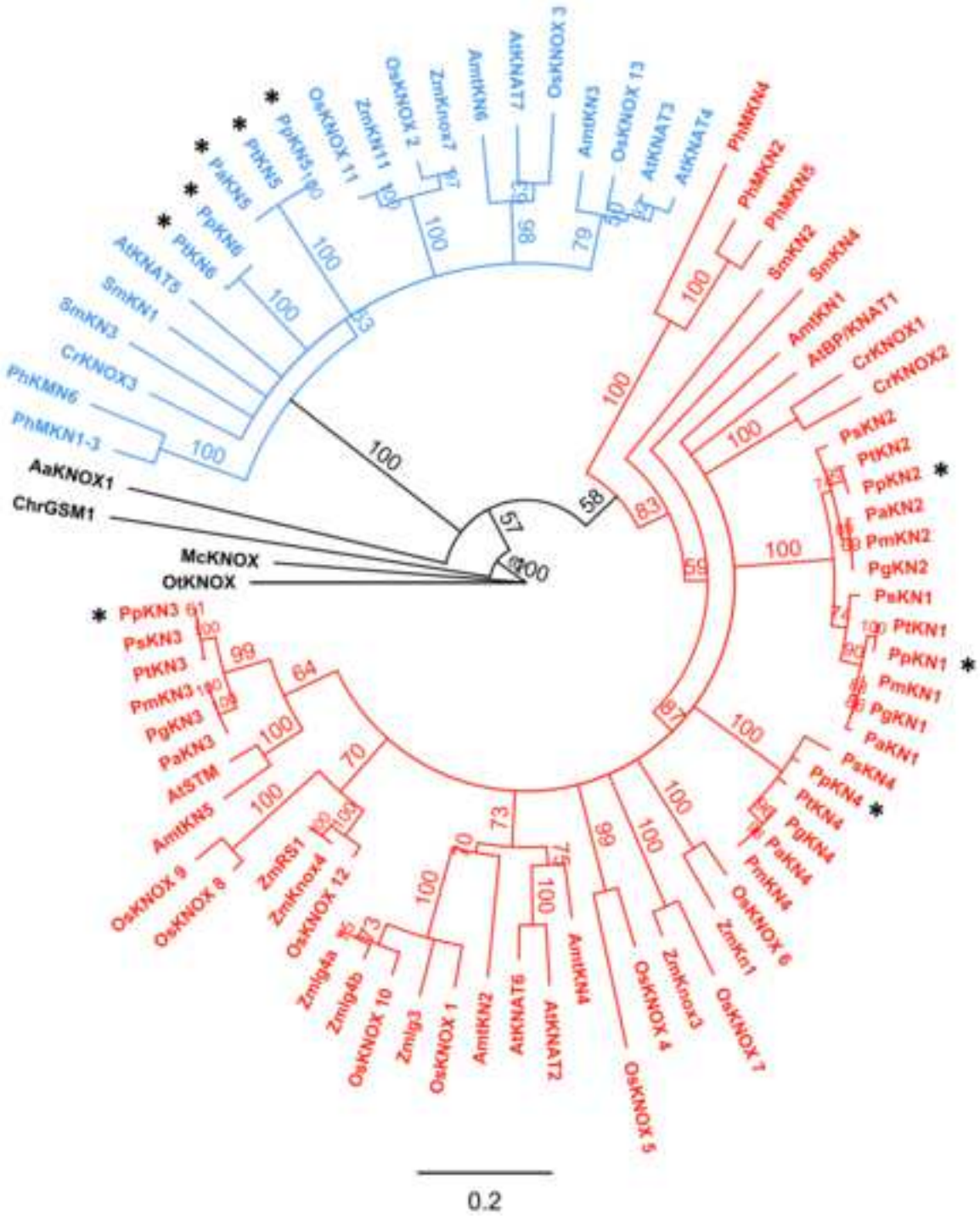
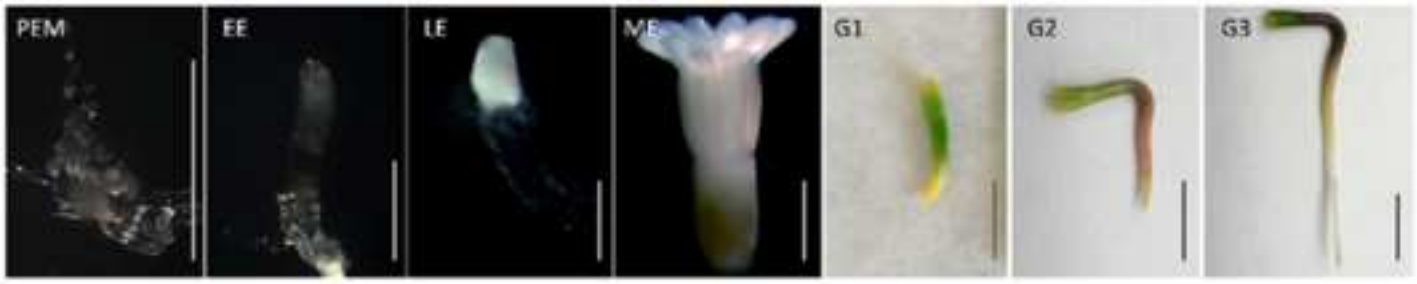
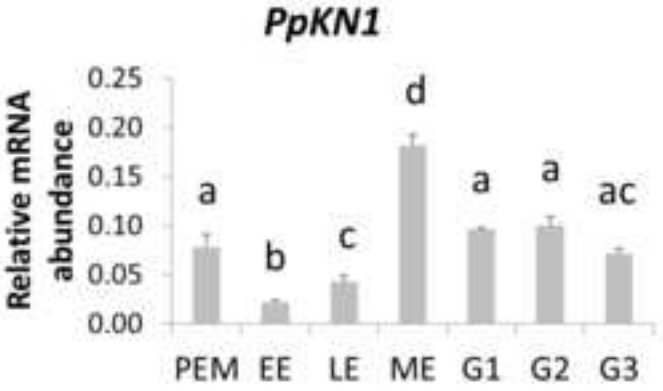
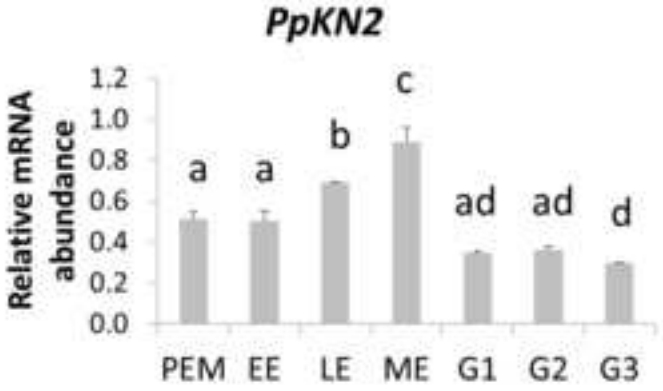
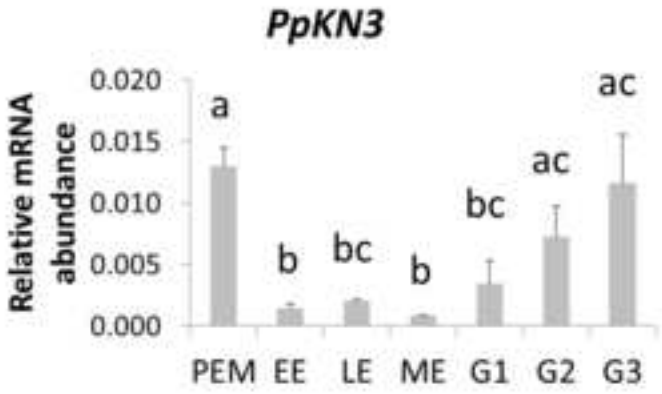

PpKN5
PaKN5
PtKN5
PpKN6
PtKN6
AtKNAT5
SmKN1
SmKN3
CrKNOX3
PhKMN6
PhMKN1-3
OsKNOX 11
ZmKnox11
OsKNOX 2
ZmKnox7
AmtKN6
AtKNAT7
OsKNOX 3
AmtKN3
OsKNOX 13
AtKNAT3
AtKNAT4
PhMKN4
PhMKN2
PhMKN5
SmKN2
SmKN4
AmtKN1
AtBP/KNAT1
CrKNOX1
CrKNOX2
PsKN2
PtKN2
PpKN2
PaKN2
PmKN2
PgKN2
PsKN1
PtKN1
PpKN1
PmKN1
PgKN1
PaKN1
PsKN4
PpKN4
PtKN4
PgKN4
PaKN4
PmKN4
OsKNOX 6
ZmKn1
ZmKnox3
OsKNOX 7
OsKNOX 4
OsKNOX 5
AmtKN4
AtKNAT2
AtKNAT6
AmtKN2
OsKNOX 1
ZmIg3
OsKNOX 10
ZmIg4b
ZmIg4a
OsKNOX 12
ZmKnox4
ZmRS1
OsKNOX 8
OsKNOX 9
AmtKN5
AtSTM
PaKN3
PgKN3
PmKN3
PtKN3
PsKN3
PpKN3
McKNOX
OtKNOX
ChrGSM1
AaKNOX1
0.2

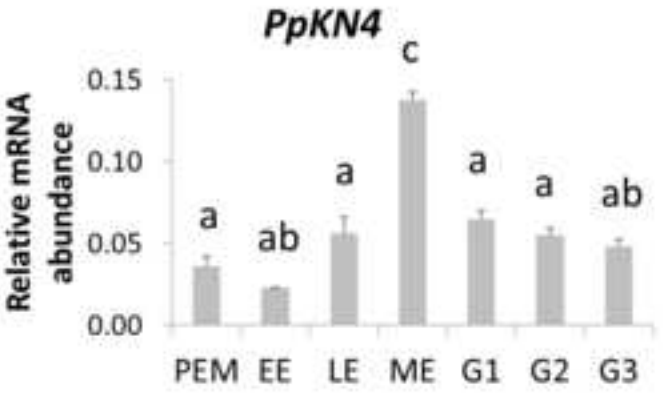
Figure 3

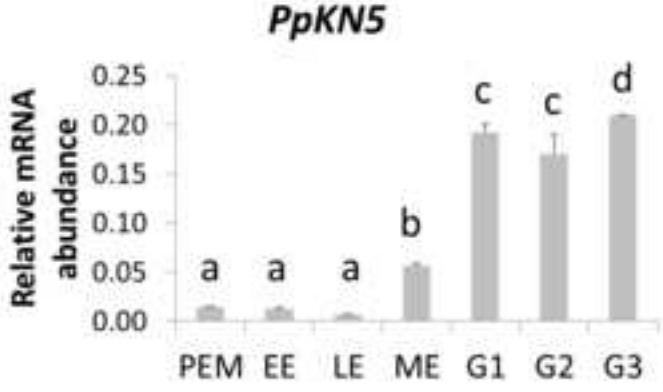

PEM
EE
LE
ME
G1
G2
G3


PpKN1
Relative mRNA abundance
0.25
0.20
0.15
0.10
0.05
0.00
a
b
c
d
a
a
ac
PEM EE LE ME G1 G2 G3


PpKN2
Relative mRNA abundance
1.2
1.0
0.8
0.6
0.4
0.2
0.0
a
a
b
c
ad
ad
d
PEM EE LE ME G1 G2 G3


PpKN3
Relative mRNA abundance
0.020
0.015
0.010
0.005
0.000
a
b
bc
b
bc
ac
ac
PEM EE LE ME G1 G2 G3


PpKN4
Relative mRNA abundance
0.15
0.10
0.05
0.00
a
ab
a
c
a
a
ab
PEM EE LE ME G1 G2 G3


PpKN5
Relative mRNA abundance
0.25
0.20
0.15
0.10
0.05
0.00
a
a
a
b
c
c
d
PEM EE LE ME G1 G2 G3

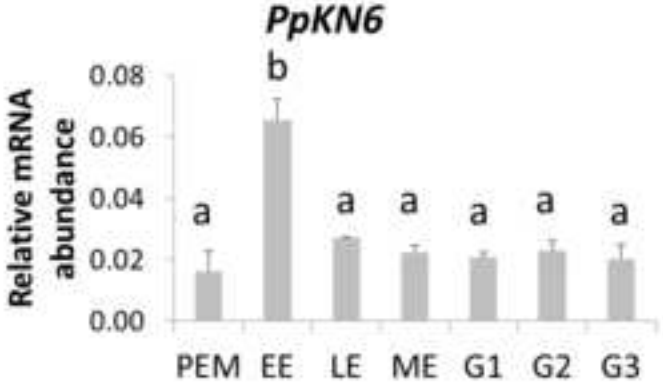

PpKN6
Relative mRNA abundance
0.08
0.06
0.04
0.02
0.00
a
b
a
a
a
a
a
PEM EE LE ME G1 G2 G3

Figure 4

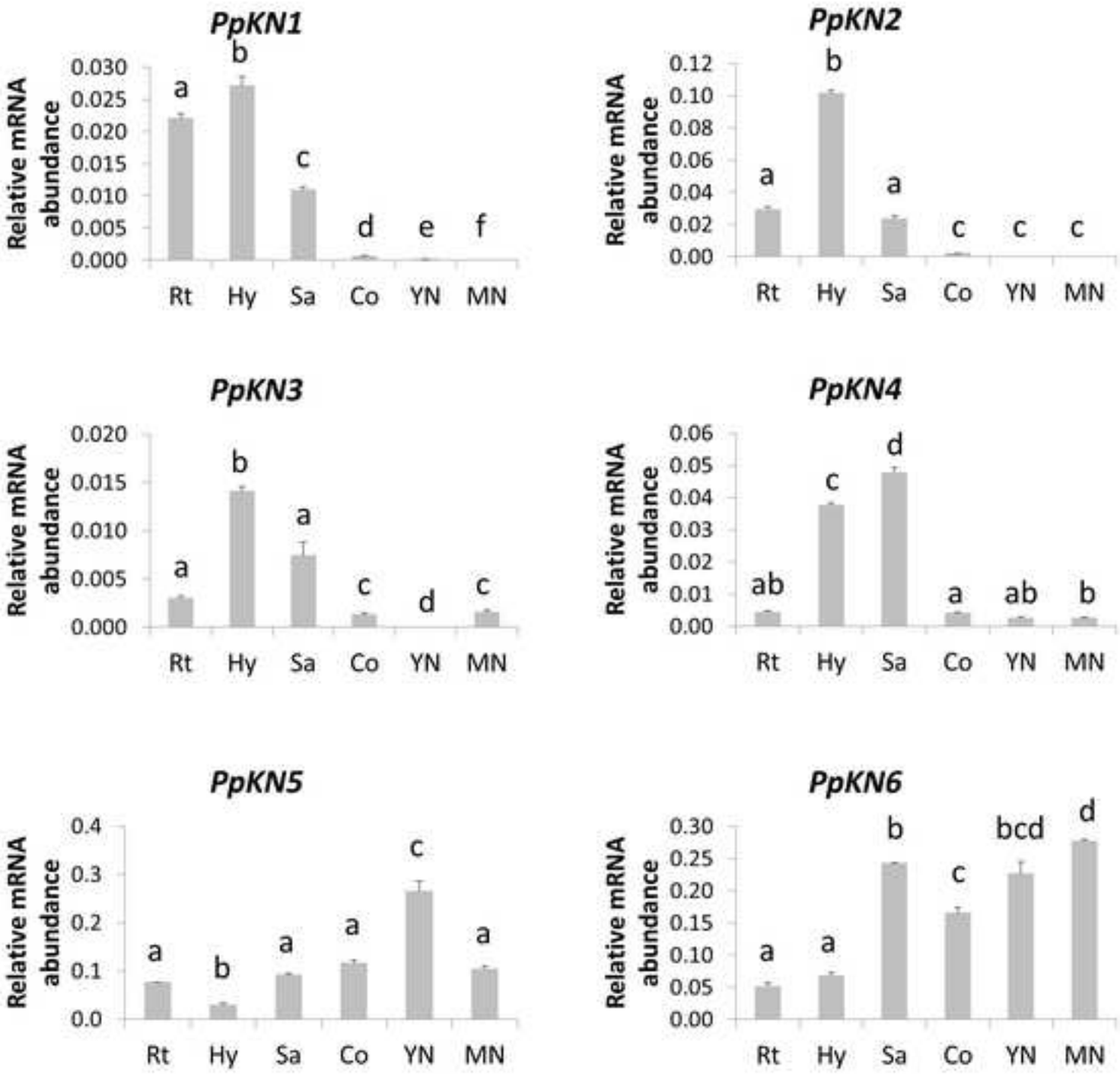

PpKN1
Relative mRNA abundance
0.030
0.025
0.020
0.015
0.010
0.005
0.000
a
b
c
d
e
f
Rt
Hy
Sa
Co
YN
MN
PpKN2
Relative mRNA abundance
0.12
0.10
0.08
0.06
0.04
0.02
0.00
a
b
a
c
c
c
Rt
Hy
Sa
Co
YN
MN
PpKN3
Relative mRNA abundance
0.020
0.015
0.010
0.005
0.000
a
b
a
c
d
c
Rt
Hy
Sa
Co
YN
MN
PpKN4
Relative mRNA abundance
0.06
0.05
0.04
0.03
0.02
0.01
0.00
ab
c
d
a
ab
b
Rt
Hy
Sa
Co
YN
MN
PpKN5
Relative mRNA abundance
0.4
0.3
0.2
0.1
0.0
a
b
a
a
c
a
Rt
Hy
Sa
Co
YN
MN
PpKN6
Relative mRNA abundance
0.30
0.25
0.20
0.15
0.10
0.05
0.00
a
a
b
c
bcd
d
Rt
Hy
Sa
Co
YN
MN

Figure 5

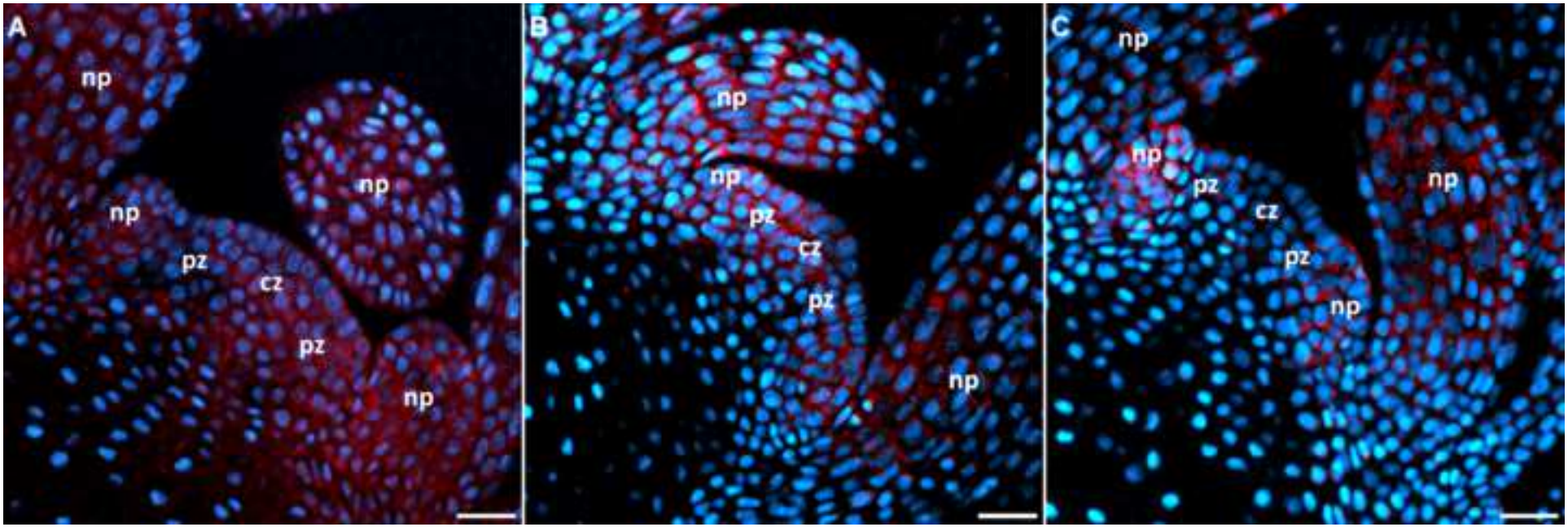
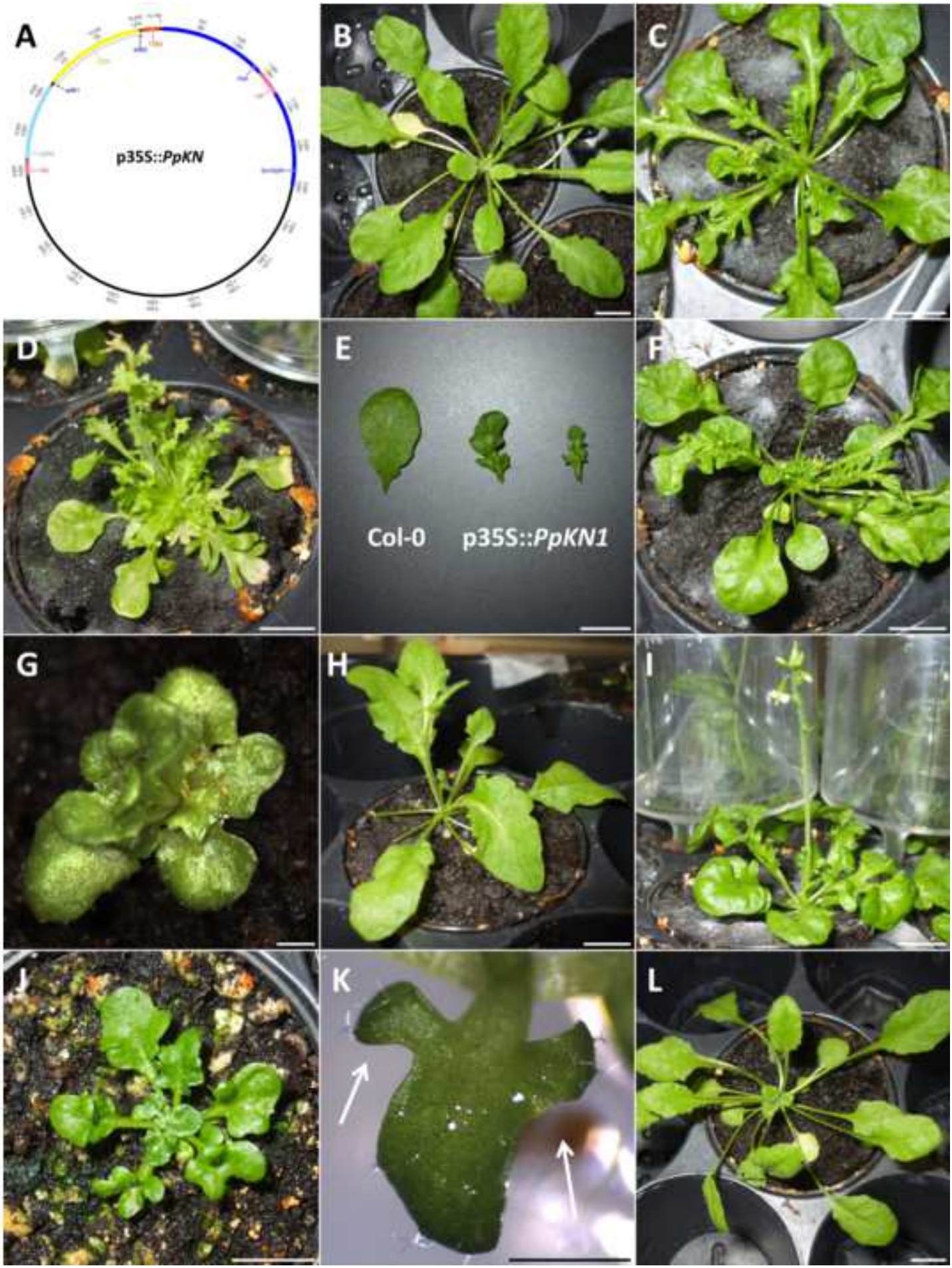


Figure 6



CONFLICT OF INTEREST

The authors declare that they have no conflict of interest. All authors revised and approved the final manuscript.