

Identification and Classification of LEA Family Genes in Orchids and Characterization of Their Role in Callus Formation

Chunmei He

South China Botanical Garden, CAS

Xuncheng Liu

South China Botanical Garden, CAS

Jaime A. Teixeira da Silva

Independent researcher

Haobin Wang

South China Botanical Garden, CAS

Tao Peng

South China Botanical Garden, CAS

Mingze Zhang

South China Botanical Garden, CAS

Can Si

South China Botanical Garden, CAS

Zhenming Yu

South China Botanical Garden, CAS

Jianwen Tan

South China Agricultural University

Jianxia Zhang

South China Agricultural University

Jianping Luo

Hefei University of Technology

Jun Duan (✉ duanj@scib.ac.cn)

Research article

Keywords: Callus induction · Classification Gene expression · Protocorm-like body ·Western blot

Posted Date: August 28th, 2019

DOI: <https://doi.org/10.21203/rs.2.12901/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The plant late embryogenesis abundant (LEA) proteins are abundant in seeds, play an important role in various abiotic stresses. However, there is still no information on genome-wide identification of LEA genes in orchids and their function in callus formation is almost unknown. In this study, the LEA genes from two orchids (*Phalaenopsis equestris* and *Dendrobium officinale*), were genome-wide identified, classified and characterized. A total of 57 and 59 LEA genes were identified in the genomes and these were divided into 8 and 9 groups for *P. equestris* and *D. officinale*, respectively. The LEA_1 and LEA_4 genes from *P. equestris* and *D. officinale* showed strong expression in seeds, but were significantly down-regulated in flowers and absent in vegetative organs (leaves, stems and roots). In addition, the LEA_1 and LEA_4 genes from *D. officinale* were abundant in the protocorm-like body (PLB) stage, while weak signals that were detected in in vitro shoots could not be detected in plantlets. The expression of these genes highlights PLBs in orchids are somatic embryos. The DoLEA36 from LEA_4 and DoLEA43 from LEA_1 were further characterized. The GFP signal of the DoLEA36-GFP fusion protein was only detected in the cytoplasm, while the GFP signal of the DoLEA43-GFP fusion protein was detected in both the cytoplasm and nucleus. This indicates that DoEA36 localizes in the cytoplasm while DoEA43 localizes in both the cytoplasm and nucleus. Both DoLEA36 and DoLEA43 stimulated callus formation in transgenic *Arabidopsis*. The percentage of callus formation from 35S::DoLEA43 transgenic lines was higher than in wild type plants in two callus induction methods. Our results provide comprehensive information about the LEA gene family in orchids and genetic evidence for the involvement of LEA genes in the induction of callus, which may reveal their positive role in the maintenance of PLBs in orchids.

Background

Plant seeds are composed of an embryo and an endosperm, which is not only the major source of energy for humans, but also the main form of dispersal for plants. Maturation drying is the last period of seed development, and is the developmental stage that is associated with a major loss of water [1]. To cope with this dramatic loss in water, plant seeds have developed a complicated mechanism to mitigate water loss and maintain cellular stability [2]. Among the actors involved in maintaining cellular stability during seed dehydration, late embryogenesis abundant (LEA) proteins are well known for their important role in assisting seeds survive under high dehydration conditions at later stages of seed maturation [3, 4].

In plants, LEA proteins are widely present, their expression is seed-specific and induced by abiotic stresses. LEA genes display very high level of expression at later stages of seed maturation when their encoded proteins accumulate considerably at this time [3]. For example, *PsLEAm* from pea (*Pisum sativum* L.), was not detected in whole seedlings 44 h after imbibition or in leaves from 3-week-old pea plants, but its expression could be re-induced in leaves after water stress [5]. The LEA gene *HVA1* from barley (*Hordeum vulgare* L.) and *WAP27* from mulberry tree (*Morus bombycina* Koidz.) could not be detected in vegetative tissues, but appeared in vegetative tissues after exposure to cold acclimation [6, 7]. This indicates that LEA genes are absent in vegetative tissues but that they reemerge in vegetative tissues after stress treatments, including water deficit and cold stress.

Multiple studies have demonstrated that LEA genes are involved in a wide range of water deficit and cold stress responses [8-12]. Although the defense mechanisms by LEA proteins remain largely unknown, many studies have demonstrated that LEA proteins help plants to deal with water deficit and cold stress by reducing water loss and maintaining cellular stability. On one hand, the LEA proteins contain a high proportion of polar amino acids and have a grand average hydropathy (GRAVY) value of less than 0 [13]. This hydrophilic property effectively allows LEA proteins to assist plants by reducing water loss under water deficit stress. On the other hand, LEA is an intrinsically disordered protein that can fold into an amphipathic α -helix under stresses [14]. LEA protein from liverwort (*Marchantia polymorpha* L.) shifted toward a structure with abundant α -helices and assisted α -casein in not aggregating during desiccation-rehydration events [15]. Moreover, LEA proteins are able to modulate membrane stability through binding and folding [16], and protect the activity of enzymes such as lactate dehydrogenase [17], malate dehydrogenase (MDH) and lactate dehydrogenase [18, 19], and β -D-galactosidase [19]. A recent study showed that rice (*Oryza sativa* L.) *OsLEA5* acts as a co-regulator to modulate the expression of ascorbate peroxidase *OsAPX1* [8].

In addition to the functions of LEA genes in response to abiotic stress, LEA genes might be involved in zygotic embryo development. The LEA_5 group of LEA genes from shrimp (*Artemia franciscana*) first appeared in diapause-destined embryos, declined slowly in the desiccation-resistant encysted stages after embryonic development resumed, but dropped rapidly when the embryo emerged from its shell [20]. Some LEA genes in plants are only present in seeds and in the stage between the seed and seedlings. For example, MLG3 polypeptide from maize (L.) was detected at an early stage of germination, but disappeared in the germinated seedling within 5 d of radicle emergence [21]. In addition, a mutation of a T-DNA insertion allele of *Arabidopsis thaliana* (L.) Heynh. *ATEM6* displayed incomplete seed development at the distal end of siliques, indicating that *ATEM6* is necessary for normal seed development [22].

The Orchidaceae is one of the largest families of flowering plants and *Phalaenopsis* and *Dendrobium* are two of the economically most important genera. The seeds of Orchidaceae only have an embryo without an endosperm [23]. Seed-specific LEA genes in orchids have been poorly explored, although 17 LEA genes were identified from *D. officinale* and their role in the response to salt and heat stress has been characterized [24]. In this study, we performed a global-scale identification of LEA genes in the genomes of two orchids, *P. equestris* and *D. officinale*, classified and explored their expression pattern in different organs and developmental stages, and also characterized their participation in callus formation. Our study provides comprehensive information of LEA genes in these two orchids as well as genetic evidence for the involvement of LEA genes in callus formation. This information will assist in future studies related to the mechanism of plant regeneration.

Results

Identification, classification, exon-intron composition of LEA genes in *P. equestris* and *D. officinale*

To obtain a comprehensive view of LEA genes in two orchids, genome-wide identification and analysis of LEA genes from *P. equestris* and *D. officinale* were conducted. In this study, 57 and 59 LEA genes from *P. equestris* and *D. officinale* were identified, respectively (Additional file 1 table 1). In Arabidopsis, LEA proteins have been classified into nine groups, namely LEA_1-5, AtM, dehydrin, SMP and PvLEA8 [13]. Base on a phylogenetic analysis of orchid and Arabidopsis LEA proteins, the *P. equestris* LEA proteins were clustered into eight major phylogenetic groups (LEA_1-4, PvLEA18 and NG1-3), while the LEA proteins from *D. officinale* contained nine groups (LEA_1-4, PvLEA18, SMP and NG1-3). Three groups (NG1-3) were found exclusively in these two orchids while AtM, LEA_5 and Dehydrin groups were absent in both *D. officinale* and *P. equestris* (Figure 1). In Arabidopsis, the SMP group contains six members. However, only one SMP gene (*DoLEA22*) was found in the *D. officinale* genome, but none in *P. equestris* (Figure 1).

Alternative splicing of mRNA precursors plays an important role in forming diverse mRNAs [25]. The size of exons/introns can affect splice-site recognition [26]. Hence, the composition of exons/introns of LEA genes in *P. equestris* and *D. officinale* was explored. Interestingly, the majority of LEA genes from *P. equestris* and *D. officinale* have conserved intron/exon structures. For example, 64.9% (37 of 57) of *PeLEAs* and 62.7% (37 of 59) of *DoLEAs* contain only one exon (Figure 2). This suggests that the structure of the LEA gene in orchids is conserved and may have resulted in functional conservation during their evolution.

Expression analysis of *PeLEAs* and *DoLEAs* in different tissues

To characterize the LEA genes, the expression pattern of LEA genes among vegetative organs (roots, stems and leaves) and reproductive organs (flowers and seeds) were surveyed by comparing the FPKM values for each gene in *P. equestris* and semi-quantitative RT-PCR in *D. officinale*. The NG1-3, PvLEA18, LEA_2 and LEA_3 groups in both orchids did not display specific expression patterns (Figure 3). *DoLEA12*, -18, -23, -31 and *DoLEA53* from NG1-3 groups were widely expressed in roots, stems, leaves, flowers, and seeds (Figure 3). Only one PvLEA18 gene was found in both *P. equestris* and *D. officinale*, although they shared different expression patterns. *PeLEA43* from the PvLEA18 group in *P. equestris* was strongly expressed in seeds and flowers, but *D. officinale DoLEA7* from the PvLEA18 group was absent in flowers (Figure 3). In contrast, the LEA_1 group members in both orchids displayed a seed-specific expression pattern. All seven LEA_1 group genes (*PeLEA1*, -23, -24, -50, -6, -7 and -8) in *P. equestris*, and all LEA_1 group genes from *D. officinale* (except for *DoLEA3*), showed a strong expression pattern in seeds (Figure 3). In addition, LEA_1 group members such as *PeLEA6*, *DoLEA43* and *DoLEA48* were also strongly expressed in flowers (Figure 3). Two LEA_2 group genes (*PeLEA4* and *PeLEA47*) were found in *P. equestris*, *PeLEA7* displayed no specific expression pattern in all the detected tissues, while *PeLEA4* showed higher expression in seeds and lower expression in flowers, similar to the LEA_1 group genes (Figure 3). *DoLEA13*, -17, -36, -4 and -5 from the *D. officinale* LEA_4 group were highly expressed in seeds, but expression was absent in all vegetative organs (Figure 3). LEA_1 and LEA_4 displayed a similar

expression pattern, suggesting that they might have a similar function, whereas the other groups had diverse expression patterns.

Expression of LEA genes from *D. officinale* at different developmental stages

PLBs are a specific developmental stage in orchids, and are regarded as somatic embryos as a result of their similar macroscopic and microscopic features [27]. We used four developmental stage samples namely PLBs, multiple shoots from PLBs, and two developmental stage plantlets of *D. officinale* (see details in the materials and methods section), to survey the changes of LEA genes during development. Very interestingly, most genes in LEA_1 and LEA_4 groups showed high expression at the PLB stage (T1), while their expression became lower as PLBs developed to plantlets in the two detection methods (Figure 4). For example, *DoLEA4*, -13, -17, -36 and -45 from the LEA_4 group, and *DoLEA9*, -11, -27, -43 and -48 from LEA_1 were only strongly detected at the PLB stage but were significantly down-regulated or absent at the other three developmental stages namely T2, T3 and T4 (Figure 4). The LEA_2 group has a close phylogenetic relationship with the LEA_1 and LEA_4 groups, and most genes in this clade also displayed a similar expression pattern with genes from LEA_1 and LEA_4 groups. For instance, *DoLEA2*, -19 and *DoLEA55* from the LEA_2 group were highly expressed at the PLB stage after semi-quantitative RT-PCR analysis (Figure 4). However, the genes from NG1-3 groups displayed diverse expression patterns in the four developmental stages.

The localization of DoLEA36 and DoLEA43

To further analysis the function of the LEA_1 and LEA_4 group genes, the *DoLEA36* gene from LEA_4 and *DoLEA43* from LEA_1 were selected to explore their localization and further function analysis based on their expression patterns. The empty vector of pCAMBIA1302 transformed into Arabidopsis plants was used as the positive control. The green fluorescent signals of the seedlings harboring an empty vector observed in the cytoplasm, plasma membrane and nucleus were strong (Figure 5A). In 35S:*DoLEA36-GFP* seedlings, green fluorescence filled cells, which suggests that the DoLEA36-GFP fusion protein is located in the cytoplasm (Figure 5A). In contrast, the 35S:*DoLEA43-GFP* lines showed green fluorescent signals in the cytoplasm and nucleus (Figure 5A). To further verify the localization of the two proteins, the cytoplasm proteins and nucleus proteins were isolated from both 35S:*DoLEA36-GFP* and 35S:*DoLEA43-GFP* transgenic plants. Western blot analysis showed that the GFP of 35S:*DoLEA36-GFP* plants was only detected in the cytoplasm (Figure 5B). However, the GFP of 35S:*DoLEA43-GFP* plants was detected in the cytoplasm and nucleus (Figure 5B). This result indicates that DoLEA36 localized in the cytoplasm while DoLEA43 localized in the cytoplasm and nucleus. DoLEA36 and DoLEA43 had different localization although they shared a similar expression pattern.

DoLEA36 and *DoLEA43* have different roles in stimulating callus formation

Both *DoLEA36* and *DoLEA43* were strongly detected in seeds and PLBs of *D. officinale*, but displayed different localization. Hence, the callus formation rate in the transgenic lines of *DoLEA36* and *DoLEA43* driven by a 35S promoter were analyzed. The RT-PCR result (Figure 6A) and the western blot analysis (Figure 6B) revealed that both *DoLEA36* and *DoLEA43* genes were transcribed and expressed successfully in transgenic lines. The hypocotyl without a cotyledon and roots was planted to plant growth regulators (PGR)-free half-strength MS medium and cultured in the dark for two days, then transferred to a 16-h photoperiod. Callus formed at the wounding site in a small number of explants. The WT plants and transgenic lines displayed a different callus induction rate (Figure 6C-F). The callus induction rate of 35S:*DoLEA36* lines was about two-fold higher than WT plants (Figure 6C and D) while the callus induction rate of 35S:*DoLEA43* lines was three-fold higher than WT plants 14 d after wounding (Figure 6E and F). The callus initiation of all transgenic lines exceeded 15%, while that of WT plants was only about 6% (Figure 6).

To further explore callus induction, 6 d-old seedlings after stratification were dissected and the upper end of hypocotyls were removed, while the hypocotyl containing roots were incubated on PGR-free half-strength MS medium in the dark. The 35S:*DoLEA36* lines and WT did not have statistically significantly different callus induction 12 d after wounding (Figure 7A and B). In contrast, the callus induction rate of all 35S:*DoLEA43* lines was higher than WT plants 12 d after wounding (Figure 7C). The 35S:*DoLEA43* lines showed increased callus initiation compared with WT plants (Figure 7D). These results suggest that *DoLEA36* and *DoLEA43* play a role in callus formation, while *DoLEA43* is more effective in callus induction than *DoLEA36*.

The expression patterns of *DoLEA36* and *DoLEA43* under wounding stress

Increasing data proved that wounding triggers callus formation [28, 29]. Hence, the expression of *DoLEA36* and *DoLEA43*, which played a role in callus formation, was explored under wounding stress. Total RNA at 0, 2, 5, 10 and 25 h after wounding were applied to qRT-PCR analysis. As expected, *DoLEA36* and *DoLEA43* genes were up-regulated in these time points (Figure 8A). *DoLEA36* displayed an increasing trend at 2, 5, 10 and 25 h after wounding, while *DoLEA43* initially increased within 10 h then maintained its transcriptional level (Figure 8A). Both *DoLEA36* and *DoLEA43* showed about 3.5-fold higher expression 2 h after wounding, while became about 15- and 52-fold higher at 10 h after wounding, respectively (Figure 8A). This indicates that both genes were modulated by wounding.

No wounding responsive element (WUN-motif, AAATTACTA) was found in the putative promoter of *DoLEA36*, while one WUN-motif was present at the -453 bp site of the *DoLEA43* promoter (Figure 8B). In addition, two WUSATAg motifs (TTAATGG) were found at -1014 and -1440 bp sites of the *DoLEA43* promoter (Figure 8B). These results thus establish that both *DoLEA36* and *DoLEA43* could be induced by wounding and might help explants generate callus.

Discussion

Classification of LEA gene family in plants

LEA genes exist widely in plants, including in chlorophyte, liverwort, lycophyte, gymnosperm, and angiosperm plant lineages [30]. In many higher plants, the LEA genes were classed into eight conserved groups namely LEA_1-5, PvLEA18, dehydrin and SMP. These eight groups were found in eudicot species such as Chinese plum (*Prunus mume* Siebold & Zucc.) [31] and rapeseed (*Brassica napus* L.) [32]. The monocotyledonous plant sorghum (*Sorghum bicolor* (L.) Moench) contains LEA_1-5, PvLEA18, dehydrin and SMP groups in its genome [33]. Moreover, seven groups, including LEA_1-5, dehydrin and SMP, were detected in genome of a gymnosperm, Manchurian red pine (*Pinus tabuliformis* Carr.) [34]. These findings suggest that the evolution of LEA genes are conserved and present in angiosperm and gymnosperm lineages, indicating that most LEA groups existed before the split between monocotyledonous and dicotyledonous plants. In this study, the monocotyledonous plants *P. equestris* and *D. officinale* contain five (LEA_1-4 and PvLEA18) and six (LEA-4, PvLEA18 and SMP) groups of the eight conserved groups in their genomes (Figure 1). NG1-3 groups were only found in the two orchids, but AtM, LEA_5 and dehydrin were absent in both *P. equestris* and *D. officinale* when classified with *Arabidopsis* LEA proteins by phylogenetic analysis (Figure 1). These results suggest that AtM, LEA_5 and dehydrin might have been lost and that new groups were generated during the evolution of orchids. The loss of gene families is common in orchids. For example, the FLC, AGL12 and AGL15 subfamilies of the MADS-box superfamily was absent in both *P. equestris* and *D. officinale* [35, 36]. The loss or gain of LEA groups has also been observed in other higher plants. For example, no PvLEA18 gene was detected in tea (*Camellia sinensis* (L.) Kuntze), LEA_5 and SMP genes were found in moso bamboo (*Phyllostachys edulis* (Carrière) J.Houz.), while other atypical LEA genes were found in cucumber (*Cucumis sativus* L.) and stiff brome (*Brachypodium distachyon* (L.) P.Beauv.) [37].

The dehydrin genes known to protect plants against abiotic stresses [38]. However, no dehydrin gene was found in these two orchids, indicating that the response to abiotic stresses mediated by these genes in both orchids might be different from plants containing dehydrin genes. The generation of NG1-3 in both orchids might attain new functions and help them cope with variable environments or other physiological mechanisms. Functional characterization is needed to determine the importance of these genes. No AtM gene was found in both orchids, which is consistent with other higher plants such as *Prunus mume* [31] and *Sorghum bicolor* [33]. AtM is regarded as a Brassicaceae-specific group that is found exclusively in Brassicaceae species [13].

The LEA_1 and LEA_4 group genes were abundant in seeds and PLBs

The plant LEA protein was first isolated from cotton (*Gossypium hirsutum*), and showed extremely abundant expression in seeds [39]. The LEA gene got its name from cotton Late Embryogenesis Abundant [40]. Plant LEA proteins mainly accumulate in zygotic embryos [21, 39]. In *Arabidopsis*, more

than 50% (28 of 51) of LEA genes showed abundant expression in seeds, especially genes from the SMP, LEA_1 and LEA_4 groups [13]. *PsLEAm* from pea (*Pisum sativum* L.) is a LEA_4 group member and was detected during late seed development and was strongly expressed in the desiccation period, but could still be observed during germination, but no signal was detected 44 h after imbibition [5]. Most of these two group genes in *P. equestris* and *D. officinale* show pronounced expression in seeds. This indicates that LEA_1 and LEA_4 members are seed-specific genes in plants and might play a conserved and important role in seed development.

PLBs are regarded as somatic embryos in orchids [27, 41]. Shoots can regenerate from the PLBs followed by the regeneration of roots to form a plantlet [42]. These developmental stages are similar with plants regenerated from embryonic callus. Most of the LEA_1 group genes (*DoLEA9*, -11, -29, -43 and *DoLEA48*) and LEA_4 group genes (*DoLEA4*, -13, -17, -36 and *DoLEA45*) were highly expressed at the PLB stage, but were not observed or only slightly detected in multiple shoots and plantlets of *D. officinale*. Similarly, in sweet orange (*Citrus sinensis* L. Osb.), most of the genes from the LEA_1 and LEA_4 groups showed preferential expression in fruit and embryogenic callus [43]. These results indicate that the LEA_1 and LEA_4 genes are highly expressed in embryogenic tissues such as PLBs and embryonic callus, and may play a positive role in plant somatic embryogenesis. The expression of these genes further fortifies the notion that PLBs are somatic embryos and provides the first molecular evidence to support this concept.

Different roles of *DoLEA36* and *DoLEA43* genes in triggering callus formation

LEA_1 and LEA_4 are closely related, as per a phylogenetic analysis (Figure 1). The *DoLEA36* gene from the LEA_4 group and *DoLEA43* from the LEA_1 group showed similar expression in different organs and different developmental stages (Figure 3 and 4, respectively). Both genes were strongly induced after wounding, especially the *DoLEA43* gene (Figure 8). This is similar to the expression pattern of a LEA gene from *Papaver somniferum* that displayed 4-fold higher expression than the control after wounding treatment for 5 h in the whole seedlings [44]. In addition, one WUN-motif was found in the promoter of *DoLEA43*, but none in the *DoLEA36* promoter (Figure 8B). This suggests that the regulation of expression in wounding by transcription factors might differ for the two genes. Two WUSATAg motifs were found in the *DoLEA43* promoter, but no WUSATAg motif in the *DoLEA36* promoter. WUSCHEL homeodomain transcription factors are required to maintain stem cells in the shoot apical meristem and prevent premature cell differentiation, and can recognize and bind to WUSATAg motifs [45-48]. The callus induction rate increased significantly in both callus induction methods when *DoLEA43* was over-expressed in *Arabidopsis*. The callus induction rate of 35S:*DoLEA36* plants was higher than WT plants, but lower than 35S:*DoLEA43* plants during callus induction. Taken together, these results indicate that *DoLEA36* and *DoLEA43* genes play different roles in triggering callus formation, and might be required for the maintenance or proliferation of PLBs in orchids.

Conclusions

In this study, a total of 57 and 59 the LEA genes from *P. equestris* and *D. officinale* were identified at the genome-wide scale, and these could be divided in 8 and 9 groups, respectively. The expression of LEA in different organs and different developmental stages were explored and showed a conserved expression pattern in some groups. Genes from the LEA_1 and LEA_4 groups were abundantly expressed in seeds and PLBs. Two LEA genes, *DoLEA36* and *DoLEA43*, shared similar expression patterns but different localization. These two genes stimulate callus formation but displayed different roles in callus induction in transgenic Arabidopsis. Our study provides insight into the mechanism of maintenance of PLBs in orchids and plant regeneration.

Methods

Plant materials

D. officinale was collected from the Guangxi Zhuang Autonomous Region in China and was identified by professor Jun Duan. All authors comply with the Convention on the Trade in Endangered Species of Wild Fauna and Flora. The *D. officinale* plants at the reproductive stage were grown in a greenhouse at South China Botanical Garden (Guangzhou, China) in The roots, stems, leaves, flowers and seeds (120 days after pollination) from at least six potted plants were harvested and used for gene expression analysis. Protocorm-like bodies (PLBs, T1 stage), multiple shoots from PLBs (T2 stage) and plantlets (T3, about 2 cm in height; T4, about 8 cm in height) were cultured on half-strength Murashige and Skoog (MS) [49] medium supplemented with 0.5 mg l⁻¹ 1-naphthalene acetic acid (NAA), 1 g l⁻¹ activated carbon, 20 g l⁻¹ sucrose and 6 g l⁻¹ agar at pH 5.4, were grown in a growth chamber under 40 μmol m⁻² s⁻¹ light intensity and a 12-h photoperiod. All samples were collected, frozen in liquid nitrogen, then RNA was extracted immediately or placed rapidly at -80 °C.

A. thaliana (Columbia ecotype) obtained from professor Keqiang Wu (South China Botanical Garden, Chinese Academy of Sciences) was used as the wild type (WT) and. Arabidopsis plants were cultivated in a mixture of topsoil and vermiculite (1:3, v/v) under a 16-h photoperiod, 100 μmol m⁻² s⁻¹ light intensity, and at 22 °C in a growth chamber. Arabidopsis seeds were sown on half-strength MS medium [49] containing 15 g l⁻¹ sucrose and 8 g l⁻¹ agar and grown in the same conditions as the Arabidopsis plants.

Identification of LEA gene family in two orchids, and exon-intron prediction

The peptide and General Feature Format (GFF) files of *P. equestris* and *D. officinale* were obtained from the National Center of Biotechnology Information (NCBI) provided by Zhang et al. [35-36], respectively. The proteins in the peptide files were annotation using NCBI non-redundant (NR), Gene Ontology (GO), Pfam [50], Swissprot and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The genes were annotated as the LEA genes were considered as candidate LEA genes, then rechecked in NCBI BLASTP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

For the prediction of gene structure, the GFF annotation of LEA genes was obtained from GFF files of *P. equestris* and *D. officinale* and submitted to the Gene Structure Display Server (GSDS) 2.0 (<http://gsds.cbi.pku.edu.cn/>) [51] to generate a diagram of gene structure.

Multiple sequence alignments and phylogenetic analysis

The LEA proteins were aligned using MAFFT version 7 software [52] to generate a FASTA alignment file. The alignment file was uploaded into MEGA version 7 [53] to construct a phylogenetic tree by the neighbor-joining (NJ) method.

Construction of the 35S::LEA-GFP vector and generation of transgenic lines

The coding sequences of *LEA36* and *DoLEA43* without a termination codon were amplified by KOD-Plus DNA polymerase kit (Toyobo Co., Ltd., Osaka, Japan) and inserted in the *N*col site of the pCAMBIA 1302 vector using the In-Fusion HD Cloning Kit (Takara Bio Inc., Dalian, China) according to the manufacturer's instructions. The transformation of *Agrobacterium tumefaciens* strain EHA105 with the correct construct was performed using the freeze and thaw method [54]. Thereafter, *A. tumefaciens* harboring the construct was transformed into Arabidopsis by the floral dip method [55] to generate the overexpression transgenic lines.

Subcellular localization

The transgenic seeds of 35S::*DoLEA36* and 35S::*DoLEA43* were surface-sterilized in 1% NaClO for 10 min, washed in sterile distilled water six times, then sown on half-strength MS medium in Petri dishes containing 1.5% sucrose and 0.8% agar (pH 5.7). Plates were incubated at 4 °C in the dark for 2 d for stratification, then cultured in a growth chamber. After stratification, 4 day-old seedlings were used to survey GFP fluorescence and photographed with a Zeiss LSM 510 confocal microscope (Zeiss, Jena, Germany). Transgenic plants that only contained pCAMBIA 1302 were used as the positive control.

Protein extraction from Arabidopsis

Total protein was extracted according to He et al. [56]. Briefly, one week old seedlings were harvested, cleaned and ground into a fine powder in liquid nitrogen. Thereafter, an extraction buffer [50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM MgCl₂, 1 mM DTT, 20% glycerol, 1% NP-40 and containing a protease inhibitor cocktail (Roche, Mannheim, Germany)] was added then centrifuged at 14,000 g for 20 min. The

supernatant containing total proteins was transferred into a proteinase-free 1.5 ml Eppendorf tube and used for Western blot analysis immediately or frozen in liquid nitrogen for 2 min then stored at -80 °C.

One week old seedlings (200 mg) were also used to extract the cytoplasmic proteins. The Cytoplasmic Protein Extraction Kit (BestBio Inc., Shanghai, China) was used to extracted cytoplasmic proteins according to the manufacturer's protocol. For the extraction of nuclear proteins, 4 g of seedlings were ground in liquid nitrogen into a fine powder, then 20 ml of lysis buffer [20 mM Tris-HCl (pH 7.4), 25% glycerol, 20 mM KCl, 2 mM EDTA, 2.5 mM MgCl₂, 250 mM sucrose, 1 mM DTT and 1 mM PMSF] was added, washed six times with a washing buffer [20 mM Tris-HCl (pH 7.4), 25% glycerol, 2.5 mM MgCl₂ and 0.2% Triton X-100], and washed once with nuclei resuspension buffer [20 mM Tris-HCl (pH 7.4), 25% glycerol and 2.5 mM MgCl₂]. The nuclei storage buffer (200 µL) containing the Roche protease inhibitor cocktail was used to suspend the nuclear proteins. The cytoplasmic and nuclear proteins were used for Western blot analysis immediately or frozen in liquid nitrogen for 2 min then stored at -80 °C within 1 month.

SDS-PAGE and Western blot

The protein samples were degenerated in 1× SDS loading buffer and incubated in boiling water for 10 min, then centrifuged at 12,000 rpm for 2 min. The supernatant, which included the degenerated proteins, was used immediately for SDS-PAGE analysis. SDS-PAGE was carried out in 12% gels, then proteins were transferred from gels to a polyvinylidene fluoride (PVDF) membrane by a wet transfer apparatus (Bio-Rad Laboratories, Hercules, CA, USA). After blocking with 5% skimmed milk for 1.5 h at room temperature, the membrane was incubated with primary antibody for 1.5 h at room temperature, then washed with 1× TBST buffer (Solarbio Science & Technology Co., Ltd., Beijing, China) three times (each wash was 10 min). The membrane was then incubated with secondary antibodies at room temperature for 1.5 h. After washing membranes in 1× TBST buffer three times, the anti-GFP mouse monoclonal antibody (1: 5,000 dilution, Thermo Fisher Scientific, Waltham, MA, USA; cat#GF28R) was used as the probe to detect the level of GFP-tagged proteins. Actin and histone H3 were probed by Anti-Plant Actin Mouse Monoclonal Antibody (A01050; Abbkine, Wuhan, China) and a-H3 (ab1791; Abcam, Cambridge, UK) at 1: 5000, respectively. The goat anti-rabbit IgG-HRP (catalog number sc-2301, Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) and goat-anti mouse IgG HRP antibody (Thermo Fisher Scientific; #62-6520) were used as the secondary antibodies depending on the primary antibody. Protein bands were visualized in a ChemiDoc™ MP Imaging system (Bio-Rad, Hercules, CA, USA). The protein marker used was EasySee® Western Marker (DM201, TransGen Biotech, Beijing, China).

Semi-quantitative RT-PCR and quantitative real-time PCR (qRT-PCR) analysis

Total RNA was isolated with an RNA extraction kit (RNAout 2.0, Tiandz Inc., Beijing, China) according to the manufacturer's protocol. One µg of RNA was used to synthesize cDNA by reverse transcription PCR (RT-PCR) of each reaction using a MonScript™ PTIII all-in-one Mix Reverse Transcriptase Kit (Monad Biotech Co., Ltd., Suzhou, China). cDNA content was diluted to 200 ng µl⁻¹ and 1 µl was used as template for PCR amplification using a 20 µl reaction system, then amplified by 35 cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 1 min. After amplification, 5 µl of each reaction was surveyed by 1% agarose gel electrophoresis and visualized using the Gel Imaging System (GenoSens1880, Shanghai Qinxiang Scientific Instrument Co., Ltd., Shanghai, China). For quantitative real-time PCR (qPCR) analysis, total RNA was extracted, cDNA was generated as indicated above, and qPCR was performed as described by He et al. [56]. All specific primer pairs are listed in Additional file 2 table 2.

Gene expression analysis based on RNA sequencing data

For expression analysis of LEA genes from *P. equestris*, the transcriptome sequencing data of *P. equestris* roots (SRR2080194), stems (SRR2080200), leaves (SRR2080202), flowers (SRR2080204) and seeds (SRR3606718) were obtained from the NCBI Sequence Read Archive provided by Niu et al. [57]. The clean reads were mapped to the *P. equestris* genome by TopHat version 2.0.8 (Kim et al. 2013) and gene expression level was calculated by the fragments per kilobase of exon per million fragments mapped (FPKM) method using HTSeq [58]. For the expression analysis of LEA genes from *D. officinale* at different developmental stages (see plant materials section), FPKM were obtained from our *D. officinale* developmental database, which constructed in our laboratory. This data is available upon reasonable request.

Callus induction

In this study, we explored callus formation via two callus induction methods. In the first method, hypocotyls from 6-day-old Arabidopsis seedlings after stratification were cut by disposable knives about 0.5 mm from the cotyledon-hypocotyl junction and at about 0.5 mm from the hypocotyl-root junction. The center of hypocotyls was transferred carefully to Petri dishes containing half-strength MS medium with 1.5% sucrose and 0.8% agar (pH 5.7) with a sterile toothpick, incubated for 2 days in the dark and grown in a growth chamber. In the second method, hypocotyls were cut once at about 2 mm from the hypocotyl-root junction, the hypocotyl-root explants were transferred to Petri dishes containing half-strength MS medium with 1.5% sucrose and 0.8% agar (pH 5.7) and incubated in the dark at 22 °C. Callus was observed at the end of the cut site. Callus on an explant was considered to be induced if the callus was visible on a Leica S8 APO stereomicroscope (Leica Microsystems Ltd., Heerbrugg, Switzerland). Callus induction was quantified as a percentage of explants forming callus. Every experiment was repeated in triplicate. At least 50 seedlings of WT and transgenic lines from each experiment were used.

Wounding treatment

PLBs growing on half-strength MS medium containing 2.0% sucrose, 0.5 mg/L NAA and 0.6% agar (pH 5.4) were used for the wounding treatment. The PLBs were sliced into 2 mm thick slices and placed on the same medium. Explants at 0, 2, 5, 10, and 25 h after wounding were harvested and total RNA was extracted using the method described above. *DoLEA36* and *DoLEA43* expression was detected by qPCR analysis. Three biological replicates were performed for each sample.

Prediction of *cis*-responsive element in *DoLEA36* and *DoLEA43* promoters

To further explore the gene-responsive factors, 2000 bp upstream of the initiation codon from *DoLEA36* and *DoLEA43* were obtained from the *D. officinale* genome and used to predict the *cis*-responsive elements by PLANTCARE [59] and PLACE [60].

Statistical analyses

All the data of induction rate was analyzed by SigmaPlot12.5 software (Systat Software Inc., San Jose, CA, USA) using one-way analysis of variance (ANOVA) followed by Dunnett's test. $P < 0.05$ was considered to be statistically significant.

Abbreviations

LEA, late embryogenesis abundant; GRAVY, grand average hydropathy; MDH, malate dehydrogenase; FPKM, fragments per kilobase of transcript per Million fragments mapped; PGR, plant growth regulators; qRT-PCR, quantitative real time polymerase chain reaction; PLB, protocorm-like body; GFP, green fluorescent protein; N, Nuclear; NN, non-nuclear; PVDF, polyvinylidene fluoride

Declarations

Acknowledgements Many thanks to Professor Xinlan Xu (South China Botanical Garden, Chinese Academy of Sciences) for assistance in using the Zeiss LSM 510 confocal microscope.

Funding This research was funded by the Natural Science Foundation of Guangdong Province Projects (grant number 2018A030313603) and the National Science and Technology Planning Project of Guangzhou (grant number 201704020010). The funding bodies was used for experimental cost and publication fees.

Authors' contributions JD supervised the project. CH conceived the research and designed the experiments. JT and JL helped CH design the experiments. HW and MZ constructed the vector, generated the transgenic lines and conducted semi-quantitative RT-PCR. TP performed Western blot analysis. CS performed qPCR. JZ provided the plant materials. ZY performed the gene structure analyses. CH, XL and JATdS collectively interpreted the results and wrote all drafts of the manuscript. All authors approved the final draft for submission and take full public responsibility for the content of the manuscript.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declared that they have no competing interest.

Availability of data and materials The data sets supporting the results of this article are included within the article and supplementary files. The RNA-Seq data is available upon reasonable request.

References

1. Angelovici R, Galili G, Fernie AR, Fait A: **Seed desiccation: a bridge between maturation and germination.** *Trends in plant science* 2010, **15**(4):211-218.
2. Leprince O, Hendry GAF, McKersie BD: **Mechanisms of desiccation tolerance in developing seeds.** *Seed Science Research* 1993, **3**(4):231-246.
3. Leprince O, Pellizzaro A, Berri S, Buitink J: **Late seed maturation: drying without dying.** *Journal of experimental botany* 2017, **68**(4):827-841.
4. Shih M-D, Hoekstra FA, Hsing Y-IC: **Late embryogenesis abundant proteins.** In: *Advances in Botanical Research.* vol. 48: Elsevier; 2008: 211-255.
5. Johann G, Abdelilah B, Emeline T, Marie-Hélène AM, Didier G, David M: **Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying.** *Plant Physiology* 2005, **137**(1):157-167.
6. Sutton F, Ding X, Kenefick DG: **Group 3 LEA gene *HVA1* regulation by cold acclimation and deacclimation in two barley cultivars with varying freeze resistance.** *Plant physiology* 1992, **99**(1):338-340.
7. Ukaji N, Kuwabara C, Takezawa D, Arakawa K, Fujikawa S: **Cold acclimation-induced WAP27 localized in endoplasmic reticulum in cortical parenchyma cells of mulberry tree was homologous to group 3 late-embryogenesis abundant proteins.** *Plant physiology* 2001, **126**(4):1588-1597.
8. Huang L, Zhang M, Jia J, Zhao X, Huang X, Ji E, Ni L, Jiang M: **An Atypical Late Embryogenesis Abundant Protein OsLEA5 Plays a Positive Role in ABA-Induced Antioxidant Defense in *Oryza sativa* L.** *Plant & Cell Physiology* 2018, **59**(5):916.
9. Liu Y, Liang J, Sun L, Yang X, Li D: **Group 3 LEA protein, ZmLEA3, is involved in protection from low temperature stress.** *Frontiers in Plant Science* 2016, **7**:1011.

10. NDong C, Danyluk J, Wilson KE, Pocock T, Huner NP, Sarhan F: **Cold-regulated cereal chloroplast late embryogenesis abundant-like proteins. Molecular characterization and functional analyses.** *Plant physiology* 2002, **129**(3):1368-1381.
11. Sharma A, Kumar D, Kumar S, Rampuria S, Reddy AR, Kirti PB: **Ectopic expression of an atypical hydrophobic group 5 LEA protein from wild peanut, *Arachis diogoi* confers abiotic stress tolerance in tobacco.** *PloS one* 2016, **11**(3):e0150609.
12. Xu D, Duan X, Wang B, Hong B, Ho T-HD, Wu R: **Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice.** *Plant physiology* 1996, **110**(1):249-257.
13. Hundertmark M, Hincha DK: **LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*.** *BMC genomics* 2008, **9**(1):118.
14. Bremer A, Wolff M, Thalhammer A, Hincha DK: **Folding of intrinsically disordered plant LEA proteins is driven by glycerol-induced crowding and the presence of membranes.** *The FEBS journal* 2017, **284**(6):919-936.
15. Hatanaka R, Furuki T, Shimizu T, Takezawa D, Kikawada T, Sakurai M, Sugawara Y: **Biochemical and structural characterization of an endoplasmic reticulum-localized late embryogenesis abundant (LEA) protein from the liverwort *Marchantia polymorpha*.** *Biochemical and biophysical research communications* 2014, **454**(4):588-593.
16. Hundertmark M, Dimova R, Lengefeld J, Seckler R, Hincha DK: **The intrinsically disordered late embryogenesis abundant protein LEA18 from *Arabidopsis thaliana* modulates membrane stability through binding and folding.** *Biochimica et Biophysica Acta (BBA)-Biomembranes* 2011, **1808**(1):446-453.
17. Gilles GJ, Hines KM, Manfre AJ, Jr WRM: **A predicted N-terminal helical domain of a Group 1 LEA protein is required for protection of enzyme activity from drying.** *Plant Physiology & Biochemistry* 2007, **45**(6):389-399.
18. Jiang S, Wang J, Liu X, Liu Y, Guo C, Zhang L, Han J, Wu X, Xue D, Gomaa AE: **DrwH, a novel WHy domain-containing hydrophobic LEA5C protein from *Deinococcus radiodurans*, protects enzymatic activity under oxidative stress.** *Scientific reports* 2017, **7**(1):9281.
19. Ried JL, Walker-Simmons MK: **Group 3 Late Embryogenesis Abundant Proteins in Desiccation-Tolerant Seedlings of Wheat (*Triticum aestivum* L.).** *Plant Physiology* 1993, **102**(1):125-131.
20. Warner AH, Miroshnychenko O, Kozarova A, Vacratsis PO, MacRae TH, Kim J, Clegg JS: **Evidence for multiple group 1 late embryogenesis abundant proteins in encysted embryos of *Artemia* and their organelles.** *The Journal of Biochemistry* 2010, **148**(5):581-592.
21. Thomann EB, Sollinger J, White C, Rivin CJ: **Accumulation of group 3 late embryogenesis abundant proteins in *Zea mays* embryos.** *Plant Physiology* 1992, **99**(2):607-614.
22. Manfre AJ, Lanni LM, Marcotte WR: **The *Arabidopsis* group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development.** *Plant physiology* 2006, **140**(1):140-149.

23. Kodahl N, Johansen BB, Rasmussen FN: **The embryo sac of *Vanilla imperialis* (Orchidaceae) is six-nucleate, and double fertilization and formation of endosperm are not observed.** *Botanical journal of the Linnean Society* 2015, **177**(2):202-213.
24. Ling H, Zeng X, Guo S: **Functional insights into the late embryogenesis abundant (LEA) protein family from *Dendrobium officinale* (Orchidaceae) using an *Escherichia coli* system.** *Scientific Reports* 2016, **6**(1):39693.
25. Maniatis T, Tasic B: **Alternative pre-mRNA splicing and proteome expansion in metazoans.** *Nature* 2002, **418**(6894):236-243.
26. De Conti L, Baralle M, Buratti E: **Exon and intron definition in pre-mRNA splicing.** *Wiley Interdisciplinary Reviews: RNA* 2013, **4**(1):49-60.
27. Lee YI, Hsu ST, Yeung EC: **Orchid protocorm-like bodies are somatic embryos.** *American Journal of Botany* 2013, **100**(11):2121-2131.
28. Ikeuchi M, Iwase A, Rymen B, Lambolez A, Kojima M, Takebayashi Y, Heyman J, Watanabe S, Seo M, De Veylder L: **Wounding triggers callus formation via dynamic hormonal and transcriptional changes.** *Plant physiology* 2017, **175**(3):1158-1174.
29. Iwase A, Mitsuda N, Koyama T, Hiratsu K, Kojima M, Arai T, Inoue Y, Seki M, Sakakibara H, Sugimoto K: **The AP2/ERF transcription factor WIND1 controls cell dedifferentiation in *Arabidopsis*.** *Current Biology* 2011, **21**(6):508-514.
30. Artur MAS, Zhao T, Ligterink W, Schranz E, Hilhorst HW: **Dissecting the genomic diversification of LATE EMBRYOGENESIS ABUNDANT (LEA) protein gene families in plants.** *Genome biology and evolution* 2018, **11**(2):459-471.
31. Du D, Zhang Q, Cheng T, Pan H, Yang W, Sun L: **Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*.** *Molecular biology reports* 2013, **40**(2):1937-1946.
32. Liang Y, Xiong Z, Zheng J, Xu D, Zhu Z, Xiang J, Gan J, Raboanatahiry N, Yin Y, Li M: **Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*.** *Sci Rep* 2016, **6**:24265.
33. Nagaraju M, Kumar SA, Reddy PS, Kumar A, Rao DM, Kavi Kishor PB: **Genome-scale identification, classification, and tissue specific expression analysis of late embryogenesis abundant (LEA) genes under abiotic stress conditions in *Sorghum bicolor*L.** *PLoS One* 2019, **14**(1):e0209980.
34. Gao J, Lan T: **Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabuliformis* (Pinaceae) in *Escherichia coli*.** *Scientific reports* 2016, **6**:19467.
35. Zhang G-Q, Xu Q, Bian C, Tsai W-C, Yeh C-M, Liu K-W, Yoshida K, Zhang L-S, Chang S-B, Chen F: **The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution.** *Scientific reports* 2016, **6**:19029.
36. Zhang GQ, Liu KW, Li Z, Lohaus R, Hsiao YY, Niu SC, Wang JY, Lin YC, Xu Q, Chen LJ: **The *Apostasia* genome and the evolution of orchids.** *Nature* 2017, **549**(7672):379.

37. Wang W, Gao T, Chen J, Yang J, Huang H, Yu Y: **The late embryogenesis abundant gene family in tea plant (*Camellia sinensis*): Genome-wide characterization and expression analysis in response to cold and dehydration stress.** *Plant Physiology and Biochemistry* 2019, **135**:277-286.
38. Lv A, Su L, Liu X, Xing Q, Huang B, An Y, Zhou P: **Characterization of Dehydrin protein, CdDHN4-L and CdDHN4-S, and their differential protective roles against abiotic stress in vitro.** *BMC plant biology* 2018, **18**(1):299.
39. Dure L, Greenway SC, Galau GA: **Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis.** *Biochemistry* 1981, **20**(14):4162-4168.
40. Galau GA, Hughes DW, Dure L: **Abscisic acid induction of cloned cotton late embryogenesis-abundant (Lea) mRNAs.** *Plant molecular biology* 1986, **7**(3):155-170.
41. Teixeira da Silva JA, Aceto S, Liu W, Yu H, Kanno A: **Genetic control of flower development, color and senescence of *Dendrobium* orchids.** *Scientia Horticulturae* 2014, **175**(1):74-86.
42. Ishii Y, Takamura T, Goi M, Tanaka M: **Callus induction and somatic embryogenesis of *Phalaenopsis*.** *Plant Cell Reports* 1998, **17**(6-7):446-450.
43. Pedrosa AM, Martins CdPS, Gonçalves LP, Costa MGC: **Late embryogenesis abundant (LEA) constitutes a large and diverse family of proteins involved in development and abiotic stress responses in sweet orange (*Citrus sinensis* L. Osb.).** *PLoS One* 2015, **10**(12):e0145785.
44. Mishra S, Triptahi V, Singh S, Phukan UJ, Gupta M, Shanker K, Shukla RK: **Wound induced transcriptional regulation of benzylisoquinoline pathway and characterization of wound inducible PsWRKY transcription factor from *Papaver somniferum*.** *PLoS One* 2013, **8**(1):e52784.
45. Dolzblasz A, Nardmann J, Clerici E, Causier B, van der Graaff E, Chen J, Davies B, Werr W, Laux T: **Stem cell regulation by *Arabidopsis* WOX genes.** *Molecular plant* 2016, **9**(7):1028-1039.
46. Kamiya N, Nagasaki H, Morikami A, Sato Y, Matsuoka M: **Isolation and characterization of a rice WUSCHEL-type homeobox gene that is specifically expressed in the central cells of a quiescent center in the root apical meristem.** *The Plant Journal* 2003, **35**(4):429-441.
47. Mayer KF, Schoof H, Haecker A, Lenhard M, Jürgens G, Laux T: **Role of WUSCHEL in regulating stem cell fate in the *Arabidopsis* shoot meristem.** *Cell* 1998, **95**(6):805-815.
48. van der Graaff E, Laux T, Rensing SA: **The WUS homeobox-containing (WOX) protein family.** *Genome biology* 2009, **10**(12):248.
49. Murashige T, Skoog F: **A revised medium for rapid growth and bio-assays with tobacco tissue cultures.** *Plant Physiology* 1962, **15**(3):473-497.
50. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A: **The Pfam protein families database in 2019.** *Nucleic acids research* 2018, **47**(D1):D427-D432.
51. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server.** *Bioinformatics* 2014, **31**(8):1296-1297.

52. Kazutaka K, Standley DM: **MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.** *Molecular Biology & Evolution* 2013, **30**(4):772-780.
53. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Molecular Biology and Evolution* 2016, **33**(7):1870-1874.
54. Weigel D, Glazebrook J: **Transformation of agrobacterium using the freeze-thaw method.** *CSH protocols* 2006, **2006**(7):1031-1036.
55. Clough SJ, Bent AF: **Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*.** *The plant journal* 1998, **16**(6):735-743.
56. He C, Yu Z, Teixeira da Silva JA, Zhang J, Liu X, Wang X, Zhang X, Zeng S, Wu K, Tan J et al: **DoGMP1 from *Dendrobium officinale* contributes to mannose content of water-soluble polysaccharides and plays a role in salt stress response.** *Scientific reports* 2017, **7**:41010-41010.
57. Niu S-C, Xu Q, Zhang G-Q, Zhang Y-Q, Tsai W-C, Hsu J-L, Liang C-K, Luo Y-B, Liu Z-J: **De novo transcriptome assembly databases for the butterfly orchid *Phalaenopsis equestris*.** *Scientific data* 2016, **3**:160083.
58. Anders S, Pyl PT, Huber W: **HTSeq—a Python framework to work with high-throughput sequencing data.** *Bioinformatics* 2015, **31**(2):166-169.
59. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: **PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.** *Nucleic Acids Res* 2002, **30**(1):325-327.
60. Higo K, Ugawa Y, Iwamoto M, Korenaga T: **Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999.** *Nucleic Acids Research* 1999, **27**(1):297-300.

Additional File

Additional file 1 table S1 Information of LEA genes in *Arabidopsis*, *P. equestris* and *D. officinale*.

Additional file 2 table 2 The pair specific primers used for Semi-quantitative RT-PCR and qPCR.

Figures

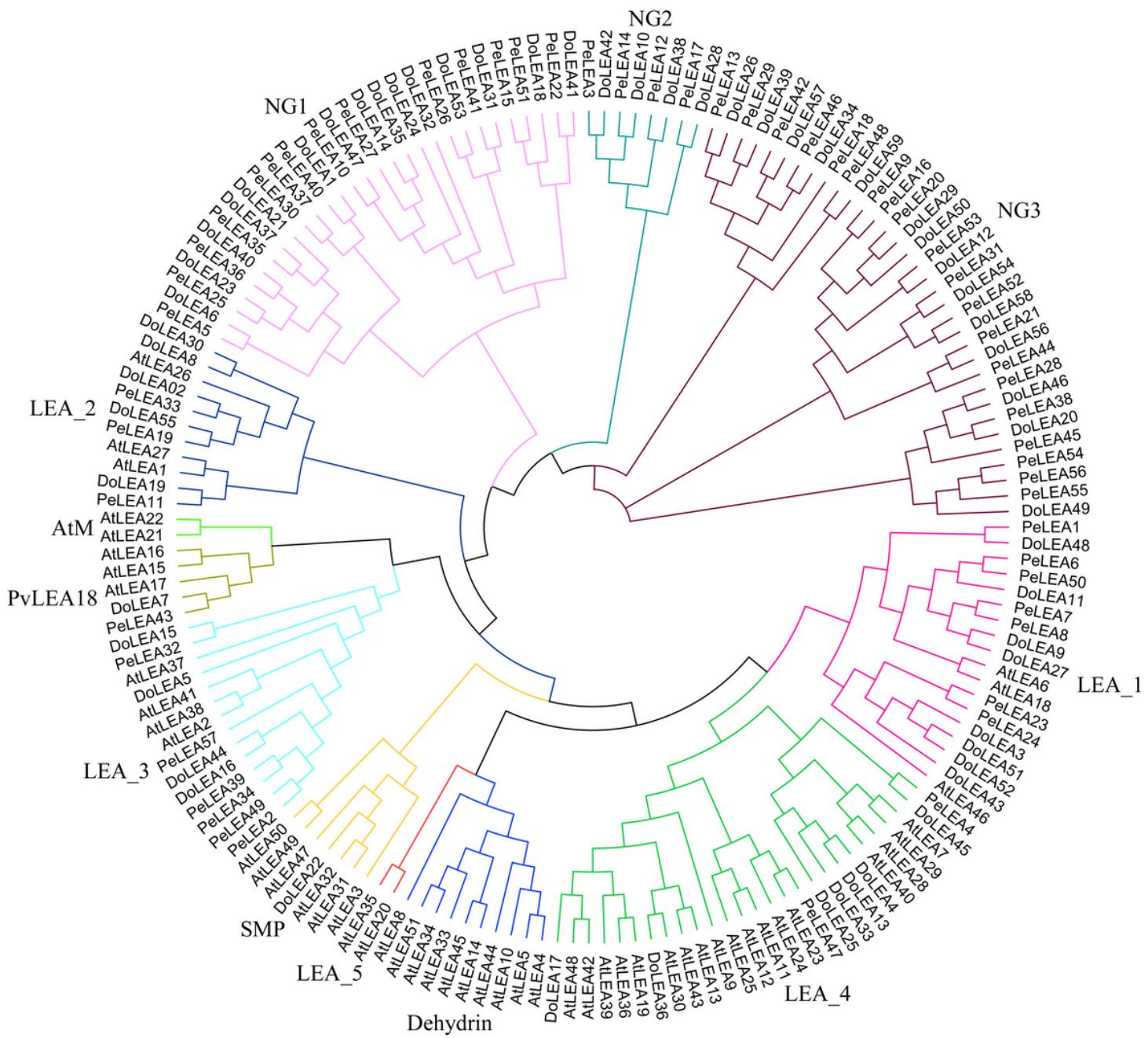


Figure 1

Molecular phylogenetic tree of all the LEA proteins from *Arabidopsis*, *P. equestris* and *D. officinale*. The tree was constructed using MEGA version 7 by the Neighbor-Joining method based on alignment by MAFFT.

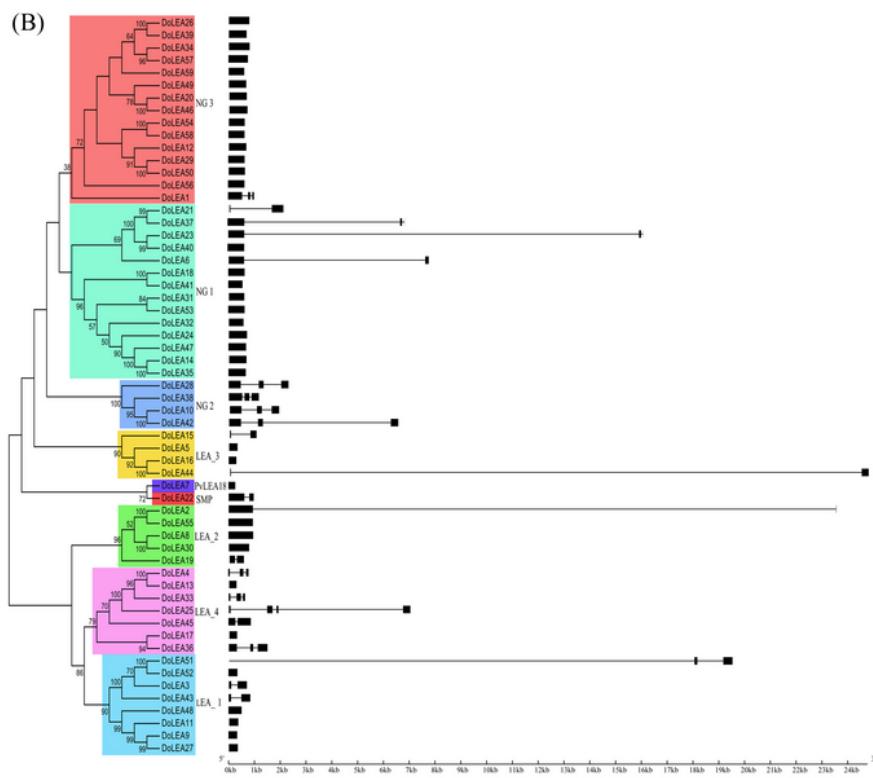
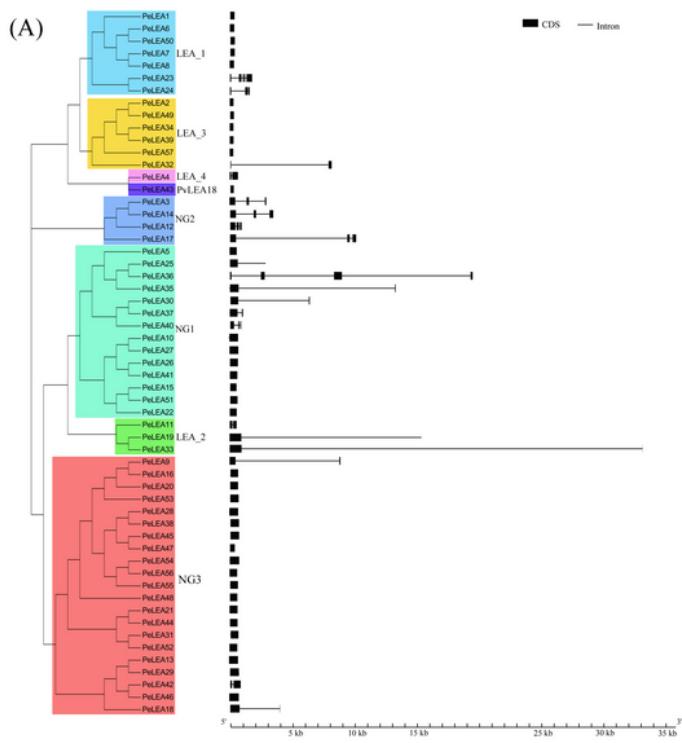


Figure 2

Phylogenetic analysis and gene structures of LEA genes in *P. equestris* (A) and *D. officinale* (B). The phylogenetic tree was constructed by MEGA version 7 with the Neighbor-Joining method and 1000 bootstrap replicates method based on alignments of complete LEA protein sequences by MAFFT. The gene structure diagram was generated by Gene Structure Display Server (GSDS) 2.0 using the GFF annotation of each LEA gene.

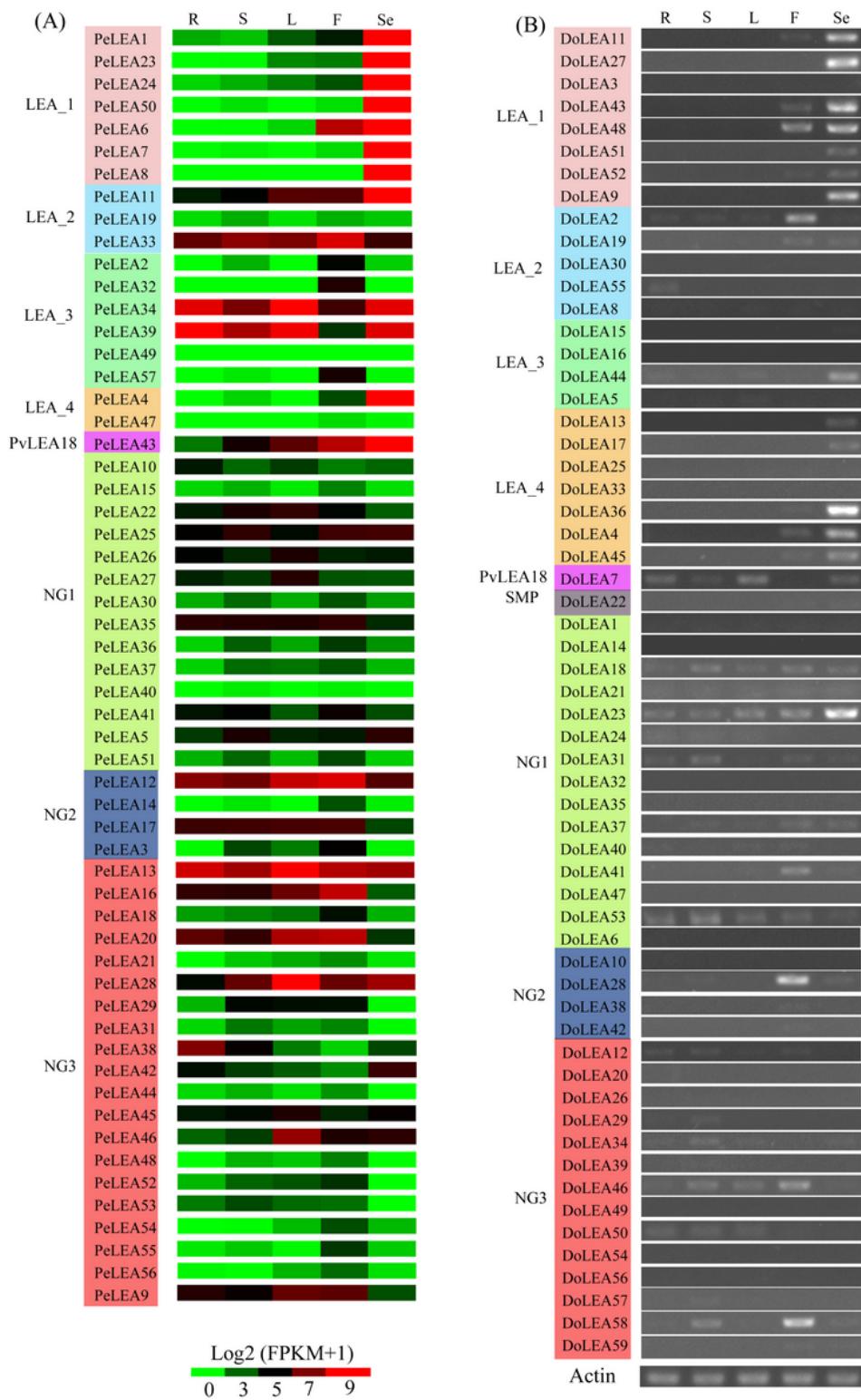


Figure 3

Expression patterns of LEA genes from *P. equestris* (A) and *D. officinale* (B) among roots, stems, leaves, flowers and seed. Concentration gradient heat maps generated using the gene expression level [log2 (FPKM+1)] of each organ. Red and green in concentration gradient heat maps indicate high and low gene expression level, respectively.

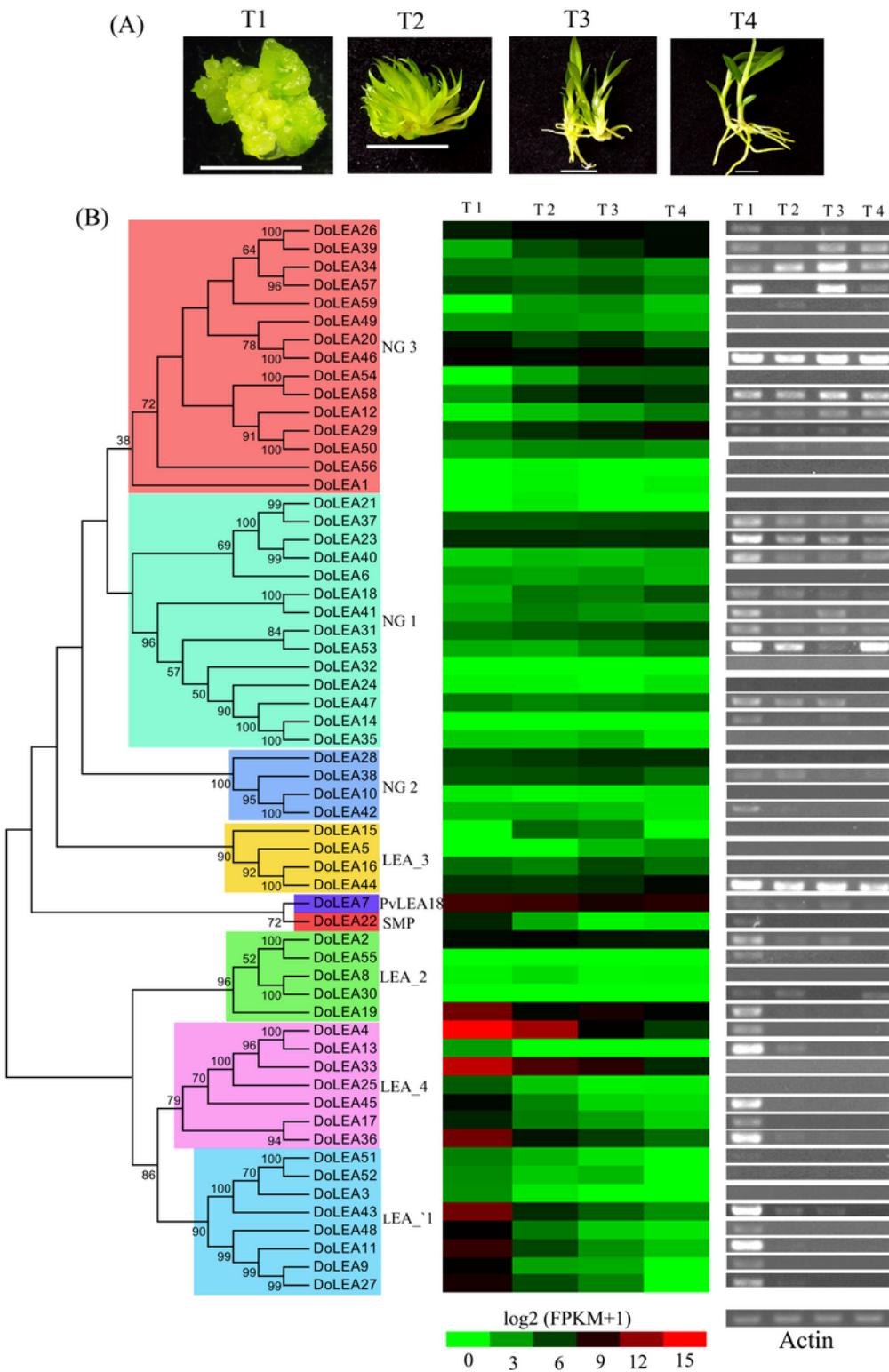


Figure 4

Expression patterns of LEA genes from *D. officinale* at different developmental stages. (A) The four developmental stages (see details in materials and methods) of *D. officinale* using to explore the DoLEAs. Bar = 1 cm. (B) The expression of DoLEAs in T1-T4 stages by heat map (left) and semi-quantitative RT-PCR (right) analysis.

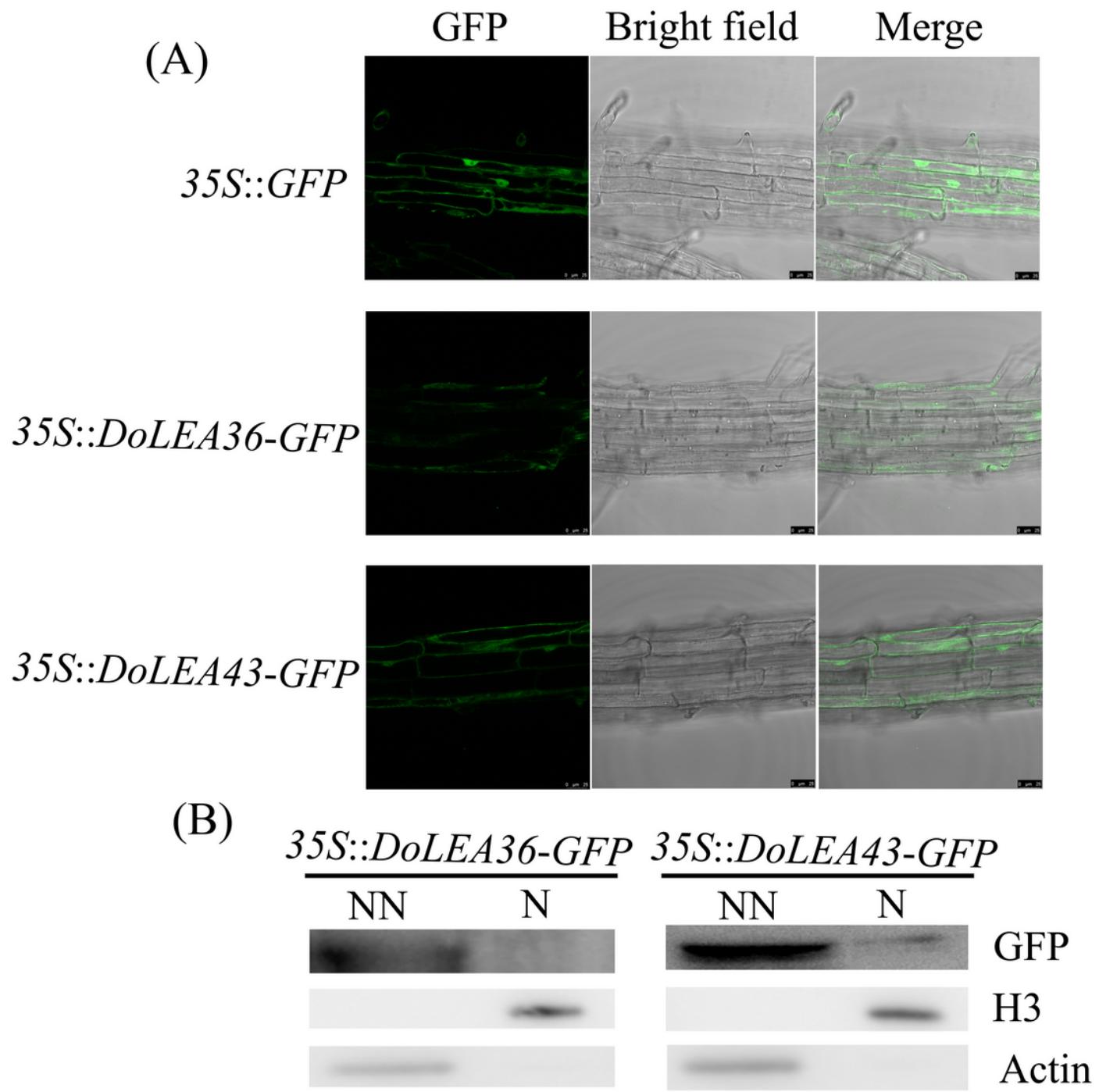


Figure 5

Localization of DoLEA36 and DoLEA43 proteins. (A) Fluorescent microscopic images of GFP and DoLEA36-GFP and DoLEA43-GFP fusion proteins in the roots of transgenic lines, respectively (bar = 25 μ m). (B) Localization of DoLEA36 and DoLEA43 by Western blot analysis. Nuclear (N) and non-nuclear (NN) proteins were extracted from *35S::DoLEA36-GFP* and *35S::DoLEA43-GFP* transgenic seedlings grown for 7 d, respectively, and subjected to immunoblot analysis with anti-GFP antibody. Histone H3 and actin were detected as nuclear and non-nuclear protein controls, respectively.

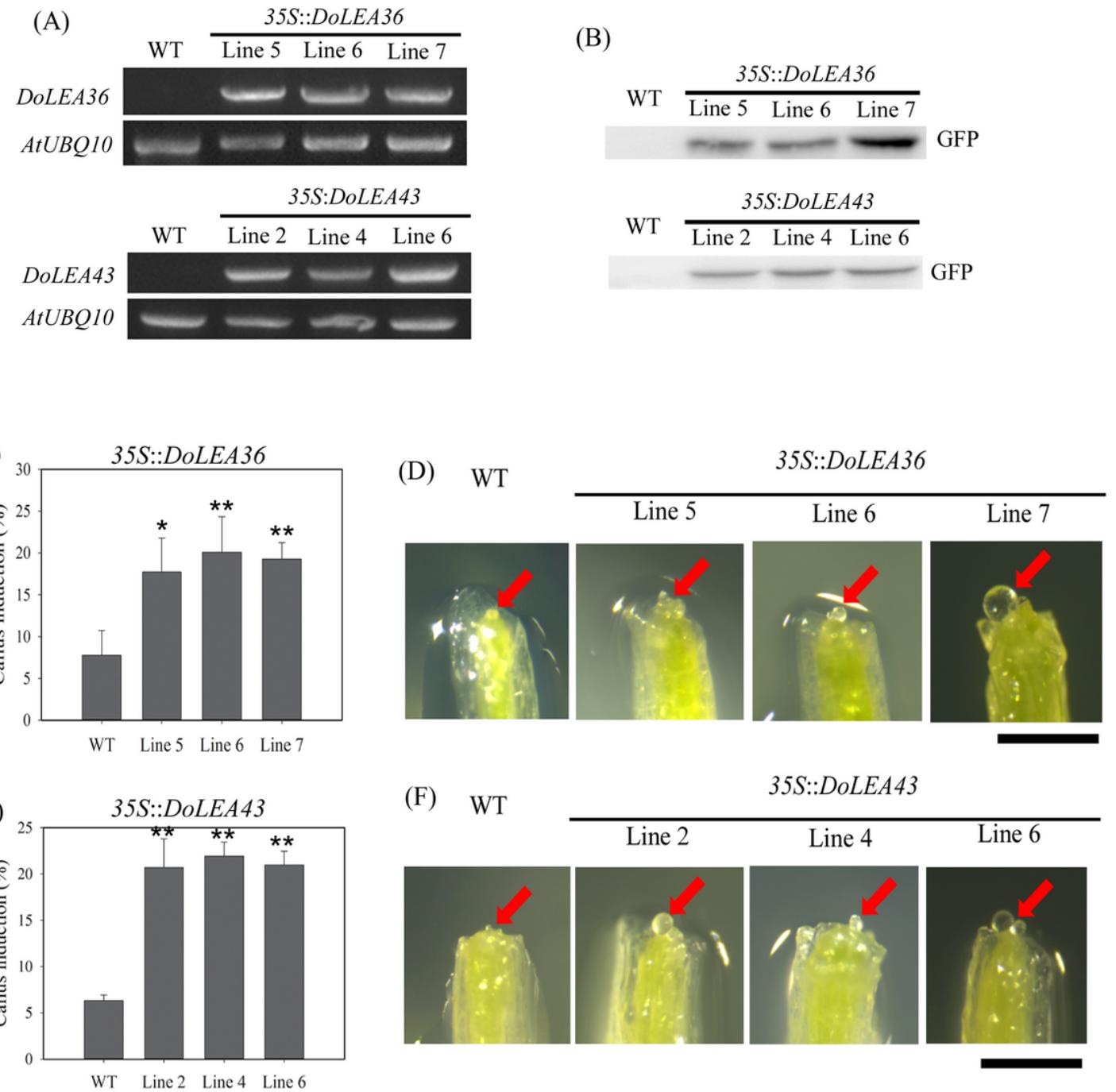


Figure 6

The *DoLEA36* and *DoLEA43* genes display a positive role in callus initiation at 14 d after wounding. (A) Analysis the expression of *DoLEA36* and *DoLEA43* among 35S::*DoLEA36*-GFP and 35S::*DoLEA43*-GFP transgenic seedlings, respectively, by semi-quantitative RT-PCR. (B) Western blot analysis the *DoLEA36* and *DoLEA43* protein among 35S::*DoLEA36*-GFP and 35S::*DoLEA43*-GFP transgenic seedlings, respectively. (C) The 35S::*DoLEA36*-GFP lines show a higher callus induction than WT plants. (D) Representative images of callus initiation in WT and 35S::*DoLEA36*-GFP lines. (E) The 35S::*DoLEA43*-GFP lines showed increased callus initiation phenotypes at 14 d after wounding. (F) Representative images of callus initiation in WT and 35S::*DoLEA43*-GFP lines. Scale bar = 0.05 cm. Hypocotyl explants

were cut at about 0.5 mm from the cotyledon-hypocotyl junction and at about 0.5 mm from the hypocotyl-root junction. The middle of hypocotyls was used to analyze the callus induction. Bars represent mean \pm SD of three biological replicates. ** and * indicate P < 0.001 and P < 0.05 between WT and transgenic lines, respectively. WT indicates wild type.

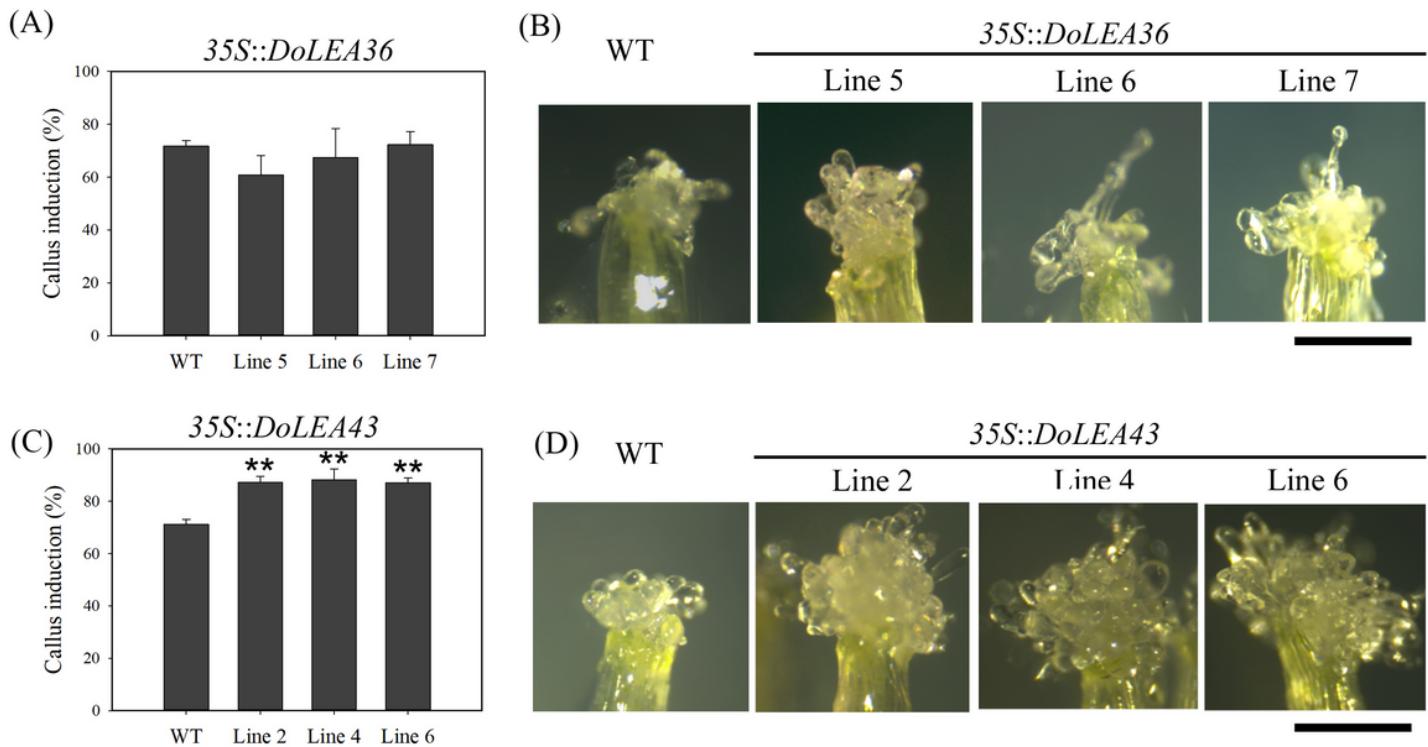


Figure 7

The different role of DoLEA36 and DoLEA43 genes in callus initiation at 12 d after wounding under dark. (A) No significant difference of callus induction in the *35S::DoLEA36*-GFP lines and WT plants. (B) Representative images of callus initiation in WT and *35S::DoLEA36*-GFP lines. (C) The *35S::DoLEA43*-GFP lines showed increased callus initiation phenotypes at 12 d after wounding. (D) Representative images of callus initiation in WT and *35S::DoLEA43*-GFP lines. Scale bar = 0.05 cm. The hypocotyls were cut once at about 2 mm from the hypocotyl-root junction, the hypocotyl-root explants were used to study the callus induction. Bars represent mean \pm SD of three biological replicates. ** indicate P < 0.001 between WT and transgenic lines, respectively.

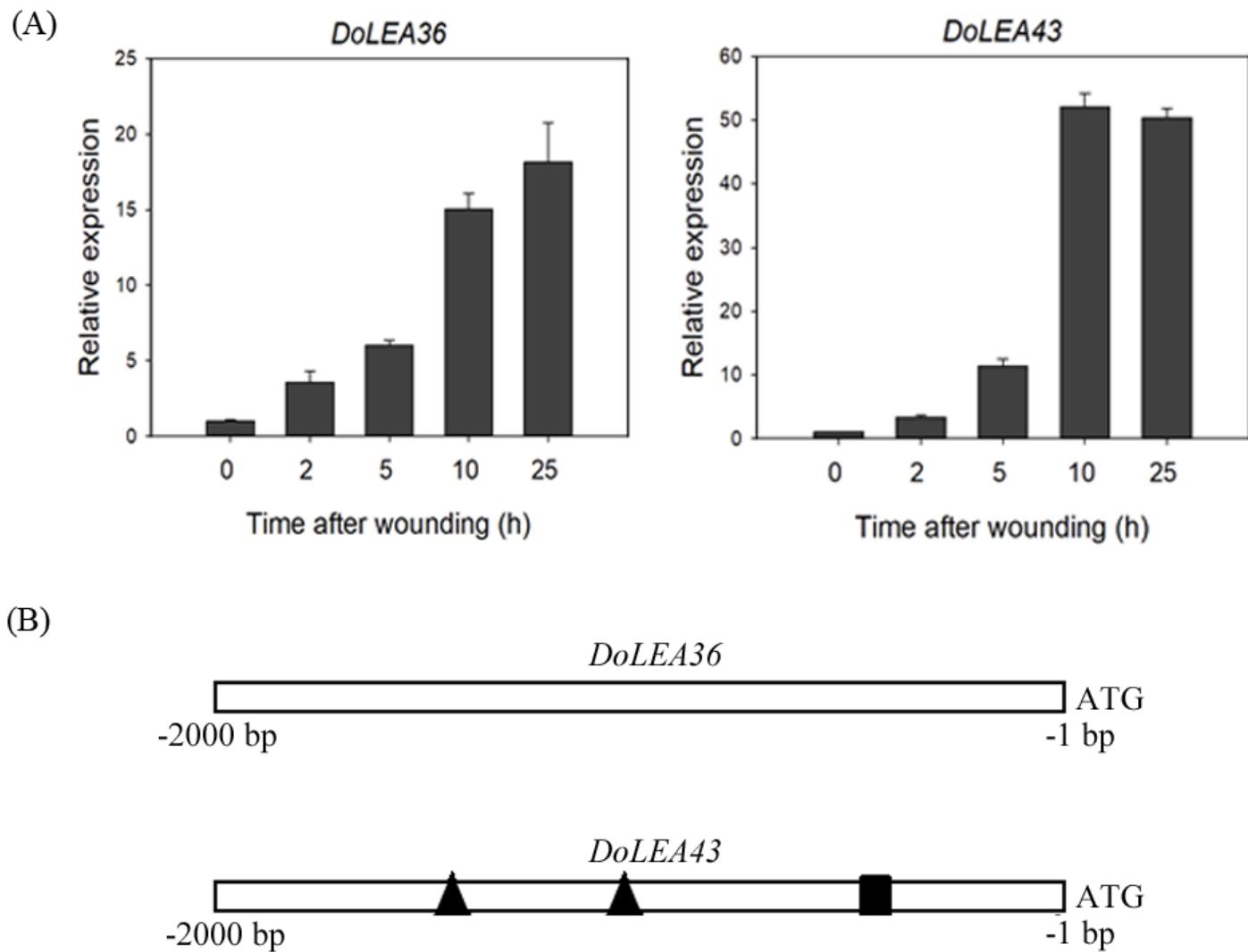


Figure 8

(A) The expression of DoLEA36 and DoLEA43 genes at 0, 2, 5, 10 and 25 h after wounding. (B) Prediction of WUN-motif and WUSATAg motif in the 2-k upstream regulatory regions of DoLEA36 and DoLEA43 genes. Triangle indicates the WUSATAg motif, rectangle indicates WUN-motif.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.pdf