LEA家族基因参与仁用杏花期调控的分子机制

李少锋[1*]， 乌云塔娜[2]，王琳[2]，朱高浦[2]，夏永秀[1]，王少丽[1]，王飞[1]，张耀丹[1]

（1. 中国林业科学研究院华北林业实验中心 北京 102300； 2. 中国林业科学研究院经济林研究所 河南郑州 450003）

摘 要:【目的】仁用杏具有很高的经济价值和生态价值，是我国“三北地区”重要的生态经济型干果树种、木本油料树种和植物蛋白饮料树种。由于仁用杏休眠期短、花期早，易遭受晚霜为害，造成减产甚至绝产，因此了解低温条件下仁用杏花期调控的响应机制，实现花期的人为调控是解决仁用杏低产和不稳产的有效途径之一。【方法】该研究以拟南芥 AtLEAs 和杨树 PtLEAs 序列为种子序列，通过 blast 在仁用杏“龙王帽”基因组中鉴定胚胎发育晚期丰富蛋白（Late embryogenesis abundant，LEA 蛋白）家族成员，并对仁用杏、拟南芥和杨树基因组内的 LEA 家族成员开展了系统发育分析和共线性分析，对 PasLEAs 家族基因的染色体分布、基因复制事件、保守 Motif、顺式作用元件进行分析。在对 PasLEAs 家族基因在不同组织（花芽和茎段）和不同休眠、萌发时期（生理休眠、生态休眠、发芽期和萌发期）进行表达模式分析的基础上，开展了具有代表性的家族成员的亚细胞定位和基因共表达网络分析，并进行了低温胁迫下 *PasLEA3-2* 和 *Pasdehydrin-3* 转酵母和拟南芥的表型和基因表达研究，以及通过酵母单、双杂交、BiFC 等技术挖掘 *PasLEA3-2* 和 *Pasdehydrin-3* 的互作信息和调控机制。【结果】在仁用杏“龙王帽”基因组中共鉴定到 54 个 PasLEAs 成员，发现 14 个 LEA 共线性模块存在于拟南芥和拟南芥，9 个共线性模块在拟南芥和杨树之间被鉴定到。全基因组复制（WGD）、片段复制和串联复制可能在 LEA 家族的基因复制和加倍中起重要作用。PasLEAs 可与参与低温胁迫或休眠、休眠解除相关的 *PasEREBP*、*PasSer/Thr*、*PasWRKY* 等基因协同作用，调控仁用杏抗寒性、芽休眠和芽萌发。酵母双杂交和 BiFC 研究发现， PasLEA1-2 与 PasERD4，PasLEA3-1 与 PasXTH9，PasSMP-3 与 PasEXPA22，PasDehydrin-3 与 PasCBF1 之间有蛋白相互作用。*PasLEA3-2* 和 *Pasdehydrin-3* 转化的酵母菌在低温条件（-20 °C）下能正常生长。在低温胁迫下，*PasLEA3-2* 和 *Pasdehydrin-3* 转基因拟南芥的存活率、根长、叶长、叶宽、果荚长度、种子结实率相比对照均提高，开花时间和抽薹时间相比对照均提前，SOD 活性和脯氨酸含量显著高于对照，MDA 含量明显比对照低，并且花发育相关基因 *AtAP1*, *AtBTF3L*, *AtEBF1-1, AtLFY, AtPUB26*, *AtGl* 和 *AtRAV1* 的表达相比对照明显提高。酵母单杂交研究表明，PasGBF4 和 PasDOF3.5 能转录调控 *PasLEA3-2* 的表达，PtoBLH8 和 PtoWRKY40 能转录调控 *Pasdehydrin-3* 的表达。【结论】上述研究表明，*PasLEA3-2* 和 *Pasdehydrin-3* 过表达的转基因拟南芥抗寒性明显提高，并且 *PasLEA3-2* 和 *Pasdehydrin-3* 的过表达促进了转基因拟南芥提前开花、提高了转基因植株的种子结实率，有利于低温下优异种质资源的保存和后代的繁衍。这些研究为了解仁用杏花期调控的低温响应机制、获得抗寒育种的基因资源以及花期的精准调控奠定良好基础。

关键词: 仁用杏；LEA 蛋白；抗寒性；花芽休眠；作用机制

Genome-wide analysis and functional analysis of late embryogenesis abundant (LEA) genes during dormancy and sprouting periods of kernel-using apricots (*P. armeniaca* L. × *P. sibirica* L.)

Shaofeng Li [1*], Ta-na Wuyun [2], Lin Wang [2], Gaopu Zhu [2], Yongxiu Xia [1], Shaoli Wang [1], Yaodan Zhang [1]

([1] State Key Laboratory of Tree Genetics and Breeding, Experimental Center of Forestry in North China, National Permanent Scientific Research Base for Warm Temperate Zone Forestry of Jiulong Mountain in Beijing, Chinese Academy of Forestry, Beijing 100091, P.R. China; [2] State Key Laboratory of Tree Genetics and Breeding, Non-timber Forestry Research and Development Center, Chinese Academy of Forestry, Zhengzhou 450003, P.R. China;)

Abstract:【Objective】The aim of this study was to Understand the response mechanism of regulating the flowering period of kernel-using apricots (*P. armeniaca* L. × *P. sibirica* L.) under low temperature and achieving artificial regulation of flowering period

基金项目：中国林科院院基金（CAFYBB2018MA003、CAFYBB2020QD001）、国家自然科学基金项目（31770705、31400570）

* 李少锋为通讯作者

is one of the effective ways to solve the problem of low and unstable yield of kernel apricot.【Method】The predicted peptide sequences of LEA members were acquired from the genome database (unpublished) of kernel-using apricot (*P. armeniaca* L. × *P. sibirica* L.), to construct a local protein database. BLASTP searches were performed using LEA protein sequences of *A. thaliana* and poplar as queries. Phylogenetic analysis and collinearity analysis were carried out for LEA family members in the genomes of *P. armeniaca* ×. *sibirica*, *A. thaliana*, and *P. trichocarpa* , and the chromosome distribution, gene replication events, conservative Motif, and Cis-regulatory element of PasLEAs family genes were analyzed. Based on the analysis of expression patterns of *PasLEAs* genes in different tissues (flower buds and stems) and different dormancy and germination periods (Physiological dormancy (PD), ecological dormancy (ED), sprouting period (SP), and germination stage (GS)), subcellular localization and Gene co-expression network analysis of important family members were carried out, and the phenotype and gene expression of *PasLEAs* transgenic yeast and *Arabidopsis* under low temperature stress, as well as Yeast two-hybrid (Y2H) assay, Yeast one-hybrid (Y1H) assay, Bimolecular fluorescence complementation assay (BiFC) were used to explored the interaction information and regulatory mechanisms of *PasLEA3-2* and *Pasdehydrin-3*. 【Result】 In this study, 54 *LEA* members were genome-wide identified from the kernel-using apricots (*P. armeniaca* L. × *P. sibirica* L.) genome and clustered into eight subgroups according to phylogenetic relationship. A total of 14 collinear blocks containing *LEA* genes between *P. armeniaca* ×*P. sibirica* and *Arabidopsis thaliana* were observed by synteny analysis, and 9 collinear blocks were identified between *P. armeniaca* and poplar. Gene structure and conserved motif composition analyses showed that a conserved exon-intron pattern and conserved motifs are exist within members of each subgroup. Whole-genome duplication (WGD), segmental duplication and tandem duplication are likely played an vitrol role in the gene cluster and duplication of the LEA family in *P. armeniaca* × *P. sibirica*. Expression analysis using both RNA-seq data and quantitative real-time RT-PCR (qRT-PCR) showed that *PasLEA2-20*, *PasLEA3-2*, *PasLEA6-1, Pasdehydrin-3* and *Pasdehydrin-5* were induced in flower buds during the dormancy and sprouting periods of *P. armeniaca* ×*P. sibirica*. A *PasLEA* coexpression network showed that *PasLEAs* were coexpressed with 15 cold resistance genes or bud dormancy/dormancy release related genes such as *PasEREBP*, *Pas*Ser/Thr and *PasWRKY*. The expression pattern of all of the *PasLEAs* coexpressed with cold stress-associated genes was consistent with the four periods (Physiological dormancy (PD), ecological dormancy (ED), sprouting period (SP), and germination stage (GS)) of flower buds of *P. armeniaca* ×*P. sibirica*. An *A. thaliana* ortholog-based protein–protein interaction network was constructed, combined yeast two-hybrid (Y2H) and BiFC experiments revealed that PasLEA proteins had strong connections with the cold resistance pathway, such as PasLEA1-2 could interact with PasERD4, PasLEA3-1 could interact with PasXTH9, PasSMP-3 could interact with PasEXPA22, and PasDehydrin-3 could interact with PasCBF1. Remarkably, overexpression of *PasLEA3-2* and *Pasdehydrin-3* accerated the protein accumulation in yeast in low temperature. *PasLEA3-2* and *Pasdehydrin-3* overexpressed in *Arabidopsis thaliana* in cold stress further revealed extended pod length, shortened the bolting time and flowering time, improved survival rate and seed setting rate, and increased accumulation of proline and antioxidative enzyme activities, and a higher expression of flower development associated genes such as *AtAP1*, *AtBTF3L*, *AtEBF1-1, AtLFY, AtPUB26*, *AtGl* and *AtRAV1* compared with wild-type plants. Yeast one-hybrid (Y1H) assay confirmed PasGBF4 and PasDOF3.5 can serve as upstream regulatory factors bind to the promoter of *PasLEA3-2*, and PasDOF2.4, PasDnaJ2 and PasAP2 can bind to *Pasdehydrin-3* promoter to regulate the expression of downstream genes. 【Conclusion】 Overall, this study presents a comprehensive overview of evolutionary relationship, protein interaction, and functional analysis of *PasLEA* genes during dormancy and sprouting periods of *P. armeniaca* × *P. sibirica*, and identifies candidate genes for improving cold resistance and manipulated flower bud dormancy in *P. armeniaca* × *P. sibirica*.

Key words: kernel-using apricots (*P. armeniaca* L. × *P. sibirica* L.); Late embryogenesis abundant (LEA) protein; cold resistance; flower bud dormancy; regulation mechanism

Introduction

Late embryogenesis abundant (LEA) proteins were first observed to accumulate during the late embryogenesis in cotton seed (Galau et al. 1986). Subsequently, they have been identified in angiosperms and gymnosperms, including *Arabidopsis* (Hundertmark et al. 2008), rice (Wang et al. 2007), sunflower (Prietodapena et al. 1999), wheat (Sasaki et al. 2014), maize (Amara et al. 2012), *Brassica napus* (Liang et al. 2016), barley (Choi et al. 1999), and *Pinus tabuliformis* (Gao et al. 2016). Additionally, LEA proteins also exist in lower plants

such as algae (Velten and Oliver 2001; Campos et al. 2006), and other species such as fungi, nematodes and chironomids (Gal et al. 2014; Kikawada et al. 2006). Subcellular localization analysis has revealed that LEA proteins are mainly located in cytoplasm parts and nuclear regions, such as chloroplast (Ndong et al.,2002) and mitochondria (Grelet et al., 2005; Tolleter et al., 2010), which may beneficial to protect the cell stability under stress environment (Hundertmark et al., 2008; Avelange-Macherel et al., 2014). LEA proteins are detected in various plant organs or tissues and observed in the seeds, buds, roots and seedlings of plants (Hong-Bo et al., 2005; Hundertmark et al., 2008). LEA proteins are mostly consist of hydrophilic amino acids ordered by repetitive sequence, forming a highly hydrophilic structure and protect enzymes and biomolecules from induced aggregation by space-filling of LEA proteins, which can reduce the collision rate between accumulated proteins (Campos et al. 2006; Chakrabortee et al. 2012). LEA proteins would help the the isolation of metal ions and calcium, and play a vital role in signaling transduction in plants (Krüger et al., 2012). LEA proteins may promote the formation of the glassy state, contributing to sugar accumulation in the cytoplasm of plants during conditions of drought stress (Shimizu et al., 2010).

LEA proteins contain high proportion of glycine or alanine, serine and threonine amino acids, and lack or contain small amounts of tryptophan and cysteine residues (Campos et al. 2006), and are characterized by conserved repeat motifs and structure arrangement rules (Hundertmark et al. 2008). Based on the eight conserved Pfams (PF03168, PF04927, PF10714, PF02987, PF03760, PF03242, PF00477, PF00257), LEA proteins can be classified into at least eight families: LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, seed maturation protein (SMP) and dehydrin (Bateman et al. 2010). The genome-wide characterization of LEA genes has been performed in various plant species, such as 51 LEAs in *Arabidopsis* (Hundertmark et al. 2008), 53 LEAs in *Populus trichocarpa* (Lan et al. 2013), 36 LEAs in *Brachypodium distachyon* (Filiz et al. 2013), 39 LEAs in *O. sativa* (Wang et al. 2007), 23 LEAs in Moso Bamboo (Huang et al. 2016), and 108 LEAs in *Brassica napus* (Liang et al. 2016). However, less is known about this family in the important economic forest trees kernel-using apricots (*P. armeniaca* × *P. sibirica*).

Kernel-using apricot (*P. armeniaca* L. × *P. sibirica* L., also considered by some researchers to be *P. armeniaca* L., syn. *Armeniaca vulgaris* Lam.) is an economic trees species with ecological and economic value in the northern China, northwestern China, or northeastern China. Because of its wide adaptability, strong drought resistance, and windbreak and sand fixation, it is one of the preferred pioneer tree species used for ecological protection in shallow mountain, sandy lands and hilly areas, especially in the western regions of China. It is one of the six major nut-producing species in the world and is an emerging species for the production of woody grain and oil, and protein beverages. It can also contribute to the promoting rural revitalization and economic development of poverty-stricken mountainous areas in China. The first high-quality apricot plant (Chuanzhihong) genome has been sequenced and layed a solid foundation for research on the important agronomic characteristics and molecular breeding of apricot (Jiang et al., 2019). Owing to its short dormancy period, early-flowering period,and vulnerable to late frost, kernel-using apricots (*P. armeniaca* L. × *P. sibirica* L.) easily suffers from late spring cold damage, resulting in reduced or even no production. One effective approach to solve the scientific problem of a low and unstable yield of kernel-using apricot is to identifify effective genes involved in flower bud dormancy, bud break and cold resistance, analysis gene expression patterns during dormancy and germination periods, elucidate the relevant physiological and molecular mechanisms and determine how the dormancing and flowering period can be artificially manipulated.

Physiological dormancy (PD) (which is often called endodormancy) is defined as the period in which buds will not burst even under optimal environmental conditions, while ecological dormancy (ED) (which is also called ecodormancy) is the stage in which buds still dormant and cannot burst due to environmental factor conditions are not met (especially the temperature requirement accumulation for flower buds in ED periods is not enough).

Endodormant buds, which cannot burst under preferred conditions because of limitation of internal physiological factors (such as have a hard seed coat with poor permeability, embryos that are not mature physiologically, or contain substances which inhibit bud sprout) (Mitrakos and Diamantoglou, 2006, Gao et al., 2020), transition into the ecodormant (ecologically dormant) period and acquire ability of bud germination after exposure to low temperature for a certain period of time. When accumulated temperature required for flower bud germination in ecologically dormant periods is enough (Mazzitelli et al., 2007, Rohde and Bhalerao, 2007), the flower buds enter the sprouting periods (SP) and germination stage (GS), showing growth signs and bud break, after which they can fully develop, germinate and bloom.

Cold stress can cause chilling injury in late spring, which obviously limits plant development and influences fruit yield in economic forests. LEA genes have been observed to be induced by abiotic stresses, such as dehydration, cold, salinity and abscisic acid stress treatment, and their overexpression in transgenic plants has improved the resistance of transgenic plants to such abiotic stresses (Skriver and Mundy 1990; Grelet et al. 2005). Wheat dehydrin gene WCOR410 is involved in protection from low temperature stress, and the freezing resistance of transgenic strawberry leaves was enhanced by overexpression of WCOR410 (Houde et al. 2004). The cold tolerance of transgenic tobacco was improved by the overexpression of citrus LEA gene CuCOR19 (Hara et al. 2003). The freezing resistance of *Arabidopsis* was enhanced by the ectopic expression of the wheat WCS19 gene (NDong et al. 2002), and overexpression of several LEA genes improved freezing tolerance in transgenic Arabidopsis (Puhakainen et al. 2004). The overexpression of the barley LEA gene (HVA1) in wheat and rice conferred drought and salt tolerance to transgenic plants (Sivamani et al. 2000). Maize group 3 LEA protein ZmLEA3 could enhance the transgenic tobacco resistance to low temperature, oxidative and drought stresses (Liu et al. 2016; Liu et al. 2013).

In this study, we identified 54 candidate LEA proteins in the genome sequence of kernel-using apricots (*P. armeniaca* × *P. sibirica*), and the phylogenetic tree relationships, evolution, chromosomal positions, gene structure, conserved regulatory elements, protein–protein interactions, and transgenic analysis of the LEA gene families were performed to identify their functions in cold stress in kernel apricot. Our results provides basic research for better understanding of the function roles and evolutionary history of PasLEAs, which will helpful for future studies of selecting candidate genes for cold stress tolerance and genetic improvement in kernel apricot.

2. Materials and methods

2.1. Identification and phylogenetic analyses of LEA gene family in *P. armeniaca* L. × *P. sibirica* L.

LEA genes were identified in kernel-using apricots (*P. armeniaca* L. × *P. sibirica* L.) based on conserved LEA protein domains downloaded from Hidden Markov Model (HMM) (PF 03760, PF03168, PF03242, PF02987, PF0477, PF10714, PF04927 and PF00257 (Finn et al. 2010). The HHM profile of LEA protein was subsequently employed as a query to perform a HMMER search (http://hmmer.janelia.org/) (Finn et al. 2011) against the *P. armeniaca* × *P. sibirica* database, with E-value <0.01. All LEA candidates were identified simple modular architecture research tool (SMART, http://smart.embl-heidelberg.de/) (Letunic et al., 2012)), NCBI CDD (Conserved Domain Database, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al., 2015) and the basic local alignment search tool (BLAST) databases (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Letunic et al., 2012)) to confirm that each LEA was considered to belong to the LEA family. Using the Pfam nomenclature, the *Populus* LEA gene family was divided into eight groups: LEA1–6, seed maturation protein (SMP) and dehydrin.

2.2. Sequence alignments and collinearity analysis

PasLEAs were aligned with AtLEAs from *Arabidopsis* (https://www.arabidopsis.org/) (Bies-Ethève et al., 2008) and PtLEAs from *Populus trichocarpa* (Lan et al., 2013) using the Clustal W program (Thompson et al.,

1994). A phylogenetic tree was constructed with Molecular Evolutionary Genetics Analysis (www.megasoftware.net), 7.0 software (Kumar et al., 2016) packages with the Neighbor Joining (NJ) method, and bootstrap values was 1000 replicates. PasLEAs were named and confirmed by sequence homology and phylogenetic analyses. The syntenic relationships between the *P. armeniaca* ×. *sibirica*, *A. thaliana*, and *P. trichocarpa* genomes were constructed by MCScan (Tang et al., 2008a, 2008b, 2009) following the protocol for the Plant Genome Duplication Database (Lee et al., 2012). To identify the syntenic regions among *P. armeniaca* ×. *sibirica*, initially, potential homologous gene pairs were detected using the BLASTP algorithm ($E < 1e^{-5}$, top five matches) between every two genomes. Second, the BLASTP results and the gene location information were input into MCScanX software (Wang et al., 2012a). *P. armeniaca* × *P. sibirica* whole genome/segmental and tandem duplications were subsequently identified. All syntenic relationship and gene location data were presented using TBtools software (Chen et al., 2018).

2.3. Protein properties and gene structure analysis

The physicochemical parameters, including putative isoelectric point (pI) and the molecular weight (Mw) of the *PasLEA* protein sequences were predicted using Compute pI/Mw tool of ExPASy (http://web.expasy.org/compute_pi/). GRAVY (grand average of hydropathy) values were calculated using the PROTPARAM tool (http://web.expasy. org/protparam/) (Gasteiger et al., 2003). WolF PSORT (http://www.genscript.com/wolf-psort.html) (Horton et al., 2007) and PlantmPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) (Chou et al., 2010) was used to predicted the subcellular localization of the PasLEA proteins. Conserved protein motifs of each PasLEA member were identified using the Multiple Expectation Maximization for Motif Elucidation (MEME) system (http://alternate.meme-suite.org/). The exon/intron organization of PasLEAs were investigated by using the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) based on genome sequences.

2.4. Promoter cis-element analysis and protein–protein interaction study

The 2.0 kb upstream region from the translation start site of all PasLEA genes were obtained from the genomic DNA sequences of the PasLEA genes. Transcriptional response elements in the promoter of PasLEA genes were predicted using the PLACE database

(http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

The prediction of interacting networks of proteins was generated from the STRING database (https://string-db.org/?tdsourcetag=s_pctim_aiom).

2.5. cDNA-library preparation and Illumina sequencing for transcriptome analysis

Total RNA was extracted from the ED, SP, and GS of the flower buds of two Youyi plants using a Plant RNA Extraction Kit (Autolab, China) following the manufacturer's protocol. The concentration and quality of each RNA sample were determined using a NanoDrop 2000TM microvolume spectrophotometer (Thermo Scientific, Waltham, MA, USA) and gel electrophoresis. Poly(A) mRNA was isolated from 10 mg of total RNA using magnetic oligo(dT) beads and then divided into short fragments by fragmentation buffer (Ambion, Austin, TX, USA). First-strand cDNA was synthesized with random hexamer primers, and then second-strand cDNA was synthesized using DNA polymerase I (New England Biolabs), RNase H (Invitrogen), buffer, and dNTPs. Short DNA fragments were purified with a QIAquick PCR Purification Kit (Qiagen, Inc., Valencia, CA, USA), subjected to end repair and poly(A) addition and ligated to sequencing adaptors. Suitable fragments (350-450 bp) were purified by agarose gel electrophoresis and gathered by PCR amplification. Six cDNA libraries (ED, SP, and GS of the flower buds of two Youyi plants) were sequenced by using the Illumina HiSeq™ 2000 platform. High-quality clean reads was mapped to the reference genome to calculate the expression level in Fragments Per Kilobase of transcript per Million fragments mapped (FPKM).

2.6. Yeast two-hybrid (Y2H) Assays

Y2H experiments were performed using the Matchmaker Gold Yeast Two Hybrid System (Clontech). The coding regions of the PasLEA genes and predicted interacting proteins were cloned into pGBKT7 and pGADT7 vectors to construct various baits and prey vectors, respectively. The different pairs of prey and bait constructs were then cotransformed into Gold Y2H strain, and yeast cells were then grown on SD/–Leu/–Trp medium for 3 days. Transformed colonies were plated onto SD/–Leu/–Trp/–His/–Ade medium containing 100 μM aureobasidin A and 3 mg mL-1 X-α-Gal at 30°C to detect protein interactions between PasLEAs and predicted interacting proteins.

2.7. Bimolecular fluorescence complementation assay

To generate vectors for the bimolecular fluorescence complementation (BiFC) assays, the full-length coding sequences of PasLEA1-2 and PasERD4, PasLEA3-1 and PasXTH9, PasSMP-3 and PasEXPA22, and PasDehydrin-3 and PasCBF1, without their stop codons were subcloned into the pUC-pSPYNE or pUC-pSPYCE vectors. The expression of target genes alone was used as negative controls. The resulting constructs were transformed into *N. benthamiana* leaves as described above (Ding et al., 2018). The signal was detected by confocal microscope. The primers used in the BiFC assay are listed in Supplementary Table S1.

2.8. Yeast one-hybrid (Y1H) assay

Y1H assay was carried out using the Matchmaker™ Gold Yeast One-Hybrid System (Clontech). The *PasLEA3-2*-pro (−323 to −1784) reporter strains, and the *Pasdehydrin-3*-pro (−179 to −1778) reporter strains termed *PasLEA3-2*- pro-Y1H and *Pasdehydrin-3*- pro-Y1H, were cloned into pAbAi vector. The resulted plasmids were linearized and transformed into Y1H gold strain. Positive yeast cells were then transformed with pGADT7-AD, which contained PasLEAs. The DNA–protein interaction was determined based on the growth ability of the cotransformants on SD/-Leu/-Ura medium with Aureobasidin A (AbA) according to the manufacturer's protocol.

2.9. Gene expression analysis and qRT-PCR

To evaluate expression patterns of PasLEAs in three periods of dormancy and germination in *P. armeniaca* × *P. sibirica* 'Youyi', RNA-seq data were collected from the transcriptome sequencing data of ecological dormancy (ED), sprouting period (SP) and germination stages (GS) of the flower buds of two 'Youyi' plants which are grow in the same condition. The bud and stem samples at each stage were collected according to Song et al. (2017). Flower buds were checked and did not burst under forced conditions on the sampling day of PD or ED period, and could burst under forced conditions on the sampling day of SP or GS period. When flower buds are in the SP period, flower buds are not covered by scales, and petals are not visible. While flower buds are in the GS period, they are not covered by scales, and petals are visible. RNA-Seq data of *P. armeniaca* × *P. sibirica* 'Youyi' in this research were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database with number SRS1042411. The heatmap and hierarchical clusters were generated using HemI 1.0 software (http://hemi.biocuckoo.org/down.php). Quantitative real-time (qRT)-PCR method were chosen for comparing the expression of PasLEA genes in flower buds and stems of three 'Youyi' plants which the growth status is almost consistent with each other during physiological dormancy (PD), ecological dormancy (ED), sprouting period (SP) and germination stages (GS) of *P. armeniaca* × *P. sibirica*. Total RNA was extracted from stems and flower buds using a Plant RNA Extraction Kit (Autolab, China), and reverse transcription reactions were performed using Superscript II reverse transcriptase (Invitrogen, Grand Island, NY, USA) according to the manufacturer's instructions. The primers were designed using Primer Premier 5.0 software (Premier Biosoft Int., Palo Alto, CA, USA) and the specific primer pairs are listed in Table S1. qRT-PCR was conducted using an ABI Prism 7500 sequence detection system (Applied Biosystems, Foster City, CA, USA) and the SYBR® Premix Ex TaqTM Kit (TaKaRa, Tokyo, Japan) to measure transcript levels. *PaElf* was used as a reference gene in qRT-PCR detection. qRT-PCR analysis of every gene was performed in triplicate for three

biological sample, and fluorescence quantitative results were analyzed using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

2.10. Coexpression network analysis

The coexpression data for PasLEAs were obtained from RNA-Seq data (SRA database number SRS1042411). For *P. armeniaca* × *P. sibirica* genome-wide coexpression network construction, transcriptome data from the flower buds of two *P. armeniaca* × *P. sibirica* plants were used. Genes with Pearson correlation coefficients ≥ 0.88 were selected, and 2350 genes were preliminary screened. Among 2350 genes coexpressed with PasLEAs, 15 genes involved in cold tolerance or flower bud development were chosed to build the coexpression network of *PasLEAs*. The coexpression network diagram was created according to Cytoscape software (The Cytoscape v1.1 Core was downloaded from http://www.cytoscape.org/).

2.11. Transient expression of PasLEAs in Arabidopsis protoplasts

The GFP fragment was amplified and ligated into the pBI221 vector (Clontech, Mountain View, CA, USA). The coding regions of PasLEA1-2, PasLEA2-18, PasLEA3-1, PasLEA4-13, PasLEA5-2, PasLEA6-2, PasSMP-2 and Pasdehydrin-3 cDNA were isolated and fused to the 5′ end of the GFP DNA sequence. The constructed vector 35S:PasLEAs-GFP were instantaneously transfected into Arabidopsis leaf protoplasts via polyethylene glycol-mediated method (Sheen, 2001). The nuclear dye DAPI (10 mg/mL; Sigma-Aldrich Corp., St. Louis, MO, USA) was used to stain the nuclei (Li et al., 2020). For chloroplast localization, the chloroplast autofluorescence was obtained through the red channel (Guo et al., 2019). The fluorescent signals were observed with a UltraVIEW VoX confocal microscope (PerkinElmer Life Sciences Inc., USA). The fluorescence of GFP and FM4-64 in the cells was analyzed with a 488-nm argon laser, and a 551-nm helium-neon laser respectively, via confocal laser-scanning microscopy (TCS SP8, Leica, Germany; GFP excitation at 488 nm and emission at 505-525 nm, FM4-64 excitation at 551 nm and emission at 498-588 nm).

2.12. Yeast transformants and low-temperature treatment

The CDSs of *PasLEA3-1*, *PasLEA3-2*, *PasLEA3-3*, *PasLEA5-1*, *PasLEA5-2*, *Pasdehydrin-3*, or *Pasdehydrin-4* were inserted in the yeast expression vector pGAPZA, respectively. The recombined vector pGAPZA::*PasLEAs* harboring the cold resistance gene was introduced into yeast strains. Positive colonies were then identified by PCR, and cold treatment was conducted in solid and liquid Simmons Citrate-Ura (SC-U) medium as described previously (Liu et al., 2015). Briefly, colonies of positive clone harboring *PasLEAs* gene or empty-vector strains were shake cultured at 200 rpm 30 □ 24 h to induce PaLEA protein expression. In order to the cold treatment, 1 mL aliquot yeast (containing the pGAPZA empty or pGAPZA::*PasLEAs* recombinant vectors) culture liquid (at an OD600 value of 1.0) was stayed at -20 □ for 24 h first. Then, the yeast liquid was diluted 1:10, 100, 1000, and 10,000, respectively, followed 4 μL of the original yeast liquid and the diluted liquid were dropped in order on solid SC-U medium (containing 2% galactose). After 48 h culture at 30 □, growth rate of the transformed yeast cells was observed and recorded. And the same time, 1 mL aliquot yeast (containing the pGAPZA empty or pGAPZA::*PasLEAs* recombinant vectors) liquid at an OD600 value of 1.0 was removed in 10 mL liquid SC-U medium (containing 2% galactose). After a 24 h -20 □ treatment at 200 rpm, 30 □, OD600 value of the transformed yeast cells was measured. Control group remained untreatment without any stress was also included.

2.13. Plant transformation and cold resistance analysis

The *PasLEA3-2* and *Pasdehydrin-3* CDS were cloned into plant expression vector PBI121 respectivelly. The recombinant plasmid PBI121::*PasLEA3-2* and PBI121::*Pasdehydrin-3* was introduced into *Agrobacterium tumefaciens* (LBA4404). The transformation into *Arabidopsis* was carry out by using floral dip method (Clough and Bent, 2010). To test the effect of cold stress on seedling growth, the seeds were allowed to germinate at normal growth conditions and were then transferred to growth chambers of 22°C, 16°C and 16°C 16 h/4°C 8 h. After 10 days of treatment, the phenotypic indexes and gene expression of transgenic *Arabidopsis* and wild type were

determined. Physiological indexes such as superoxide dismutase (SOD) activity (Wang et al., 2012b), malondialdehyde and proline levels were estimated as previously described (Bates et al., 1973). Each sample taken at least three individual plants, and all phenotypic and physiology index testing and analysis were conducted in triplicate.

2.14. Statistical analyses

Measurement data for root length, leaf length, leaf width, pod length, OD600 value and gene expression intensity, etc, were performed by one-way analysis of variance (ANOVA) followed by the Tukey's HSD post hoc analysis. All the tests were conducted in triplicate. A p-value < 0.05 was considered statistically significant between two groups of data in all comparisons. SPSS 16.0 software (SPSS Inc., Chicago, IL, USA) was used to performed statistical analyses.

3. Results

3.1. Identification, classification and collinearity analysis of PaLEA genes in *P. armeniaca* × *P. sibirica*

HMM search with eight LEA HMM profifiles (PF03760, PF03168, PF03242, PF02987, PF0477, PF10714, PF04927 and PF00257) and BLAST using well-characterized LEA sequences from *Arabidopsis* as queries. In total, 54 full-length genes encoding putative LEA proteins were identified from *P. armeniaca* × *P. sibirica* (Table 1). A phylogenetic analysis based on conserved domain structures of the predicted proteins revealed that the 54 *PasLEA* genes were clustered into eight subgroups together with the *Arabidopisis* and *Populus* LEAs, confirming that they belonged to the LEA family (Figure 1a). Eight subgroups included LEA-1, LEA-2, LEA-3, LEA-4, LEA-5, LEA-6, dehydrin, and SMP. The LEA-2 family was the largest, with 33 LEA members in *P. armeniaca* × *P. sibirica*. The dehydrin group had five members, and both LEA-1, LEA-3, LEA-5, and SMP had three members. By contrast, the LEA-4 and LEA-6 groups had only two members, respectively. All of these LEAs were grouped with *A. thaliana* and *Populus* LEAs, confirming that they belonged to the LEA family (Figure 2). The *Arabidopsis* LEA gene family had a specific group (AtM) which was absent in the *P. armeniaca* × *P. sibirica* genome.

Then, apricot genomes were collinear analyzed with *Arabidopsis thaliana* and poplar genomes (Figure 1b). The results showed that apricot and *Arabidopsis thaliana* were closely related, and there were more collinearity, while apricot and poplar had less. Collinearity gene pairs were only observed for several members of the LEA gene family of apricot with poplar LEA genes. Among the 28 apricot LEA genes, only 9 genes showed collinearity with poplar LEA genes, such as Pasdehydrin-3, Pasdehydrin-4, PasSMP-2, etc. These results indicated that the ancestral differentiation of apricot and poplar was relatively long, and the structural variation at the chromosomal level had destroyed the conserved collinear arrangement relationship of ancestral genes. While most members of the LEA gene family of apricot were collinear with *Arabidopsis thaliana* LEA genes, indicating that they still had good block conservatism despite the change of corresponding chromosomal locations. We speculated that although the fusion and breakage of chromosomes in the evolutionary process changed the relative physical positions of genes in apricot and *Arabidopsis thaliana*, and even caused changes in the number of chromosomes in the two species, the collinear arrangement of gene locations in family is still conservative. Additional, except for one-to-one collinearity, one-to-many or many-to-one phenomena also appeared in the collinearity relationship between LEA gene families of apricot and *Arabidopsis thaliana*, suggesting that obvious family gene replication events occurred after genomic differentiation between apricot and Arabidopsis thaliana.

Bioinformatics analysis showed that the molecular weight (MW) of the PasLEAs ranged from 5745.61 to 48703.71 Da, and their PIs ranged from 4.64 to 10.36 (Table 1). The LEA1s, LEA3s, and LEA5s were smaller (<20 kDa) than the LEA2s, SMPs and Pasdehydrins. The PasLEA5s, PasLEA6s and PasSMPs were acidic (except for PasLEA5-3), while the majority of the other subfamilies were alkaline. The grand average of hydropathy (GRAVY) scores of the PasLEA proteins ranged from -1.612 to 0.297, and the PIPs presented the lowest average GRAVY

value (-1.282) among the subfamilies (Table 1).

3.2. Chromosomal distribution, gene duplication and selection pressure of LEA members in apricot

The transcript length of PasLEAs ranged from 156 bp to 1392 bp, and gene length ranged from 159 bp to 3816 bp. The number of amino acids in the identified PasLEAs ranged from 52 (PasLEA1-3) to 464 (Pasdehydrin-5) (Table 1, Table S2). The 54 identified *PasLEAs* were unevenly mapped on the eight chromosomes of *P. armeniaca* × *P. sibirica*. Chromosome 6 contained the largest number (13; 24.07%) of *PasLEAs* genes, followed by chromosomes 7 and 8, which contained twelve members (22.22%) and six members (11.11%), respectively. Chromosomes 1, 2, 3 and 5 both contained five members (9.26%). Chromosome 4 contained only three member (5.56%) (Figure S1).

PasLEA2-22 was grouped together with PasLEA2-28/PasLEA2-29 in phylogenetic analysis, suggesting that these PasLEA2-22/ PasLEA2-28/ PasLEA2-29 might have descended from a common ancestor. On the basis of phylogenetic tree and positions of these three genes in chromosome 5, it could be seen that PasLEA2-22 was probably duplicated during the whole-genome duplication event, while PasLEA2-28/ PasLEA2-29 were probably duplicated by segmental duplication. Additionally, in the gene cluster, PasLEA3-1 was associated with segmental duplication by exhibiting synteny relationship with PasLEA3-2. PasLEA3-1/ PasLEA3-2 were homologous genes in phylogenetic tree, indicating that PasLEA3-1/ PasLEA3-2 might have descended from a common ancestor. PasLEA5-1 and PasLEA5-2 were homologous genes in phylogenetic tree, indicating that these two genes might have descended from a common ancestor. According to their positions in chromosome 2, we concluded that PasLEA5-1 was associated with segmental duplication by exhibiting synteny relationship with PasLEA5-2. In addition, phylogenetic analysis indicated that PasLEA2-9 was grouped together with PasLEA2-12 and PasLEA2-13, indicating these three genes might have descended from a common ancestor. Furthermore, we speculated that PasLEA2-9 was probably duplicated during the whole-genome duplication event, while PasLEA2-12 and PasLEA2-13 were tandem duplication genes in the cluster of chromosome 2.

Ka/Ks represents the ratio between the non-synonymous substitution rate (Ka) and synonymous substitution rate (Ks) of two protein-coding genes, which is used to evaluate whether there is selection pressure acting on this protein-coding gene. The Ka /Ks ratio indicates the evolutionary direction, in which Ka/Ks ratio >1, =1 and < 1 represent positive selection, neutral selection and purification selection in the process of evolution, respectively. The selection pressure of members of LEA gene family in apricot was evaluated by calculating the Ka/Ks of orthologous LEA gene pairs between the apricot and poplar genomes or between the apricot and Arabidopsis thaliana genomes. As shown in Table 2, all the PasLEA genes had Ka/Ks ratios < 1, indicating there were purification selection for all the LEA gene families of apricot during the process of evolution. The results showed that all PasLEA genes had Ka/Ks ratios greater than 0.1 with the reference genome of Arabidopsis thaliana. Of which, PasLEA2-4 had lowest Ka/Ks ratio of 0.1129, while PasLEA2-3 had highest Ka/Ks ratio of 0.4452. Additionally, the most of PasLEA genes had Ka/Ks ratios greater than 0.1 with the reference genome of poplar. PasLEA2-6 had the highest Ka/Ks ratio of 0.4395. Four genes had Ka/Ks ratio less than 0.1, including PasLEA2-20 (ka/ks=0.0870), PasLEA2-28 (ka/ks=0.0664), PasLEA2-29 (ka/ks=0.0619) and PasLEA6-1 (ka/ks=0.0444). It is speculated that LEA gene families of apricot might preferentially conserve function and structure during evolution.

3.3. Gene structure, conserved motifs and molecular modeling of PasLEAs

To characterize the phylogenetic relationships and structural diversity of the PasLEAs genes, exon-intron organization analysis of the *PasLEAs* family genes was performed. As shown in Figure S2, the distribution of introns in the *PasLEA* genes ranged from zero to two introns per gene. The number and size of the exons and introns, were conserved within both PaLEA 3 and PaLEA 5 group. Most PasLEA2 group genes contain zero or one introns, except PaLEA2-17 showing two introns. The exon-intron organization analysis approximately exhibited a similar pattern of gene structure within the same subfamilies of PasLEAs. The relatively conserved

exon-intron feature verifys the phylogenetic relationships and classification of subgroups of *P. armeniaca* × *P. sibirica* (Figure S2). Motif scan analysis were applicated by using MEME program to further analyze the conservation, relevance and diversity in each subfamily of *P. armeniaca* × *P. sibirica*. Ten conserved motifs were identified in the PasLEA family, and several members in the same subfamily shared a similar number of motifs (Figure S3). For example, all the members of PasSMPs had ten motifs, except for PasSMP-2. All the members of PasLEA3 had seven motifs, except for PasLEA3-3. In order to meet the needs of structural genome and gene function research, 3D proteins structure of *P. armeniaca* × *P. sibirica* LEAs can be effectively predicted (Groot and Grubmüller, 2005). The structural properties of all 54 PasLEAs were displayed in homology-based tertiary (3D) protein models, which were predicted via the Phyre2 server display, and the results are shown in Table S3. Most of 3D protein models were constructed greater than 90% confidence, and the percentage of aligment coverage almost 60%. These predicted 3D models provided an important basis for the structure characteristics of gene function analysis of the LEA genes of *P. armeniaca* × *P. sibirica* (Deshmukh et al., 2013; Tombuloglu et al., 2016).

3.4. Promoter cis-element analysis

Promoter sequence analysis indicated that the PasLEA genes contained several hormone-related, stress- and development-responsive cis-acting regulatory elements, including low-temperature-responsive elements (LTE), drought-inducible elements (MBS), abscisic acid (ABA)-response elements (ABREs), auxin-responsive elements (TGA AuxRE-core and AuxRR-core), gibberellin-responsive element (GARE-motif, P-box, and TATC- box), light-responsive elements (GT1, MRE, TCT-motif and G-box), and MeJA-responsiveness (CGTCA-motif, TGACG-motif). Several regulatory element-related site were found, such as AT-rich DNA binding site (AT-rich element), protein binding site (Box III). (Figure S4 and Table S4). Among the stress-related cis-acting elements, almost all PasLEAs (48 genes, except for PasLEA2-6, PasLEA2-13, PasLEA2-19, PasLEA2-23, PasLEA2-26, and Pasdehydrin-3) exhibited AREs (essential for anaerobic induction). There were 47 MBS (drought-inducible) and 31 LTR (low-temperature-responsive) elements located in 32 and 22 PasLEA promoters, respectively. This Cis-element analysis indicated that LEA proteins may play a significant role in stress tolerance and response in *P. armeniaca* × *P. sibirica*. Among the hormone-related cis-acting elements, 211 ABRE elements (ABA-responsive), 67 TGACG motifs (MeJA -responsive) and 70 CGTCA motifs (MeJA-responsive) were identified in the promoters of 46, 35, and 37 PasLEAs, respectively. Especially, a large number of elements related to light responsiveness, such as 173 G-box and 112 Box 4 were identified in the promoters of 42 and 48 PasLEAs. Some elements related to cell cycle regulation (MSA-like), meristem expression (CAT-box) elements, zein metabolism regulation (O_2-site), flavonoid biosynthesis (MBSI) and endosperm expression (GCN4_motif) were detected, indicating that PasLEAs may involved in developmental processes in *P. armeniaca* × *P. sibirica*. PasLEA genes containing different promoter Cis-element and element numbers may reflect the differences in expression profiles of *PasLEAs* and their different roles associated with plant growth and abiotic stress tolerance.

3.5. Protein-protein interaction network and interaction confirmation of PasLEAs

To further clarify the function of PasLEAs in interaction with other proteins in dormancy and sprouting periods, PasLEA1-2, PasLEA3-1, PasSMP-3, and PasDehydrin-3, which showed high expression in each subgroup of PasLEAs family (Figure 3), were used to construct a protein-protein interaction network. As shown in Figure 3a-d and Table S5, most of these tested PasLEAs did not shared the same interacting proteins, which showed the diversity of interaction forms. In addition, several LEA homologous gene interacted with PasLEA1-2 (homologous to AtLEA4-1) separately, such as LEA3, LEA7, LEA4-5, etc. This search result indicated that PasLEA1-2 could interact with LEAs members and form homodimers. In particular, the interacting zinc finger proteins and NAC transcription factors can enhance cold tolerance in transgenic plants (Jung et al., 2013). The proteins that interact with PasLEA3-1 also include some LEA homologous gene such as LEA4-1, LEA14, LEA31,

etc. Other protein such as XTH9, which encodes a member of xyloglucan endotransglucosylase/hydrolases (XTHs) that catalyze the cleavage and molecular grafting of xyloglucan chains function in loosening and rearrangement of the cell wall, may function as PasLEA1-2-interacting protein involved in dormancy and sprouting stages. AGL81, Agamous-like MADS-box protein AGL81, which is found in PasSMP-3-interacting proteins, may function in the maintenance of the proper function of the central cell in pollen tube attraction predicted by string website. Interestingly, ECP63 and LEA domain-containing protein (homologous to AT1G72100) both the interaction protein of PasLEA1-2 and PasSMP-3. Notably, PasDehydrin-3 specifically interacted with several cold-responsive genes such as CBF1, CBF2, COR15A, COR15B, etc, including KIN2, which is encodes a gene that can be induced by cold and abscisic acid and may be involved in cold acclimation and salt tolerance (Table S5). Y2H experiments showed that yeast cells cotransformed with constructs PasDehydrin-3-pGBKT7 and PasCBF1-pGADT7, PasLEA1-2-pGBKT7 and PasERD4-pGADT7, PasLEA3-1-pGBKT7 and PasXTH9-pGADT7, and PasSMP-3-pGBKT7 and PasEXPA22-pGADT7, could interact with each other in the selection medium and b-galactosidase staining (Figure 3e). The interaction between PasDehydrin-3 and PasCBF1, PasLEA1-2 and PasERD4, PasLEA3-1 and PasXTH9, and PasSMP-3 and PasEXPA22 occurred in the nuclei of leaf pavement cells revealed by BiFC experiments (Figure 3f). Thus, Y2H and BiFC assays support the protein interaction between PasDehydrin-3 and PasCBF1, PasLEA1-2 and PasERD4, PasLEA3-1 and PasXTH9, and PasSMP-3 and PasEXPA22. In particular, the *A. thaliana* gene encoding early responsive to dehydration 10 (ERD10) was induced by cold treatment. The increased level of *ERD10* is required to activate the *CBF*/*DREB1* genes and their downstream genes during cold stress in *Arabidopsis* (Sun et al., 2010). This may also be conserved within the kernel-using apricots.

3.6. Expression profiling of PasLEAs in the dormancy and sprouting stages of *P. armeniaca* × *P. sibirica*

In order to clarify the potential role of PasLEA proteins in *P. armeniaca* × *P. sibirica*, we analyzed the expression profiles of *PasLEAs* in the ecological dormancy (ED), sprouting period (SP) and germination stages (GS) of flower buds by RNA sequencing. The RNA-seq data can be found in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under number SRS1042411. The expression levels of 54 *PasLEA* genes were clustered into a heatmap (Figure S5). We observed the expression of 29 *PasLEAs* (88.89%) in at least one of the developmental stages, and all of these *PasLEAs* showed differential expression in three stages of the dormancy and sprouting periods. Among these genes, 33 *PasLEA* genes (68.75%) were expressed in all three periods of dormancy and germination, among which *PasLEA6-1*, *PasLEA2-20* and *Pasdehydrin-3* exhibited the highest expression levels. Interestingly, *PasLEA1-3*, *PasLEA2-8*, *PasLEA2-15*, *PasLEA2-19*, *PasLEA2-28*, *PasLEA2-30*, and *PasLEA5-1* were not expressed in any of the three periods of dormancy and germination. In addition to these six genes, *PasLEA2-6*, *PasLEA2-13*, *PasLEA2-3*3, *Pasdehydrin-1* and *PasLEA2-29* were not expressed during the germination stages of flower buds; *PasLEA1-1*, *PasLEA2-29*, *PasLEA2-33*, *PasLEA5-3*, *PasLEA6-2* and *PasLEA2-31* were not expressed in the sprouting period of flower buds, and *PasLEA1-1*, *PasLEA2-6*, *PasLEA2-12*, *PasLEA3-2*, *PasLEA6-2*, *PasSMP-2*, *Pasdehydrin-4* and *PasSMP-1* were not expressed in the ecological dormancy period.

The heatmap demonstrated the clustering of four main clades. The *PasLEA* genes from clade 1 corresponded to 1 LEA2 and 1 LEA6 members (*PasLEA2-20* and *PasLEA6-1*) and 2 Pasdehydrin members (*Pasdehydrin-3* and *Pasdehydrin-5*) showed the highest expression in three periods of dormancy and germination. The genes from clade 3 included 9 LEA2 members (*PasLEA2-1*, *PasLEA2-4*, *PasLEA2-2, PasLEA2-17*, *PasLEA2-10*, *PasLEA2-9*, *PasLEA2-22*, and *PasLEA2-26*), 2 LEA3 members (*PasLEA3-1* and *PasLEA3-3*) and 1 dehydrin member (*Pasdehydrin-2*) that exhibited relatively higher expression in these three periods, while those genes from clade 2 included 7 LEA2 members (*PasLEA2-3*, *PasLEA2-18*, *PasLEA2-7*, *PasLEA2-5*, *PasLEA2-23*, *PasLEA2-6* and

PasLEA2-11) and 1 SMP member (*PasSMP-3*) showing an intermediate level of expression, and clade 4 included 3 LEA1 members (*PasLEA1-1*, *PasLEA1-2*, and *PasLEA1-3*), 16 LEA2 members (*PasLEA2-31*, *PasLEA2-21*, *PasLEA2-12*, *PasLEA2-8*, *PasLEA2-19*, *PasLEA2-28*, *PasLEA2-15*, *PasLEA2-30*, *PasLEA2-29*, *PasLEA2-33*, *PasLEA2-13*, *PasLEA2-16*, *PasLEA2-32*, *PasLEA2-27*, *PasLEA2-14*, and *PasLEA2-24*), 1 LEA3 member (*PasLEA3-2*), 2 LEA4 members (*PasLEA4-13*, and *PasLEA4-23*), 3 LEA5 members (*PasLEA5-1*, *PasLEA5-2*, and *PasLEA5-3*), 1 LEA6 member (*PasLEA6-2*), 2 SMP members (*PasSMP-1* and *PasSMP-2*), and 2 dehydrin member (*Pasdehydrin-1* and *Pasdehydrin-4*) displaying the lowest expression in three developmental periods of flower buds (Figure S5).

3.7. Expression patterns of PasLEA genes in different tissues during the dormancy and sprouting stages of *P. armeniaca* × *P. sibirica*

In order to reveal the potential roles of *PasLEA* genes under low temperature stress, the expression pattern of 32 genes was analyzed in two tested tissues (flower buds and stems) by qRT-PCR. Four highly expressed genes (*PasLEA2-20*, *PasLEA6-1*, *Pasdehydrin-3*, and *Pasdehydrin-5*), nine genes with low expression (*PasLEA1-2*, *PasLEA2-16*, *PasLEA2-32*, *PasLEA2-27*, *PasLEA2-13*, *PasLEA5-2*, *PasLEA3-2*, *PasLEA2-24*, and *PasLEA6-2*), twelve genes (*PasLEA2-1*, *PasLEA2-17*, *PasLEA2-4*, *PasLEA2-22*, *PasLEA2-26*, *PasLEA3-3*, *PasLEA2-2*, *PasLEA2-10*, *Pasdehydrin-2*, *PasLEA2-25*, *PasLEA2-9* and *PasLEA3-1*) with stable expression patterns in all stages of dormancy and germination, seven genes (*PasLEA2-3*, *PasLEA2-23*, *PasLEA2-5*, *PasLEA2-7*, *PasLEA2-11*, *PasSMP-3* and *PasLEA2-6*) with differential expression patterns, were chosen for qRT-PCR analysis. The qRT-PCR results of most selected PasLEAs showed similar expression patterns with the transcript-level changes analysed by RNA-seq data in the dormancy and sprouting stages of flower buds, which suggested that the reliability of transcriptome data. However, moderate discrepancies in the transcriptional levels were detected in three PasLEAs compared to the RNA-seq data: *PasLEA3-3* and *Pasdehydrin-2* (Figure 4). The expression of *PasLEA2-4*, *PasLEA2-7*, and *PasLEA2-20* in the stems showed almost the same trends as in the flower buds in the four examined stages of *P. armeniaca* × *P. sibirica*. PasLEA2-25, PasLEA2-26, and Pasdehydrin-3 showed the highest expression during physiological dormancy in the stem, and *PasLEA1-1*, *PasLEA2-3*, *PasLEA2-12*, *PasLEA2-16*, *PasLEA2-18*, *PasLEA2-20*, *PasLEA2-21*, *PasLEA2-23*, *PasLEA2-27*, *PasLEA2-31*, *PasLEA2-32*, *PasLEA4-23*, *PasLEA6-1*, *PasLEA6-2*, *PasSMP-1*, *PasSMP-2*, *Pasdehydrin-4* showed the highest expression in the germination stage in the flower buds (Figure 4). Gene expression preferences in different tissues may suggested differ biological functions of *PasLEAs* involved in the dormancy and sprouting periods of *P. armeniaca* × *P. sibirica*.

3.8. Coexpression network of *PasLEAs* in *P. armeniaca* × *P. sibirica*

A *PasLEAs* coexpression network was constructed to clarify the mechanism of *PasLEAs* involved in dormancy and sprouting stages of *P. armeniaca* × *P. sibirica* (Figure S6a and Table S6). A total of 2350 coexpression pairs (Correlation coefficient ≥ 0.88) were identified coexpressed with *PasLEAs*. Among these, 15 genes (0.64%, Correlation coefficient ≥ 0.93) are associated with cold resistance or flower bud development. For example, transcriptome-based identification of *AP2/ERF* family genes and their cold-regulated expression detected during the dormancy phase transition of Chinese cherry flower buds (Zhu et al.,2021). Over expressing a chloroplast ω-3 fatty acid desaturase gene enhanced cold tolerance in transgenic tobacco (Khodakovskaya et al., 2006). Auxin response factor play vitol role in enhancing stress resistance in *Arabidopsis* under cold stress (Shibasaki et al.,2009). Tahtiharju (2001) also found that antisense inhibition of protein phosphatase 2C accelerates cold acclimation in *Arabidopsis thaliana*. In this research, Ethylene responsive element binding protein, fatty acid desaturase gene, auxin response gene, and protein phosphatase 2C etc, were highly coexpressed with PasLEAs in different tissues during the dormancy and sprouting stages of *P. armeniaca* × *P. sibirica* and may be participated in low temperature resistance in *P. armeniaca* × *P. sibirica*.

We constructed the network with the core of *PasLEAs*, the expression changes of the 15 genes associated

with cold tolerance in the *PasLEA* coexpression network were analysised in the PD, ED, SP, and GS of flower buds and stems of *P. armeniaca* × *P. sibirica* (Figure S6b). Among the 15 pairs of coexpressed genes, the expression trends of *PasEREBP* and *PasLEA2-9*, *PasSer/Thr* and *PasLEA2-16*, *PasSer/Thr* and *PasLEA2-23*, *PasWRKY* and *PasLEA2-22*, and *PasAUX* and *PasLEA5-2* were consistent in the four periods (PD, ED, SP, and GS) of flower buds of *P. armeniaca* × *P. sibirica*. Similarly, the expression was similar between *PasMYB* and *PasLEA2-5*, *PasSer/Thr* and *PasLEA2-3*, *PasSer/Thr* and *PasLEA2-20*, *PasAUX* and *PasLEA2-4*, *PasNAC* and *Pasdehydrin-3*. Additionally, several pairs exhibited no differences in expression patterns, such as *PasFAD* and *Pas*LEA2-1, *PasbZIP* and *Pas*LEA2-22, *PasZFCHY* and *Pas*LEA2-13, etc.

3.9. Subcellular localization of PasLEAs protein

Plant-mPLoc and WoLF PSORT prediction was used to forecast the subcellular localization of the PasLEAs. Almost all the members of the PasLEA3、PasLEA4、PasLEA5、PasLEA6 and PasSMPs subfamilies were localized in the nuclear according to Plant-mPLoc prediction (Table 1). However, the subcellular localizations predicted by Plant-mPLoc tool indicated that PasLEA6-1 was localized at the chloroplast or Cytoplasm. The Pasdehydrins were also predicted to localize to the nuclear, except for Pasdehydrin-2 and Pasdehydrin-5 were predicted to localize to the plasma membrane according to Plant-mPLoc prediction. However, the results predicted by WoLF PSORT were diverse, and included localization to the mitochondria, endoplasmic reticulum, nucleus, cytosol, Golgi, and chloroplast (Table 1). The PasLEAs with high expression in different dormancy and sprouting stages might indicate participating in key physiological functions during bud dormancy and germination of P. armeniaca × P. sibirica. To understand the subcellular localization of these highly expressed PasLEAs genes in each subgroup, vectors of EGFP-PasLEA1-2, EGFP-PasLEA2-18, EGFP-PasLEA3-1, EGFP-PasLEA4-13, EGFP-PasLEA5-2, EGFP-PasLEA6-2, EGFP-PasSMP-2, and EGFP-Pasdehydrin-3 cDNAs (representing genes with highest expression in each subfamily in buds dormancy and bud break period in *P. armeniaca* × *P. sibirica*) were constructed and transient expressed in A. thaliana protoplast cells. Subcellular localization of fusion protein by confocal microscopy revealed that green fluorescence of these fusion proteins was confined to the nuclear and chloroplast (Figure 5). This nuclear or chloroplast localization of these highly expressed genes was consistent with the protein localizations predicted by WoLFPSORT or Plant-mPLoc software.

3.10. The *PasLEA3-2* and *Pasdehydrin-3* gene contributated low-temperature resistance in yeast

To elucidate the cold responsive mechanism of *PasLEA* genes, seven genes which exhibit high expression in flower buds and stems under low-temperature stress, were expressed in yeast strain GS115 to check their cold stress resistance. The yeast strains transformed with the pGAPZA empty vector or different pGAPZA::PasLEA recombinant expression vector were incubated in solid and liquid SC-U medium (2% galactose) for low-temperature stress treatment. Growth of yeast cells transformed with pGAPZA::*PasLEA3-2* and pGAPZA::*Pasdehydrin-3* on solid SC-U medium containing 2 M sorbitol after -20 □ low-temperature treatment exhibited remarkably enhanced growth status than that of cells transformed with the empty vector, moreover, the colony of yeast cells harbored the *PasLEA3-2* and *Pasdehydrin-3* gene were larger size than those harbored empty-vector strains (Figure 6a). In 2 M liquid SC-U sorbitol media, mean OD value of positive strains harboring *PasLEA3-2* and *Pasdehydrin-3* was 2.73 and 2.55 whereas empty-vector strains was 2.33. After -20 □ low-temperature press treatment, the mean OD value were 2.54 and 2.48 vs 1.99. (Figure 6b). The results of the yeast transformation experiment indicated that *PasLEA3-2* and *Pasdehydrin-3* genes contributated low-temperature resistance of yeast cells.

3.11. Ectopic expression of *PasLEA3-2* and *Pasdehydrin-3* in *A. thaliana* contributated cold resistance

To study the function of the *PasLEA* genes participate in cold stress, two genes (*PasLEA3-2* and *Pasdehydrin-3*) which showed low temperature resistance in yeast were transferred individually into *A. thaliana* plants. *PasLEA3-2* overexpressing (OE) *Arabidopsis* lines and *Pasdehydrin-3*-overexpressing *Arabidopsis* lines

were identified by PCR amplification of the specific gene fragments of *PasLEA3-2 and Pasdehydrin-3*. Normal growing wild-type (WT) *Arabidopsis* and the *PasLEA3-2 and Pasdehydrin-3* gene-carrying *Arabidopsis* were transferred to growth chambers at 22, 16, or 16 □ 16 h/4 □ 8 h. After 10 days treatment, the leaves of wild-type plants are yellow and showed higher sensitivity to low temperature stress than transgenics. The growth state of *PasLEA3-2* (LEA1, LEA2, and LEA3) or *Pasdehydrin-3* overexpressing *Arabidopsis* lines (DEH1, DEH2 and DEH3) are better than the WT plants (Figure 7a), exhibiting good cold resistance. We further examined the basic phenotype data in plants. Root length, leaf length and width, pod length of LEA1, LEA2, LEA3, DEH1, DEH2 and DEH3 lines had significantly longer than WT plants (Figure 8a-d). While flowering time and bolting time of both *PasLEA3-2* and *Pasdehydrin-3* overexpressing plants were reduced compared to the WT lines (Figure 8e, f). Survival rates, and seed setting rate of LEA1, LEA2, LEA3, DEH1, DEH2 and DEH3 plants had significantly increased compared with the WT against low temperature stresses (Figure 8e, f). After 10 days of treatment, the SOD activity, and proline content of *PasLEA3-2* and *Pasdehydrin-3* transgenic plants were significantly higher than the WT (Figure 8i, j). Proline level of the transgenic leaves was found to be significantly higher than the WT in the stressed conditions, indicating advantageous physiological adaptability of the transgenic plants. The better performance of the *PasLEA3-2 and Pasdehydrin-3* transformed plants was further confirmed by analyzing the MDA content and relative electrical conductivity subsequent to the low temperature stress (Figure 8k, l). Furthermore, transcriptional levels of flower bud development related genes, such as *AtAP1*, *AtBTF3L*, *AtEBF1-1*, *AtLFY*, *AtPUB26*, and *AtRAV1* were signifcantly induced in transgenic plants than in the control WT line (Figure S7). These results suggest that *PasLEA3-2* and *Pasdehydrin-3* might alter the expression of these flower bud development related genes thereby regulating the dormancy and flowering stages in the transgenic plants confer to cold tolerance.

3.12. Upstream regulatory factors of *PasLEA3-2* and *Pasdehydrin-3*

We analysed 2 kbp of the promoter sequences of *PasLEA3-2* and *Pasdehydrin-3* whether they contain several important motifs such as W-boxes (-TTGACC/T-) , MYB binding site (-CAGTTA -), which are the cognate binding motifs for WRKY or MYB transcription factors, etc, and found that PasGBF4, and PasDOF3.5 protein could bind to the promoter of PasLEA3-2, and PasDOF2.4, PasDnaJ2, and PasAP2 could bind to the promoter of Pasdehydrin-3 (Figure 9), suggesting that these protein might act as an upstream regulator in a transcription factor signalling cascade. GBF3 was identified previously as drought stress response genes in *A. thaliana*

(Ramegowda et al., 2017) and DOFs play vitrol role in stress signaling and stress response and tolerance during plant growth (Liu et al., 2021). Hu et al (2022) showed that *SlDOF9* involved in controlling inflorescence meristem and floral meristem differentiation in tomato. DnaJ proteins are critical molecular chaperones that play important roles in the stress response to stabilize plant proteins, thereby maintaining protein homeostasis under stressful conditions (Cai et al., 2022).

4. Discussion

LEAs are a large and functionally diverse group of polypeptides, and are widely distributed in plant genomes. LEA proteins accumulate during late embryogenesis and play important roles in freezing and drought tolerance in plants, in protecting plants from abiotic stress, regulating plant development and growth (Shimizu et al., 2010, Hundertmark et al., 2008). Most of the genome-wide analyses and functional analysis of LEA proteins have focused on model plants such as *O. sativa*, *A. thaliana*, *Z. mays*, *Populus* , and *Hordeum vulgare* (Choi et al. 1999; Wang et al. 2007; Hundertmark et al. 2008; Amara et al. 2012). Very little understanding about the LEAs of

economic forest species such as kernel using apricots (*P. armeniaca* × *P. sibirica*). To determine the evolutionary relationships and functional differences of PasLEAs, a phylogenetic tree was construct refer to LEAs from *A. thaliana* (in which a full set of *AtLEA* genes is relatively understanding) and *Populus trichocarpa* (which based on the reported literature (Lan et al. 2013), it is well clustered with the LEA sequence of *Arabidopsis*). The

number of *LEAs* identified in this study was almost the same as that reported in *P. trichocarpa* (Lan et al. 2013), being slightly greater than the 51 *LEA* genes reported in *A. thaliana* (Hundertmark et al. 2008) and largely lower than the 108 genes found in *B. napus* (Liang et al. 2016). Eight subgroup name of *P. armeniaca* × *P. sibirica LEAs* was also the same as in *B. napus* (Liang et al. 2016) and cotton (Magwanga et al. 2018). However, the numbers of individual subfamilies varied in these three genomes because of 108 and 242 *LEA* genes were reported in *B. napus* and cotton (Liang et al. 2016; Magwanga et al. 2018), however only 54 *LEA* genes identified in *P. armeniaca* × *P. sibirica*. Three and two members of *LEA1s* and *LEA6s* have been reported in *P. trichocarpa* (Lan et al. 2013), while seven *SIPs* and ten *TIPs* have been detected in *P. armeniaca*. The number of *P. armeniaca* × *P. sibirica LEAs* in these two subgroup was the same to their distribution in *P. trichocarpa* (Table S7).

The GRAVY value was defined by the sum of the hydrophilic values of all amino acids divided by the protein length (Table 1), and the GRAVY scores of the PasLEAs proteins can reflect their hydrophobic properties. The lowest average GRAVY value was found in the PIP subfamily, indicating good hydrophilicity of PaPIPs toward water molecules. Almost all of the PasLEAs (except for 8 PasLEA2 proteins are hydrophobic) are hydrophilic, with a GRAVY value <0, indicating that a high range of the LEA proteins are hydrophilic,these results are in accordance with those of the LEA proteins in *Arabidopsis* (Hundertmark et al. 2008). An unrooted evolutionary tree was constructed to analyze the phylogenetic and evolutionary relationships of the PasLEA proteins, in which three PasSMP (PasSMP-1, PasSMP-2, and PasSMP-3), and two PasLEA4 (PasLEA4-13 and PasLEA4-23), aroused our attention and is necessary to clarify the classification and cluster analysis. Using the naming method several articles accepted (Gao et al. 2016; Pedrosa et al. 2015), the SMP and dehydrin groups were named in naming of apricot. PasSMP-2 shared the greatest similarity with PtLEA7-2, and PasSMP-1 shared the greatest similarity with PtLEA7-1. While for PasLEA4-13 and PasLEA4-23, PasLEA4-23 shared the greatest similarity of 58% with PtLEA4-23 (Table S2), and PasLEA4-13 clustered closer to PtLEA4-13 of Poplar.

The number of *PasLEA* genes (54) in the *P. armeniaca* × *P. sibirica* genome is greater than that in *A. thaliana* (51) and greater than that in *P. trichocarpa* (53). However, while the genome size of *P. armeniaca* × *P. sibirica* (222 Mb) is smaller than that of *P. trichocarpa* (480 Mb) (Tuskan et al., 2006), and is larger than that of *A. thaliana* (164 Mb). The 54 PasLEA proteins were categorized into eight clades. We further observed that most subgroup of the *PasLEA* genes were closely related to the *LEA* genes in *A. thaliana* and *P. trichocarpa*, this research coincident with the finding of phylogenetic relationships of *Populus* and *Arabidopsis* LEA genes. It is well known that gene duplication mechanisms (segmental duplication, tandem duplication, and WGD/segmental duplication) play a key role in biological evolution (Panchy et al., 2016). In the present study, we observed that PasLEA2-28/ PasLEA2-29 probably duplicated by segmental duplication, PasLEA2-12/PasLEA2-13 were possible experienced tandem duplication, and PasLEA2-9 was most likely duplicated during the whole-genome duplication event. It indicated that orthologous genes developed through tandem duplication and segmental duplication. These three styles of duplication have played a very important role in the expansion of *P. armeniaca* × *P. sibirica LEA* family genes. The identification and function analysis of orthologous *PasLEA* genes will further promote the evolutionary history of this gene family.

WoLF PSORT and plant-mPLoc predicted that all the PasLEA1 members were localized in the nuclear, chloroplast, cytosol and mitochondria (Table 1). These prediction results were partially consistent with experimental findings for *A. thaliana* LEA1 (Candat et al., 2014; Zhao et al., 2010). For example, three LEA1 members (At1g32560.1, At2g35300.1 and At5g06760.1) were showed dual localization in the cytosol and nucleus (Candat et al., 2014). Mongolian wheatgrass MwLEA1 protein was localized to the nucleus of the onion cells (Zhao et al., 2010). In this study, subcellular localization of PasLEA1-2 indicated that PasLEA1-2 was a nuclear-localized protein. The majority of Pasdehydrin members have been predicted to localize to the nuclear and cytoplasm, and Hernández-Sánchez et al., (2015) also revealed a dual cytoplasmic and nuclear localization of the GFP::OpsDHN1

protein in *Nicotiana benthamiana* epidermal cells. As for our research, the Pasdehydrin-3 protein showed a single nuclear localization in *A. thaliana* protoplast cells. PasLEA3s also exhibit different subcellular localizations in the mitochondria, chloroplast, cytosol, or nuclear in the plant-mPLoc and WoLF PSORT prediction. Liu et al., (2013) found that Maize (*Zea mays* L.) ZmLEA3-GFP fusion protein specifically accumulated in the nucleus and cytosol of epidermis guard cells of transgenic tobacco. In our research, PasLEA3-1 was localized to the chloroplast of *A. thaliana* protoplast. The reasons for the different localizations of these LEA3 genes are unclear. In addition to different plant species and protein sequences, some literature showned that protein structure such as histidine-rich motif and S-segment, abiotic factors such as salinity and drought, but not cold, influence the subcellular localization of LEAs or other proteins (Hernández-Sánchez et al., 2015; Deokar and Bunyamin, 2016; Boursiac et al., 2008). These results indicate that the subcellular localization of plant LEAs is complex and regulated by protein motif or external environmental conditions.

LEA proteins are expressed in various tissues and organs of *P. armeniaca* × *P. sibirica* , and their temporal and spatial expression patterns is highly correlated with different developmental periods and generate specific functions. The expression profiles of *LEAs* in different tissues as well as under cold stress may provide an important basis for clarifying their physiological and molecular functions. qRT-PCR results revealed the differential expression of *PasLEAs* in tissues and development stages, but their expression law are largely differed in various organs during dormancy and sprouting periods (Figure 4), suggesting that the physiological and molecular functions of *PasLEAs* might change. For example, the higher expression levels during the endodormancy stage (*PasLEA2-11*, *PasLEA2-22*, *PasLEA2-24* and *PasLEA2-25*) indicated that the effects of restrained growth were controlled by other physiological factors resulting from inadequate low-temperature accumulation. High expression levels during the ecodormancy stage (*PasLEA2-1*, *PasLEA2-2*, *PasLEA2-7*, *PasLEA2-9*, *PasLEA2-10*, *PasLEA2-14*, *PasLEA2-26*, *PasLEA2-33*, *PasLEA3-1*, *PasLEA3-3*, *PasLEA5-3*, *PasSMP-3*, *Pasdehydrin-2*, *Pasdehydrin-3*, and *Pasdehydrin-5*) may reflect the activation of growth through release from endogenous physiological inhibitors but remain suppression attribute to insufficient total accumulated temperature. Expression during sprouting period (*PasLEA2-4*, *PasLEA2-5*, *PasLEA2-6*, *PasLEA2-13*, *PasLEA2-17*, *PasLEA3-2*, *PasLEA4-13*, *PasLEA5-2*, and *Pasdehydrin-1* are high) and germination stage (*PasLEA1-1*, *PasLEA2-3*, *PasLEA2-12*, *PasLEA2-16*, *PasLEA2-18*, *PasLEA2-20*, *PasLEA2-21*, *PasLEA2-23*, *PasLEA2-27*, *PasLEA2-31*, *PasLEA2-32*, *PasLEA4-23*, *PasLEA6-1*, *PasLEA6-2*, *PasSMP-1*, *PasSMP-2*, *Pasdehydrin-4* are high) showed recovery of morphological growth rather than the release of (endo) dormancy. Moreover, the expression of *PasLEA2-1*, *PasLEA2-4*, *PasLEA2-11*, *PasLEA2-13*, *PasLEA2-16*, *PasLEA2-25*, *PasLEA2-26*, *PasLEA2-32*, *PasLEA5-2*, *PasLEA6-1*, *PasSMP-3*, *Pasdehydrin-2*, *Pasdehydrin-3*, *Pasdehydrin-5* in stem segments were higher than that in dormant buds, where they may help to store energy and materials required for growth, bud break and germination.

In the *PasLEAs* coexpression network, similar expression patterns of gene pairs (such as PasSer/Thr and PasLEA2-16, PasWRKY and PasLEA2-22, PasEREBP and PasLEA2-9, etc) during the dormancy and germination stages suggested that *PasLEAs* cooperated with various vital flower bud dormancy/sprouting or cold stress-related genes to regulate cold resistance, bud dormancy and germinate. For example, tobacco ERE binding proteins (EREBPs) showed a novel pattern of expression during seed germination (Leubner-Metzger et al., 1998). Altogether 68 *AP2/ERF* genes were first identified in dormant Chinese cherry flower buds, among which several *PpcAP2/ERF* TFs (homologous genes of *PaEREBP*) involved in dormancy transition through their responses to low temperature (Zhu et al., 2021). MYB transcription factor MYB96, regulates seed germination by controlling the expression of abscisic acid-insensitive4 (ABI4) in seed germination of *Arabidopsis thaliana* (Leubner-Metzger et al., 1998). Wu et al (2019) found that GhNAC83 regulates the ABA signaling and cytokinin biosynthesiss corm dormancy in *Gladiolus hybridus*. Microarray-based gene expression and qRT-PCR detection showed that

serine/threonine-protein kinase was differentially expressed in seed dormancy in rice cv. N22 and less dormant mutant derivatives (Wu et al., 2016). Chen et al. (2016) found that the expression of six *WRKY* genes induced during endodormancy and decreased during ecodormancy, indicating that these six WRKY genes probable essential for bud dormancy in peach. However, the expression regularities of several coexpression pairs, such as *PasZFCHY* and *Pas*LEA2-22, *PasWRKY* and *Pas*LEA3-1, *Pas*Ser/Thr and *Pas*LEA2-2, and *PasUbI* and *Pas*LEA2-7, were rarely consistent, implying that these gene pairs may have different functions in different stages during the dormancy and germination periods of *P. armeniaca* × *P. sibirica.*

Protein-protein interaction network prediction analysis of PasLEAs and Y2H experiments suggested that PasLEA3-1 and XTH9 could occur protein protein interactions. Xyloglucan endotransglucosylase/hydrolases (XTHs) are enzymes involved in the modification of cell wall components, which are chief enzymes in cell wall remodeling. The relevance of XTHs in cell expansion and morphogenesis suggests a major role in stress responses. *O. sativa* L. Xyloglucan endotransglucosylase 9 (*OsXET9*) was highly induced in panicle and leaf at different developmental stages, in response to all cold, drought and heat stresses, especially in seedling and booting stage under cold stress. Some cis-elements related to stress responses were identified in putative promoter region of *OsXET9,* and this further confirms *OsXET9* is a novel gene involved in cold tolerance in rice in addition to the function in metabolism of cell wall (Dong et al., 2011). Lurlaro et al (2016) found that endotransglucosylase (XET) of XTHs of wheat seedlings may response to drought and heat stress, thus promoting to wheat survival and plant establishment. *DkXTH1* overexpression resulted in cells with more strength and thickness to maintain structural integrity, and thus enhanced tolerance to drought, salt, and ABA stresses and delayed fruit softening in transgenic *Arabidopsis* plants (Han et al., 2017). PasSMP-3 and EXPA22 can interacted with each other also proved by Y2H technique. Expansins are cell wall proteins which mediate cell wall loosening and extension. The need for cell wall loosening and the involvement of expansins during germination and flowering have been endorsed by several documents (Choi et al., 2013; Saito et al., 2015). The expression of Japanese pear expansin gene PpEXPA2 was low before the rapid enlargement stage of flower buds, and it increased rapidly during germination of buds (Saito et al., 2015). Choi et al., (2013) found that *OsEXP4* antisense rice plants were shorter and flowered earlier than the control plants. Wood formation related proteins XTH or EXPA can interact with LEA homologous gene, which proves that these proteins can form protein complexes with LEAs, involved in bud germination, flowering and cold stress. We speculate that cell wall structure expansion and modification (XTH or EXPA gene function) can be useful in improving the ability of material transportation and information transmission, thus may contribute to the improvement of cold resistance and flower bud germination.

Several studies revealed the possible LEAs function as stress proteins during protective reactions against cold conditions. For example, transcriptomic analysis and qRT-PCR showed that *VamLEA3-4* gene was upregulated by cold stress in Grapevine (Xu et al., 2020). *Chimonanthus praecox CpLEA5* gene transferred into *Arabidopsis thaliana* and enhance low-temperature tolerance in transgenic plants (Liu et al., 2015). Three genes (*PasLEA3-1*, *PasLEA3-2* and *PasLEA3-3*) homolous to Chimonanthus praecox *CpLEA5*, which showed different transcript levels in all four dormancy and sprouting stages of *P. armeniaca* × *P. sibirica*, were transferred into yeast cells and *Arabidopsis thaliana*. Interestingly, yeast cells expressing *PasLEA3-2* increased cold resistance and accumulation in yeast (Figure 6). Transgenic *A. thaliana* plants overexpressing *PasLEA3-2* showed higher tolerance to low-temperature conditions compared to WT plants (Figure 7). Grapevine *VamDHN3* gene was significantly upregulated in cold stresses, overexpression of *VamDHN3* enhanced the stability of the cell membrane in grapevine callus, and it indicated that less membrane injury suffered and cold tolerance enhanced in transgenic grapevine callus under cold stresses (Xu et al., 2020). In this study, *Pasdehydrin-3* is homolous to Grapevine *VamDHN3,* which exhibited high expression in the dormancy and sprouting stages in the flower buds of *P. armeniaca* × *P. sibirica*. *Pasdehydrin-3* transferred into yeast cells and exhibited greater enhanced growth

compared with that of yeast cells transformed with the empty vector on SG-U medium with cold treatment (Figure 6), and *Pasdehydrin-3* overexpressed in *Arabidopsis* conferred cold stress tolerance by improving antioxidative enzyme activities and increasing flower bud development related gene expression, such as *AtAP1*, *AtBTF3L*, *AtEBF1-1*, *AtLFY*, *AtPUB26*, and *AtRAV1* (Figure 9), and shortened the flowering time in transgenic plants (Figure 8e). Therefore, based on these results, we can draw a preliminary conclusion that *PasLEA3-2* and *Pasdehydrin-3* genes might present stronger cold tolerance and sustain of normal growth, also play a key role in regulating the dormancy and flowering stages, promoting flower buds dormancy release and early flowering under cold stress.

Low temperatures are a common threat to many plants in seedling, flowering and seed setting (Huang et al. 2005). Cold-resistant species, such as oilseed rape and wheat, exhibit low-temperature damage to leaves only after prolonged exposure to low temperatures with simultaneous high light conditions, and maintain intact plastids and tissue regeneration in leaf cells under normal cold conditions(Kratsch and Wise 2000). One explanation for the change in leaf morphology is a feedback to temperature, with smaller leaves being closer to the equilibrium temperature of the air than larger leaves (Leigh et al. 2017). In the present study, we found that low temperature treatment reduced both leaf length and leaf width in *Arabidopsis*. Interestingly, the leaf length and leaf width of plants overexpressing *PasLEA3-2* and *Pasdehydrin-3* were significantly higher compared to the wild type and were closer to those of *Arabidopsis* at normal temperature. There was no significant difference in leaf length and leaf width between wild-type and overexpression plants at normal temperature (Figure 8b, c). This suggests that *PasLEA3-2* and *Pasdehydrin-3* may assist plants to cope with cold stress by regulating leaf morphology under low temperature conditions. In a similar study, overexpression of *CpLEA5* was found to result in good growth in *Arabidopsis* at low temperatures (Liu et al. 2015). Root length, diameter, total volume and biomass of plants are reduced under cold stress (Tiwari et al. 2023). Under low temperature conditions, plants overexpressing *PasLEA3-2* and *Pasdehydrin-3* showed a significant increase in root length compared to normal plants (Figure 8a). Bolting and flowering are critical for crop yield. Factors affecting the timing of bolting and flowering in cruciferous plants include temperature, hormones, genetic factors and physiological and biochemical substances. The meristematic zone of the stem is particularly sensitive to low temperature status. As the duration of low temperature increases, bolting time and flowering time will gradually shorten (Song et al. 2019).

Overexpression of *PasLEA3-2* and *Pasdehydrin-3* in this study shortened the shortening of bolting time and flowering time caused by low temperature even more (Figure 8e, f). During seed uptake, low temperature inhibits sugar transport to reduce the accumulation of soluble sugars in the seed embryo, which in turn affects seed germination (Wang et al. 2018). The most intuitive effect of low temperature on *Arabidopsis* injury is the reduction of survival rate and seed setting rate, which is mitigated by overexpression of *PasLEA3-2* and *Pasdehydrin-3*. In summary, *PasLEA3-2* and *Pasdehydrin-3* play a positive role in plant cold tolerance, promote flowering and improves the seed setting rate, which is beneficial for the preservation of excellent germplasm resources and the reproduction of offspring under low temperatures.

5. Conclusions

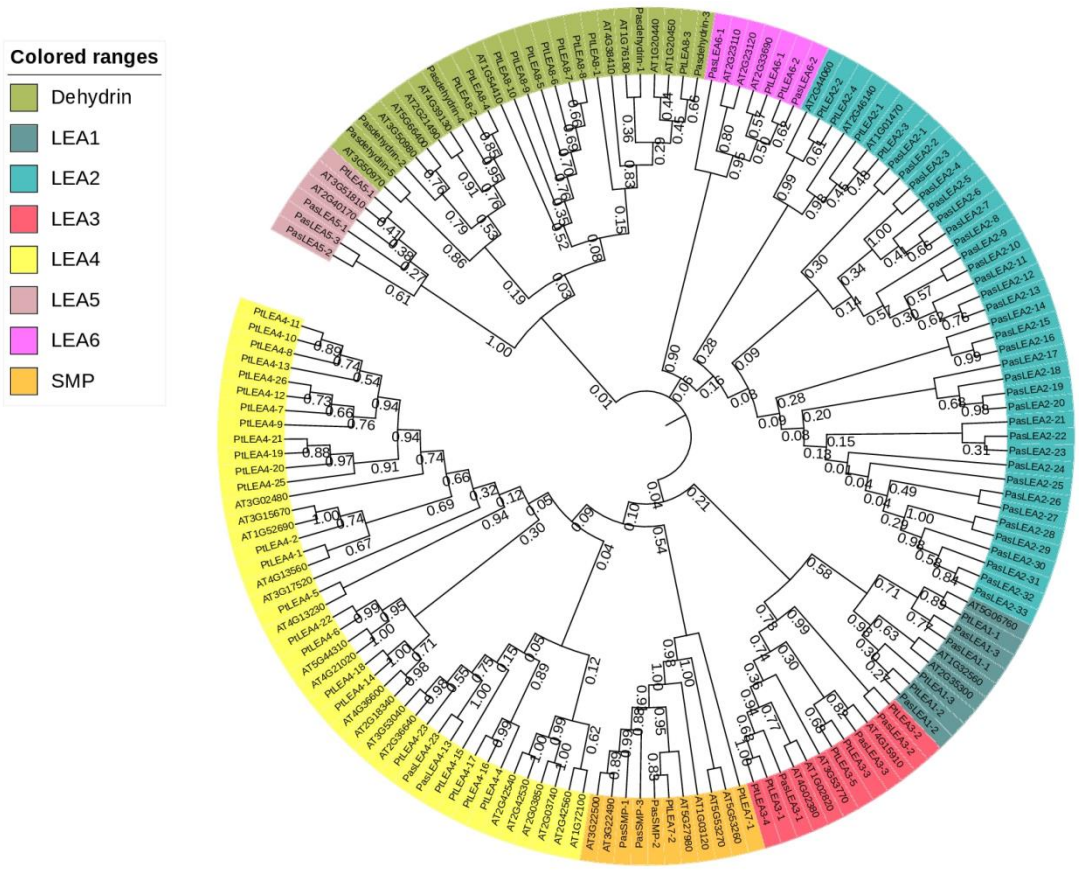
Taken together, this study demonstrates the first genome-wide research of the *P. armeniaca* × *P. sibirica LEA* gene family, 54 genes encoding LEAs (PasLEAs) were identified in the *P. armeniaca* × *P. sibirica* genome and phylogenetically summed up as eight different subgroups. Synteny analysis identified 14 collinear blocks containing *LEA* genes between *P. armeniaca* × *P. sibirica* and *A. thaliana* and 9 collinear blocks between *P. armeniaca* × *P. sibirica* and poplar. The chromosome localization, duplication event, duplication mechanism, protein features, motif composition, and 3D protein models of the *PasLEA* genes were further examined. Cis-acting regulatory elements analysis of the 2.0-kb sequences of the *LEA* promoter regions found several stress-, hormone-, and light-responsive and developmental period-specific elements included in the *PasLEAs*. The protein localization of

PasLEA1-2, PasLEA2-18, PasLEA3-1, PasLEA4-13, PasLEA5-2, PasLEA6-2, PasSMP-2, and Pasdehydrin-3 was detected, and the recombinant proteins was observed in various organelles of isolated *A. thaliana* protoplasts, suggesting potential functional diversity of PasLEAs in *P. armeniaca* × *P. sibirica*. Protein-protein interactions network showed the high correlation of PasLEA proteins with important cold resistance pathways, and specific protein interaction between PasLEA1-2 and PasERD4, PasLEA3-1 and PasXTH9, PasSMP-3 and PasEXPA22, and PasDehydrin-3 and PasCBF1 was confirmed by Y2H and BiFC assays. Gene expression analysis of *PasLEA* showed that *PasLEA2-20*, *PasLEA3-2*, *PasLEA6-1, Pasdehydrin-3* and *Pasdehydrin-5* were highly expressed in flower buds during the dormancy and sprouting periods of *P. armeniaca* × *P. sibirica.* Specially, *PasLEA3-2* and *Pasdehydrin-3* involved in cold stress tolerance in a yeast expression system was analyzed. Overexpression of *PasLEA3-2* and *Pasdehydrin-3* in *Arabidopsis* plants showed higher activity levels of antioxidative enzyme, enhanced the expression of flower development or cold-resistance genes and improved transgenic *A. thaliana* tolerance to cold stress. Research findings on upstream regulatory factors revealed that PasGBF4 and PasDOF3.5 can bind to the promoter of *PasLEA3-2*, and PasDOF2.4, PasDnaJ2 and PasAP2 can bind to *Pasdehydrin-3* promoter, and then regulate downstream gene expression. This study provides a valuable reference for further exploration of the *PasLEAs* functions and applications on genetically modified breeding of new varieties of *P. armeniaca* × *P. sibirica*.

Figure Captions

Figure 1. Phylogenetic analysis and synteny analyses of *LEA* genes among the *P. armeniaca* × *P. sibirica*, *P. trichocarpa*, and *A. thaliana* genomes. (a) LEA proteins were classified according to their homology analysis by using the maximum likely hood method with bootstrap analysis (1,000 replicates) based on MEGA 7.0 software. Denim blue, peacock blue, red, yellow, lavender, pink, beige, grass green color represent the *LEA1*, *LEA2*, *LEA3*, *LEA4*, *LEA5*, *LEA6, SMP* and *dehydrin* subfamily members of *P. armeniaca* × *P. sibirica*, *P. trichocarpa* and *A. thaliana*. The scale bar indicates the distance calculated by method of multiple alignment. (b) The chromosome numbers of these three species have different colors, and yellow, blue, and red represent the *P. trichocarpa, A. thaliana*, and *P. armeniaca* × *P. sibirica* chromosomes, respectively. The chromosome locate information is indicated on the inside with the chromosome sequence lengths in megabases. Gene pairs with syntenic relationships are connected by different color lines representing the gene linkage relationship between the *P. armeniaca* × *P. sibirica* and *P. trichocarpa* chromosomes and the *P. armeniaca* × *P. sibirica* and *A. thaliana* chromosomes, respectively.

(a)



(b)

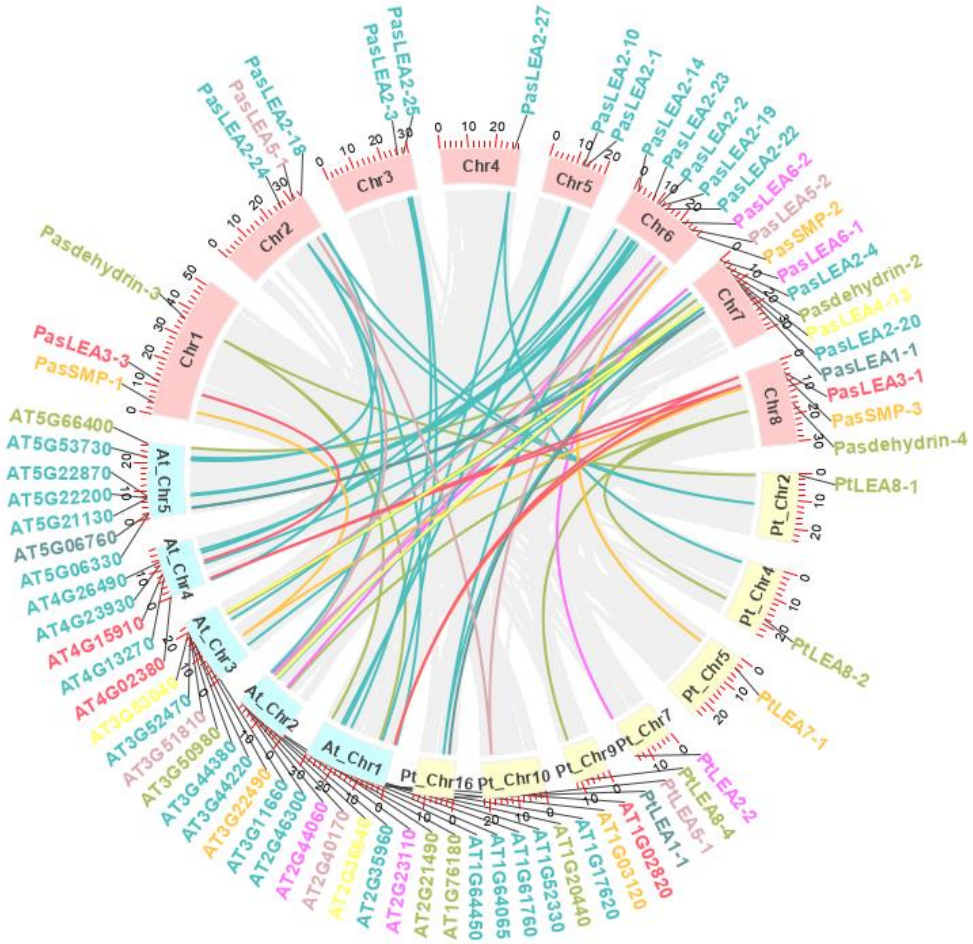


Figure 2. Phylogenetic relationships and hypothetical evolutionary progress of the clustering of *PasLEA* genes in *P. armeniaca* × *P. sibirica* chromosome. (a) Phylogenetic relationships (left) and putative evolutionary progress (right) of the clustering of *PasLEA* genes in cluster in *P. armeniaca* × *P. sibirica* chromosome 5. The letters S, and W represent putative segmental duplication and whole-genome duplication, respectively. (b) Phylogenetic relationships (left) and putative evolutionary progress (right) of the clustering of *PasLEA* genes in cluster in *P. armeniaca* × *P. sibirica* chromosome 2. The letters S, and W represent putative segmental duplication and whole-genome duplication, respectively. The letters T, S, and W represent putative tandem duplication, segmental duplication and whole-genome duplication, respectively.

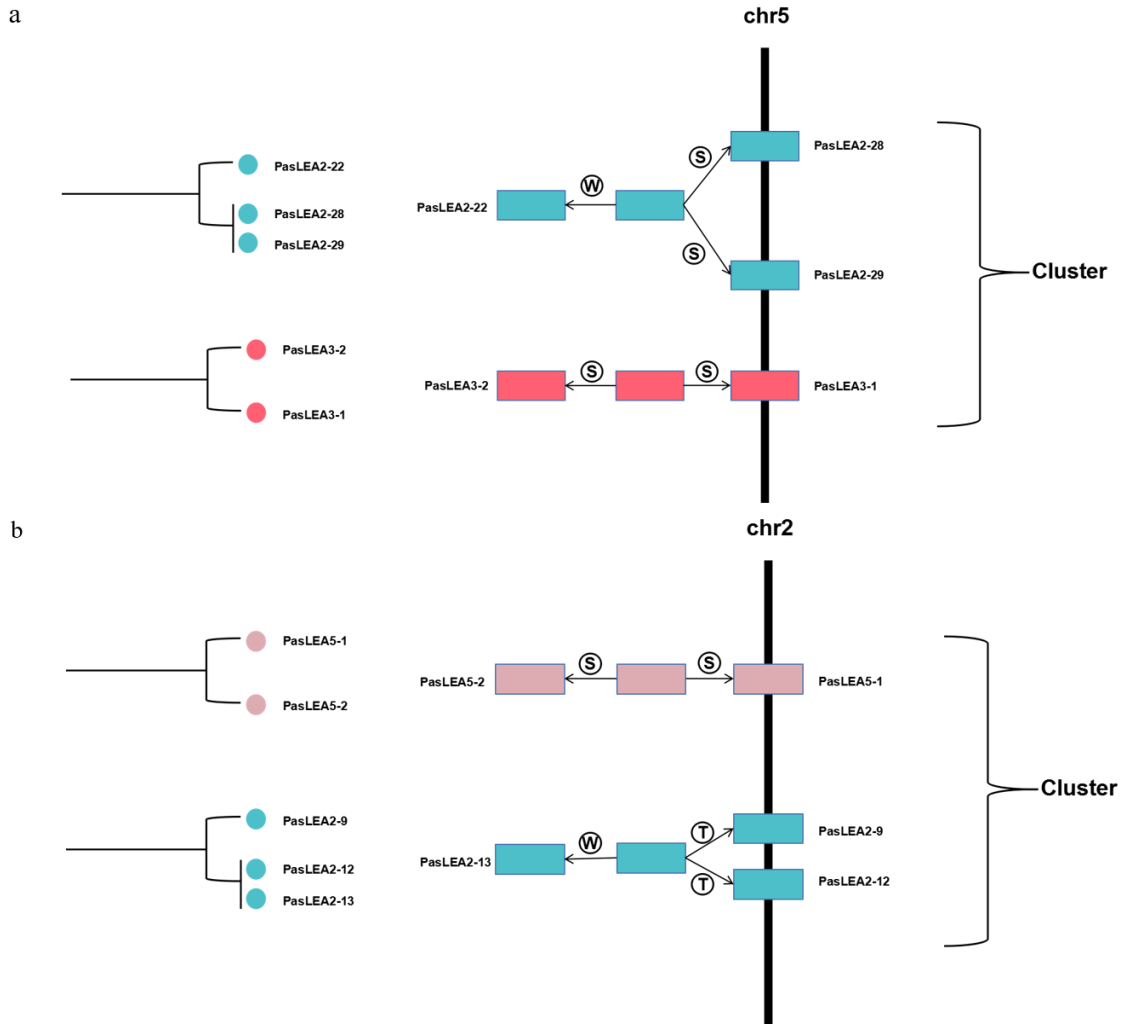


Figure 3. Interaction network of PasLEA proteins were predicted and confirmed based on the protein interactions of their homologous genes in *A. thaliana*. (a) The interaction network of PasLEA1-2. (b) The interaction network of PasLEA3-1. (c) The interaction network of PasSMP-3. (d) The interaction network of Pasdehydrin-3. (e) Y2H assays showing the protein interaction between PasLEA1-2 and PasERD4, PasLEA3-1 and PasXTH9, PasSMP-3 and PasEXPA22, and Pasdehydrin-3 and PasCBF1. The ability of yeast cells to grow on SD/–Leu/–Trp/–His/–Ade medium but containing 100 μM aureobasidin A and to turn blue in the presence of the X-α-Gal was scored as a positive interaction. (f) Interactions between PasLEA1-2 and PasERD4, PasLEA3-1 and XTH9, PasSMP-3 and EXPA22, and Pasdehydrin-3 and CBF1 proteins in the nucleus as shown by BiFC assay. Scale bar = 25 μm. STRING database was used to construct the interaction network. The red circles show the protein being queried, while the other circles represent the interacting proteins. The thickness of the lines represents the strength of data support.

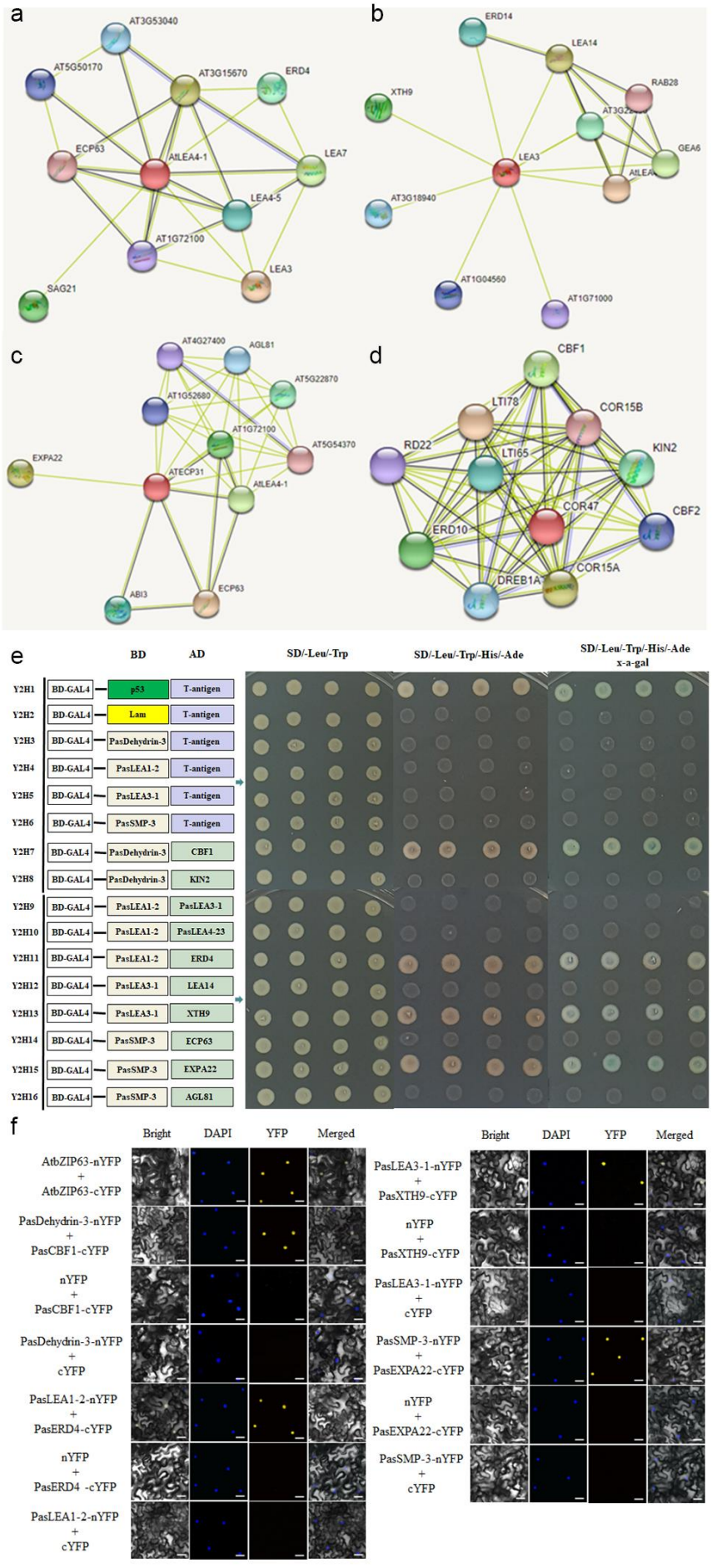

a
AT3G53040
AT5G50170
AT3G15670
ERD4
ECP63
AtLEA4-1
LEA7
LEA4-5
AT1G72100
LEA3
SAG21
b
ERD14
LEA14
RAB28
XTH9
GEA6
LEA3
AT3G18940
AT1G04560
AT1G71000
c
AT4G27400
AGL81
AT5G22870
AT1G52680
AT1G72100
AT5G54370
EXPA22
ATECP31
AtLEA4-1
ABI3
ECP63
d
CBF1
LTI78
COR15B
RD22
LTI65
KIN2
COR47
CBF2
ERD10
COR15A
DREB1A
e
BD
AD
SD/-Leu/-Trp
SD/-Leu/-Trp/-His/-Ade
SD/-Leu/-Trp/-His/-Ade x-a-gal
Y2H1 BD-GAL4 p53 T-antigen
Y2H2 BD-GAL4 Lam T-antigen
Y2H3 BD-GAL4 PasDehydrin-3 T-antigen
Y2H4 BD-GAL4 PasLEA1-2 T-antigen
Y2H5 BD-GAL4 PasLEA3-1 T-antigen
Y2H6 BD-GAL4 PasSMP-3 T-antigen
Y2H7 BD-GAL4 PasDehydrin-3 CBF1
Y2H8 BD-GAL4 PasDehydrin-3 KIN2
Y2H9 BD-GAL4 PasLEA1-2 PasLEA3-1
Y2H10 BD-GAL4 PasLEA1-2 PasLEA4-23
Y2H11 BD-GAL4 PasLEA1-2 ERD4
Y2H12 BD-GAL4 PasLEA3-1 LEA14
Y2H13 BD-GAL4 PasLEA3-1 XTH9
Y2H14 BD-GAL4 PasSMP-3 ECP63
Y2H15 BD-GAL4 PasSMP-3 EXPA22
Y2H16 BD-GAL4 PasSMP-3 AGL81
f
Bright
DAPI
YFP
Merged
AtbZIP63-nYFP + AtbZIP63-cYFP
PasDehydrin-3-nYFP + PasCBF1-cYFP
nYFP + PasCBF1-cYFP
PasDehydrin-3-nYFP + cYFP
PasLEA1-2-nYFP + PasERD4-cYFP
nYFP + PasERD4 -cYFP
PasLEA1-2-nYFP + cYFP
PasLEA3-1-nYFP + PasXTH9-cYFP
nYFP + PasXTH9-cYFP
PasLEA3-1-nYFP + cYFP
PasSMP-3-nYFP + PasEXPA22-cYFP
nYFP + PasEXPA22-cYFP
PasSMP-3-nYFP + cYFP

Figure 4. Expression patterns of 32 *PasLEA* genes in different tissues during the dormancy and sprouting periods of *P. armeniaca*×*P. sibirica*. The flower buds and stem were collected during the PD period, ED period, SP period, and GS period of three individual *P. armeniaca*×*P. sibirica* plants. The relative mRNA abundance of *PasLEA* genes were normalized to the reference gene expression of *PaElf* (elongation factor-1α). The $2^{-\Delta\Delta CT}$ method was chosen to calculate the relative gene expression levels. Mean values of data are based on three replicates ± SDs. The asterisk indicate the tested corresponding genes significantly up- or downregulated based on one-way ANOVA with Tukey's HSD post hoc analysis.

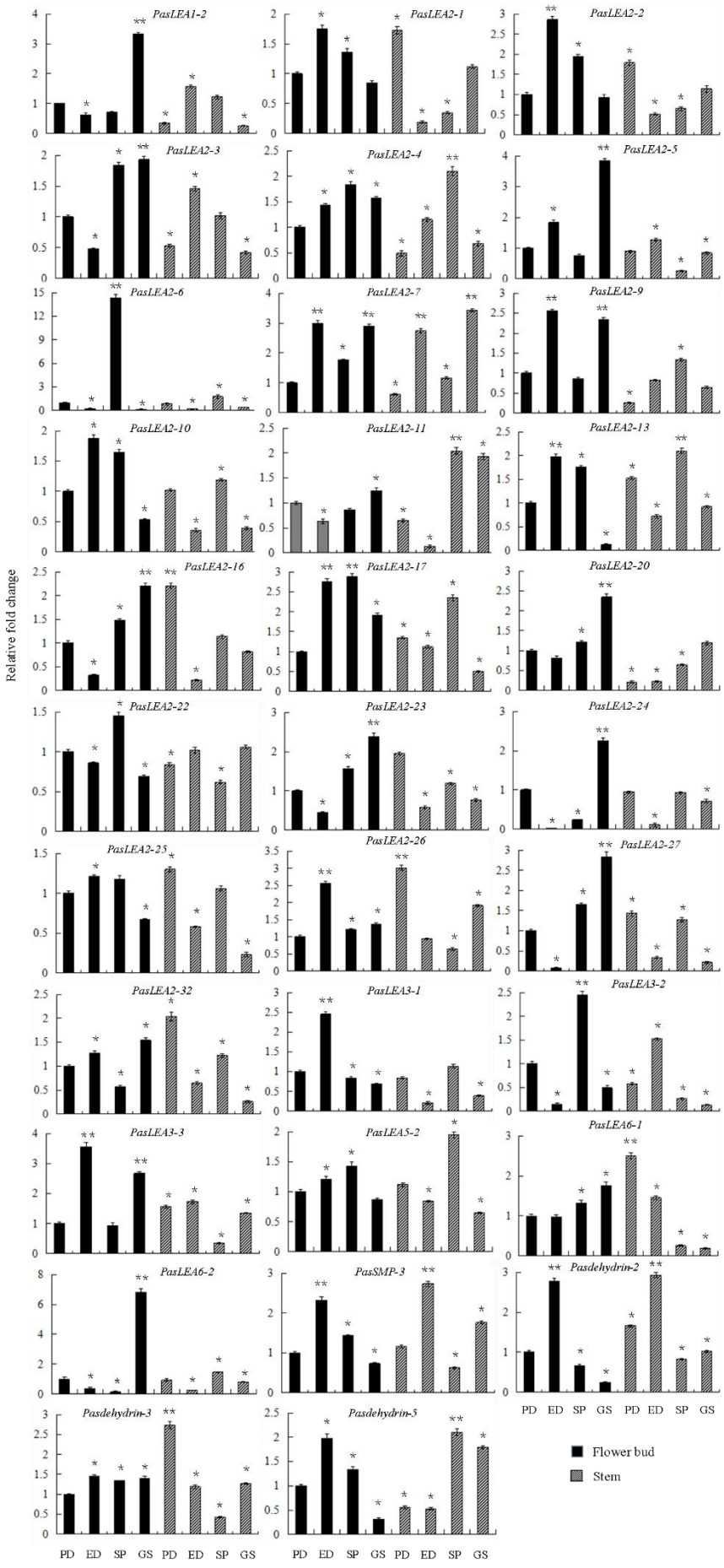

PasLEA1-2
PasLEA2-1
PasLEA2-2
PasLEA2-3
PasLEA2-4
PasLEA2-5
PasLEA2-6
PasLEA2-7
PasLEA2-9
PasLEA2-10
PasLEA2-11
PasLEA2-13
PasLEA2-16
PasLEA2-17
PasLEA2-20
PasLEA2-22
PasLEA2-23
PasLEA2-24
PasLEA2-25
PasLEA2-26
PasLEA2-27
PasLEA2-32
PasLEA3-1
PasLEA3-2
PasLEA3-3
PasLEA5-2
PasLEA6-1
PasLEA6-2
PasSMP-3
Pasdehydrin-2
Pasdehydrin-3
Pasdehydrin-5
Relative fold change
PD ED SP GS PD ED SP GS
Flower bud
Stem

Figure 5. Subcellular localization of PasLEA1-2, PasLEA2-18, PasLEA3-1, PasLEA4-13, PasLEA5-2, PasLEA6-2, PasSMP-2, and Pasdehydrin-3 in different cell organelles. Chloroplast, autofluorescence of the chloroplast. DAPI, the protoplasts are stained with 4′,6-diamidino-2-phenylindole (DAPI) to mark the nuclear. Bright, Bright-field images. Merge, overlap of GFP (green), DAPI (blue) and autofluorescence of the chloroplast (magenta). Row 1 shows the protoplasts expressing *GFP* vector alone. Row 2 shows protoplasts expressing the *PasLEA1-2*::*GFP* fusion protein with nucleus localization. Row 3 shows the protoplasts expressing the *PasLEA2-18*::*GFP* fusion protein with chloroplast localization. Row 4 shows protoplasts expressing the *PasLEA3-1*::*GFP* fusion protein with chloroplast localization. Row 5 shows the protoplasts expressing the *PasLEA4-13*::*GFP* fusion protein with nucleus localization. Row 6 shows the protoplasts expressing the *PasLEA5-2*::*GFP* fusion protein with nucleus localization. Row 7 shows protoplasts expressing the *PasLEA6-2*::*GFP* fusion protein with nucleus localization. Row 8 shows protoplasts expressing the *PasSMP-2*::*GFP* fusion protein with chloroplast and nucleus localization. Row 9 shows protoplasts expressing the *Pasdehydrin-3*::*GFP* fusion protein with nucleus localization. Bars = 5 μm.

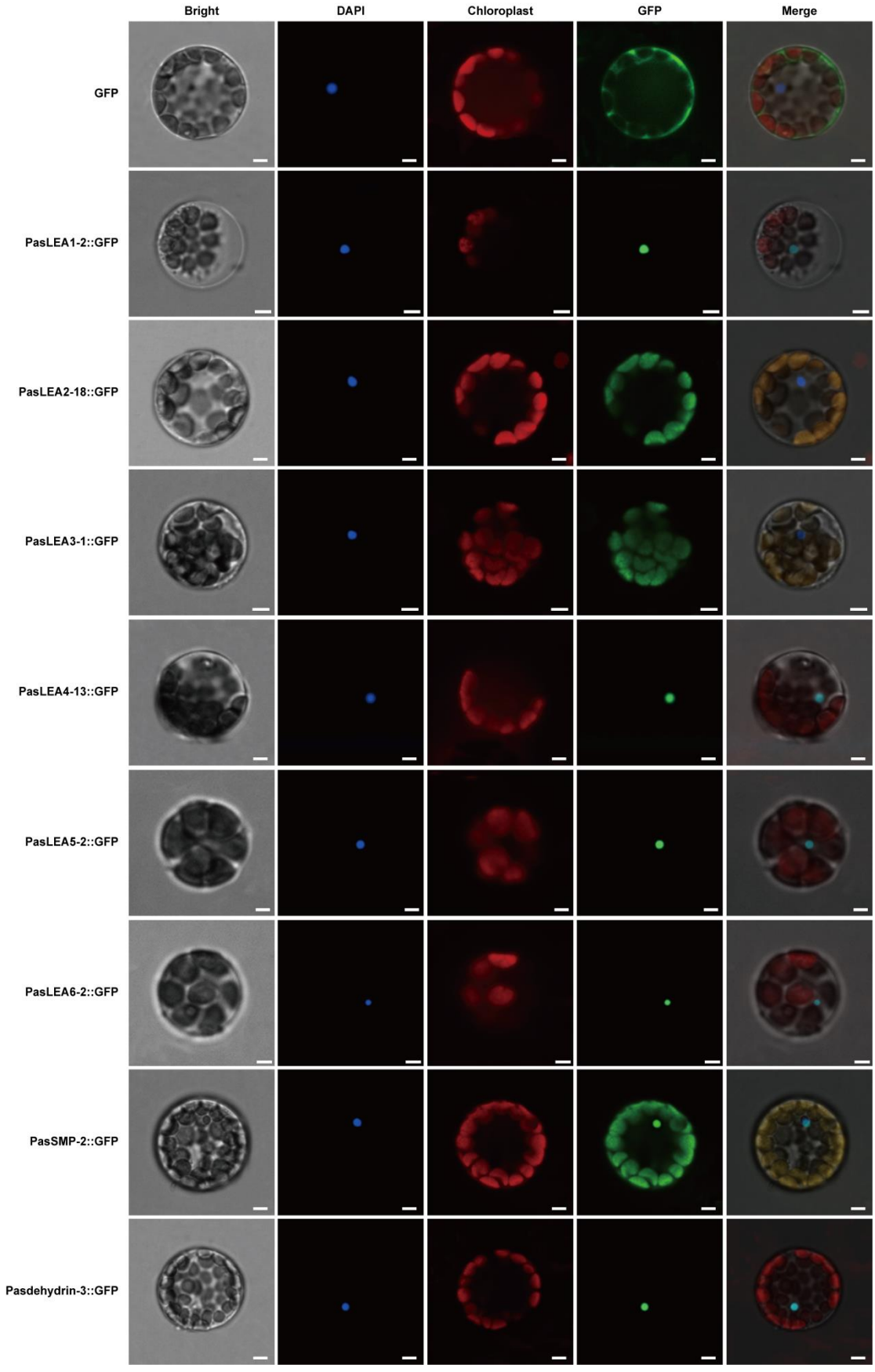
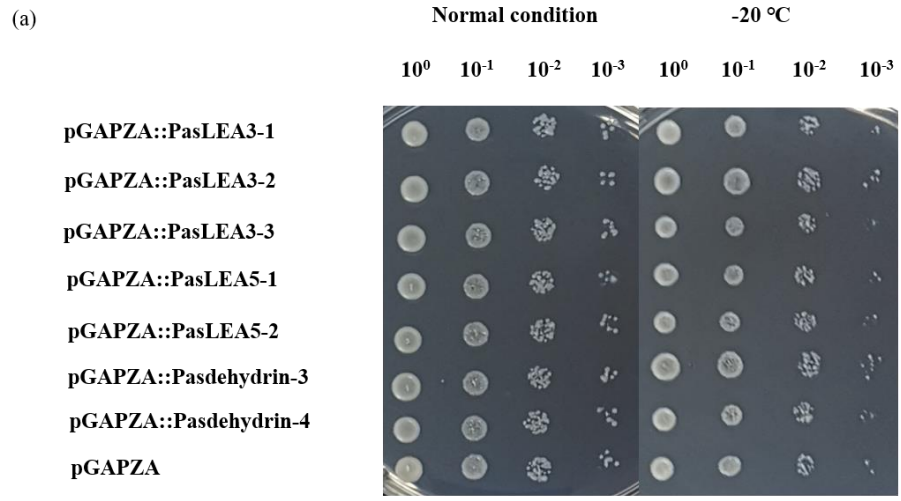

Bright
DAPI
Chloroplast
GFP
Merge
GFP
PasLEA1-2::GFP
PasLEA2-18::GFP
PasLEA3-1::GFP
PasLEA4-13::GFP
PasLEA5-2::GFP
PasLEA6-2::GFP
PasSMP-2::GFP
Pasdehydrin-3::GFP

Figure 6. Overexpression of *PasLEA3-2* and *Pasdehydrin-3* induces cold resistance and yeast accumulation in cold stress. (a) Growth of the GS115 yeast strain transformed with pGAPZA ligated *PasLEA3-1*, *PasLEA3-2*, *PasLEA3-3*, *PasLEA5-1*, *PasLEA5-2*, *Pasdehydrin-3* or *Pasdehydrin-4*. (b) OD600 value of *PasLEAs* transformed yeast and control yeast strain in response to cold stress. The data are means of three replicates ± SDs. n.s. means not significant. The different lowercase letters indicate statistical significance based on one-way ANOVA ($P < 0.05$).



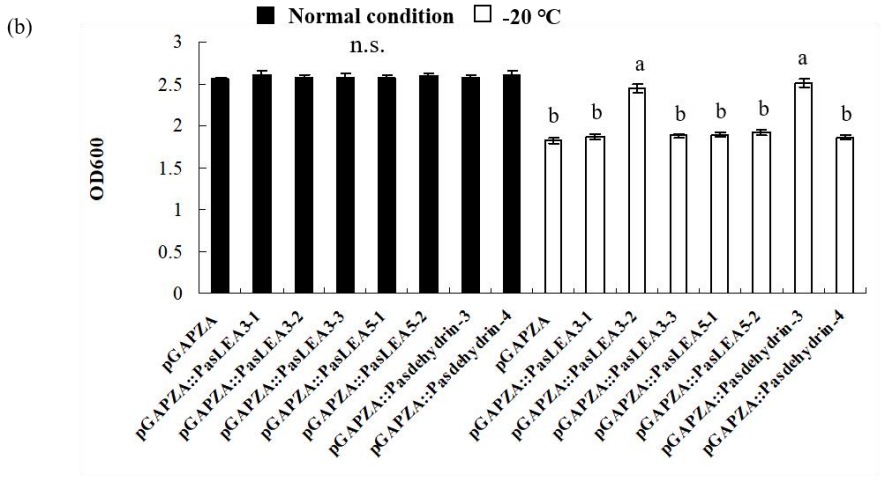


Figure 7. Cold resistance phenotypic analysis of *PasLEA3-2* and *Pasdehydrin-3* overexpressed *Arabidopsis* and wild-type plants. (a) Image of the transgenic plants and wild-type in growth chambers at 22°C, 16°C and 16°C for 16 h/4°C 8 h. Three independent *PasLEA3-2* transgenic *Arabidopsis* lines (LEAOE-1, LEAOE-2 and LEAOE-3) and three independent *Pasdehydrin-3* transgenic lines (DEHOE-1, DEHOE-2 and DEHOE-3), were identified by qRT-PCR. (b) After 10 days of treatment, the root length (upper row) and leaf phenotype (lower row) of *PasLEA3-2* or *Pasdehydrin-3* transgenic plants and wild-type *Arabidopsis* were examined. The scale bars correspond to 0.6 cm in upper row of (b), and correspond to 1 cm in lower row of (b). Pod length of *PasLEA3-2* (c) or *Pasdehydrin-3* (d) transgenic plants and wild-type *Arabidopsis* were analysized. Bars = 0.1 cm in (c) and (d). LEAOE-1-, LEAOE-2-, LEAOE-3-, and *PasLEA3-2*-OE transgenic *Arabidopsis* lines; DEHOE-1, DEHOE-2

and DEHOE-3-, and *Pasdehydrin-3*-OE transgenic *Arabidopsis* lines.

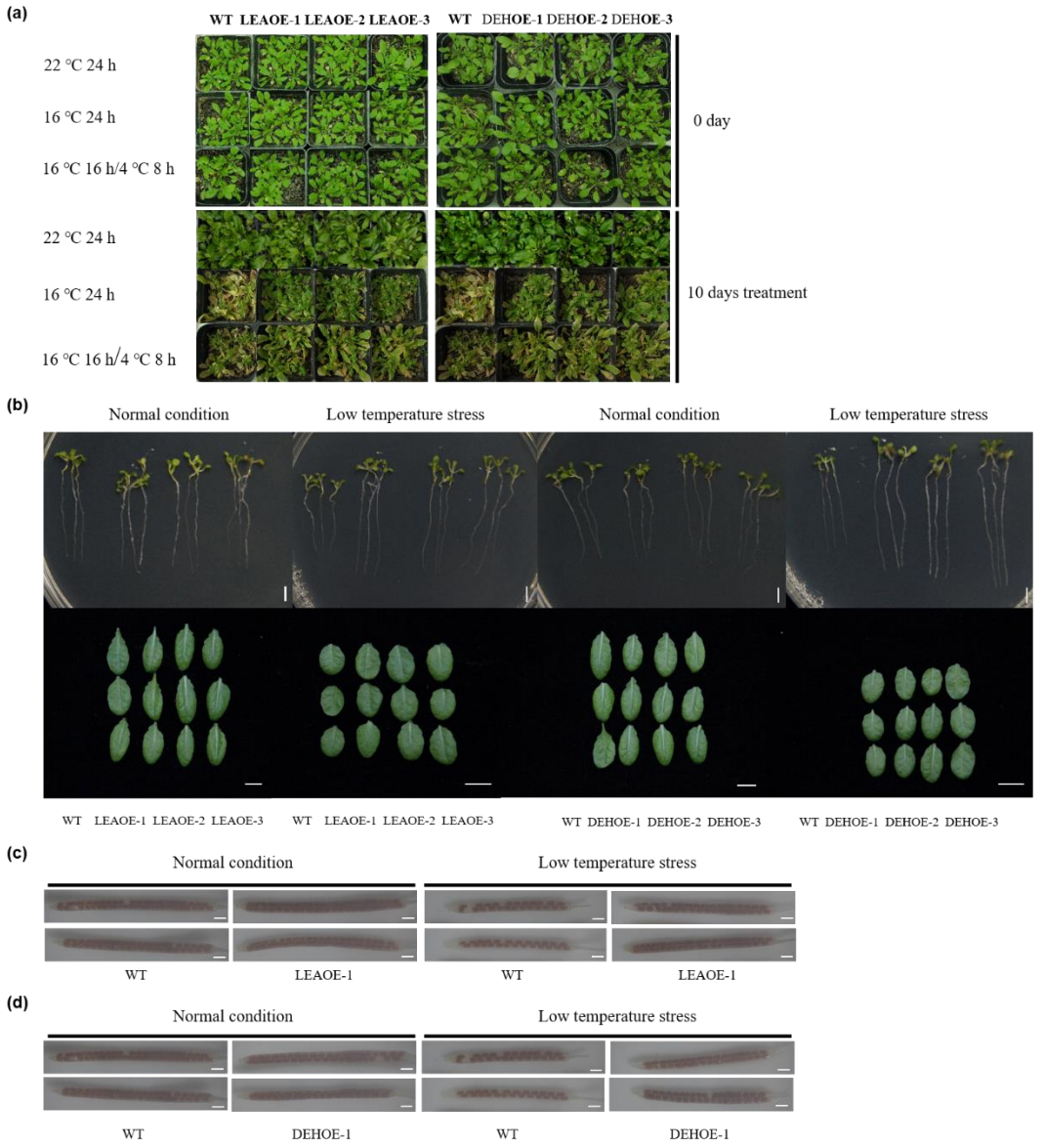


Figure 8. Phenotypic data of *PasLEA3-2* and *Pasdehydrin-3* overexpressed *Arabidopsis* and wild-type plants under low temperature stress. Root length (a), leaf length (b), leaf width (c), pod length (d), flowering time (e), bolting time (f), survival rates (g), and seed setting rate (h) of *PasLEA3-2* or *Pasdehydrin-3* overexpressed in *Arabidopsis* compared with the wild type against low temperature stresses. After 10 days of treatment, the SOD activity (i), proline content (j), MDA content (k) and relative electrical conductivity (l) of *PasLEA3-2* or *Pasdehydrin-3* transgenic plants and wild-type *Arabidopsis* were examined. LEAOE-1-, LEAOE-2-, LEAOE-3-, and *PasLEA3-2*-OE transgenic *Arabidopsis* lines; DEHOE-1, DEHOE-2 and DEHOE-3-, and *Pasdehydrin-3*-OE transgenic *Arabidopsis* lines. n.s. means not significant. Each bar represents the mean ± SD of three independent experiments. The lowercase letters represent significant differences at $P < 0.05$ (one-way ANOVA).

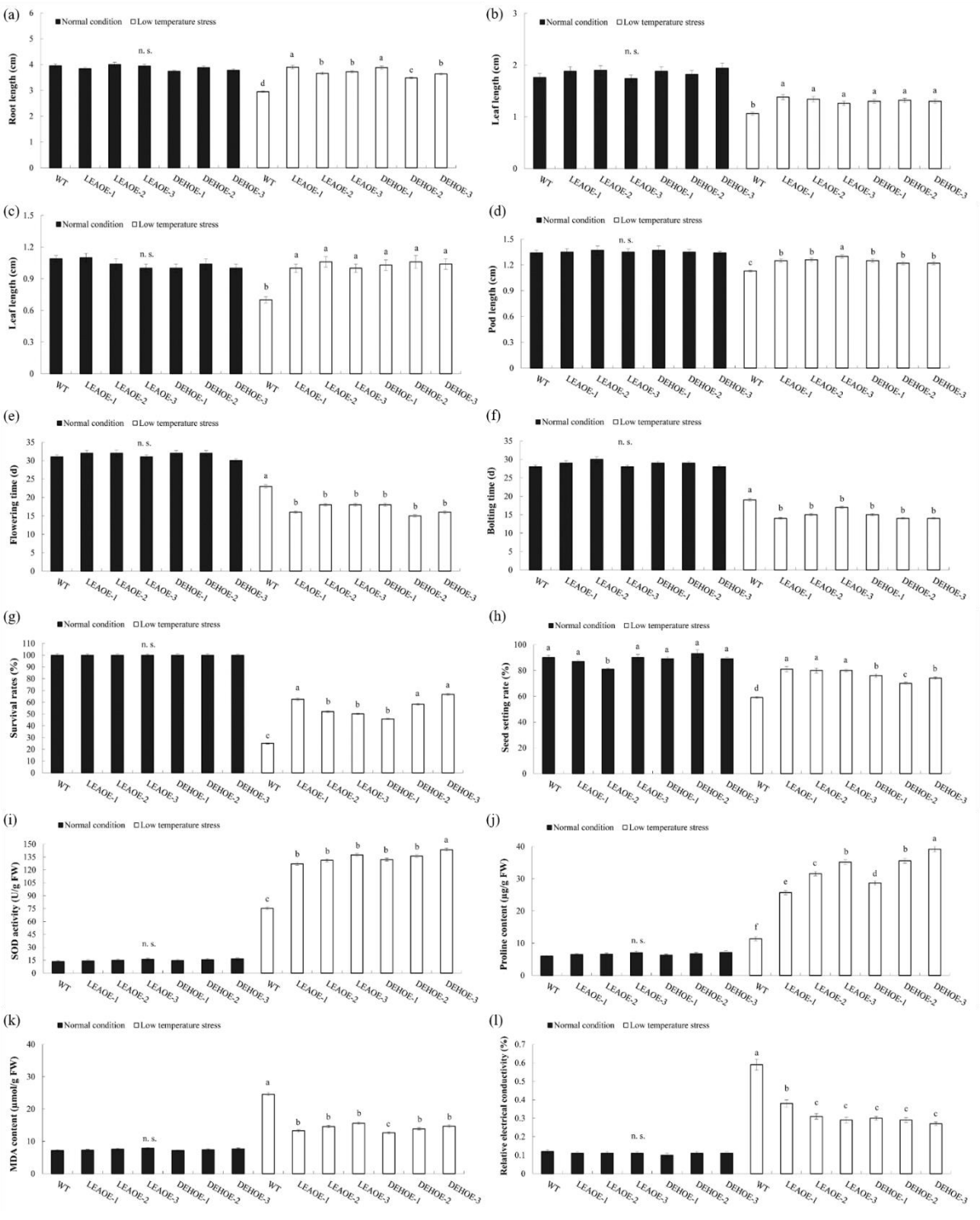


Figure 9. Yeast one-hybrid (Y1H) assay showing the interaction between PasGBF4, PasDOF3.5 and the *PasLEA3-2* promoter (pABAi-*PasLEA3-2*), and between PasDOF2.4, PasDnaJ2, PasAP2 and the *Pasdehydrin-3* promoter (pABAi-*Pasdehydrin-3*). The yeast strains were grown on SD/-Leu/-Ura +300 ng/ml Aureobasidin A (AbA) and SD/-Leu/-Ura +400 ng/ml ABA medium for 3 d.

No	Effector	Reporter	Interaction
1	AD-Rec-P53	pABAi-*p53*	Positive control
2	AD-EMPTY	pABAi	Negative control
3	AD-PasMYB1	pABAi-PasLEA3-2	−
4	AD-PasNFYB	pABAi-PasLEA3-2	−
5	AD-PasWRKY	pABAi-PasLEA3-2	−
6	AD-PasGBF4	pABAi-PasLEA3-2	+
7	AD-PasSOSEKI	pABAi-PasLEA3-2	−
8	AD-PasSPEECHLESS	pABAi-PasLEA3-2	−
9	AD-PasbHLH111	pABAi-PasLEA3-2	−
10	AD-PasDOF3.5	pABAi-PasLEA3-2	+
11	AD-PasDnaJ	pABAi-PasLEA3-2	−
12	AD-PasABI5.1	pABAi-PasLEA3-2	−
13	AD-EMPTY	pABAi -PasLEA3-2	Negative control
14	AD-Rec-P53	pABAi-*p53*	Positive control
15	AD-EMPTY	pABAi	Negative control
16	AD-PasMYB2	pABAi-Pasdehydrin-3	−
17	AD-PasDOF2.4	pABAi-Pasdehydrin-3	+
18	AD-PasDnaJ2	pABAi-Pasdehydrin-3	+
19	AD-PasMYB	pABAi-Pasdehydrin-3	−
20	AD-PasABI5.2	pABAi-Pasdehydrin-3	−
21	AD-PasCPN60-2	pABAi-Pasdehydrin-3	−
22	AD-PasARF17	pABAi-Pasdehydrin-3	−
23	AD-PasAP2	pABAi-Pasdehydrin-3	+
24	AD-EMPTY	pABAi-Pasdehydrin-3	Negative control

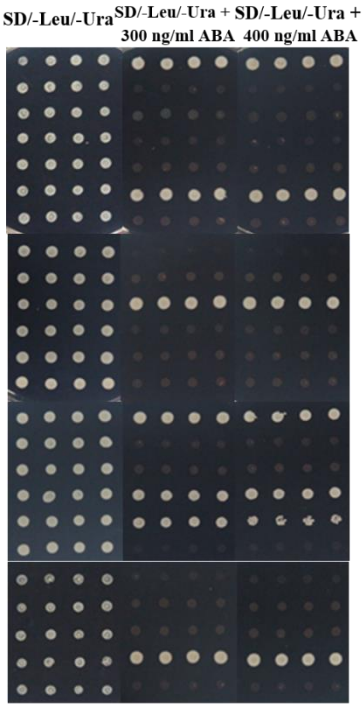


Table 1 Characteristics of 54 *PasLEA* genes in *P. armeniaca* × *P. sibirica*.

Table 2 The Ka/Ks ratios of *PasLEA* genes in *P. armeniaca* × *P. sibirica*.

Supplementary Materials

Table S1 Primer sets used in this study.

Table S2 Complete coding sequences and corresponding amino acid sequences of the *LEA* genes of *P. armeniaca* × *P. sibirica*.

Table S3 Predicted 3D structure of the 54 PasLEA proteins.

Table S4 Analysis of cis-acting elements in the 2,000-bp sequence upstream of the translation initiation codon of *PasLEA* genes.

Table S5 Predicted functional partners of PasLEA1-2, PasLEA3-1, PasSMP-3, and Pasdehydrin-3. Function of the top 10 highest scoring interacting proteins in each subfamily.

Table S6 Cold tolerance genes determined in the *PasLEA* coexpression network.

Table S7 Amino acid sequences of *LEA* genes in *Arabidopsis* and poplar.

Figure S1 Distribution of *PasLEA* genes on the eight *P. armeniaca* × *P. sibirica* chromosomes.

Figure S2 Exon-intron structures of the 54 *PasLEA* genes.

Figure S3 Distribution of conserved motifs in the PasLEA proteins.

Figure S4 Cis-acting elements analysis in the promoter of *PasLEA* genes.

Figure S5 Hierarchical clustering of the expression profiles of *PasLEA* family at three developmental stages.
Figure S6 Coexpression network of *PasLEAs*.

Figure S7 The expression levels of flowering related genes was analyzed in *PasLEA3-2* and *Pasdehydrin-3* overexpressed *Arabidopsis* and wild-type plants under low temperature stress.

Captions of supplementary figures

Figure S1. Distribution of *PasLEA* genes on the eight *P. armeniaca* × *P. sibirica* chromosomes. Forty-four *PasLEAs* are scattered across all eight *P. armeniaca* × *P. sibirica* chromosomes. The right side of the

chromosomal bar represents the *PasLEA* genes, and the corresponding positions on the chromosome (megabase pairs; Mb) are given on the left side. The *PasLEA1* genes are in denim blue, the *PasLEA2* genes are in peacock blue, the *PasLEA3* genes are in red, the *PasLEA4* genes are in yellow, the *PasLEA5* genes are in lavender, the *PasLEA6* genes are in pink, the *PasSMP* genes are in beige, and the *Pasdehydrin* genes are in grass green. Tandemly duplicated genes are underlined. Whole genome/segmental duplications are lined and underlined.

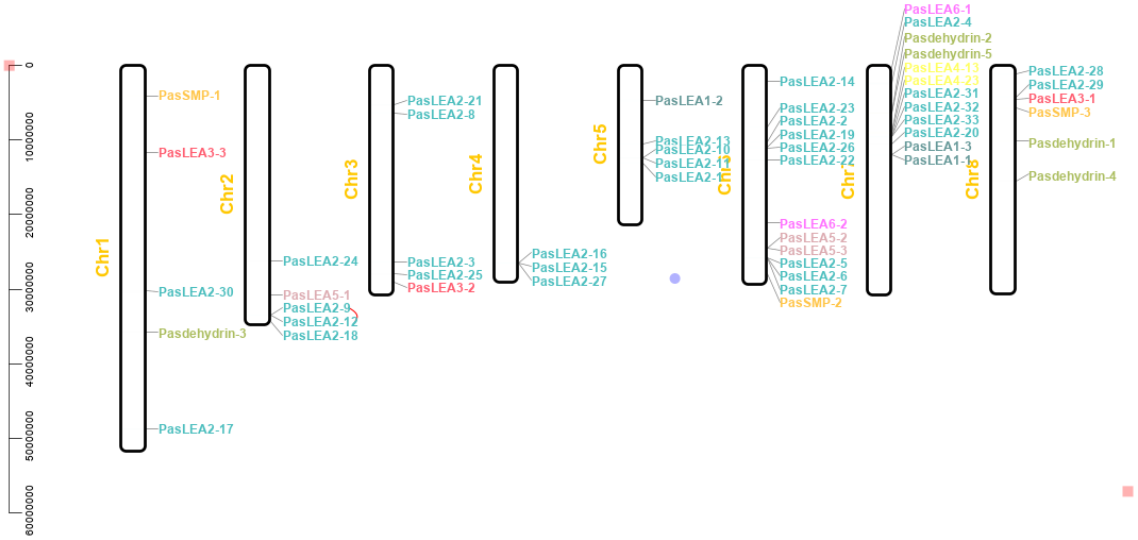


Figure S2. Exon-intron structures of the 54 *PasLEA* genes. The yellow boxes denote exons, and the lines denote introns. The exon-intron structure was determined via GSDS 2.0 (http://gsds.cbi.pku.edu.cn/) on the basis of gene and CDSs.

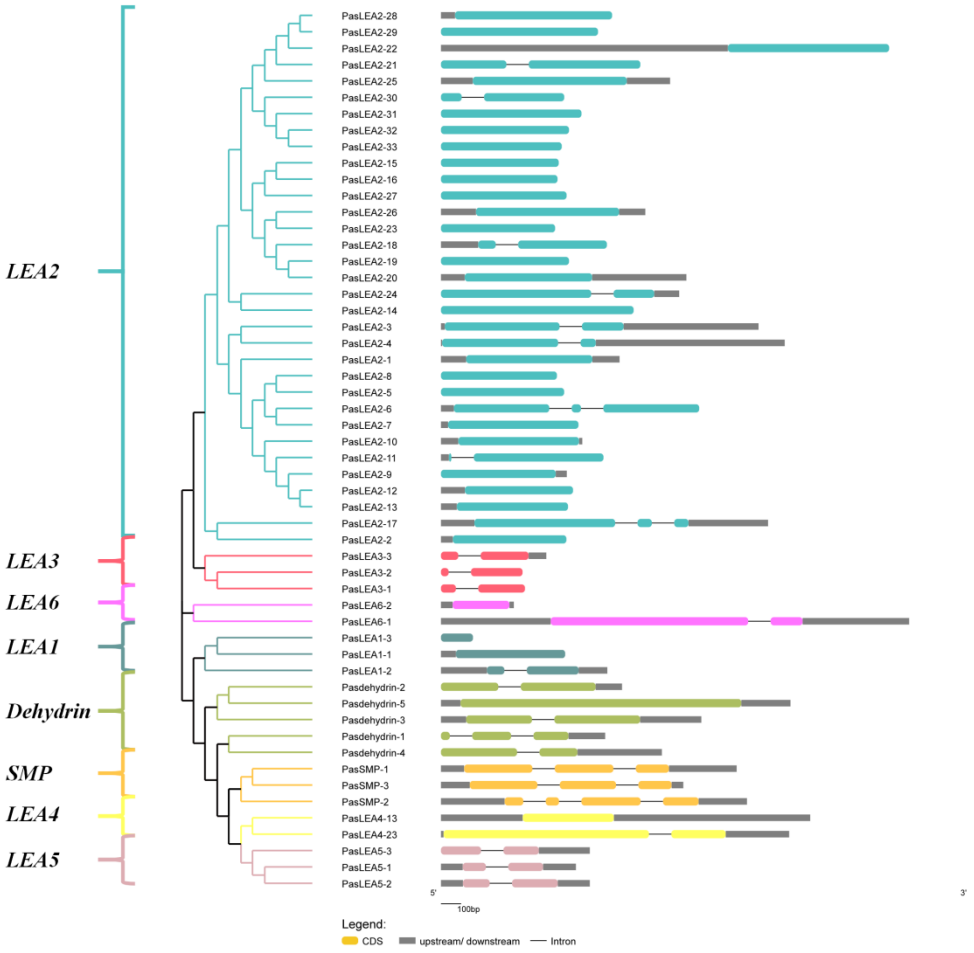


Figure S3. Distribution of conserved motifs in the PasLEA proteins. The motifs were identified via the online MEME program. The different colored boxes indicate different motifs at the corresponding positions.


LEA_1
Name p-value Motif Locations
PasLEA1-2 4.44e-79
PasLEA1-3 2.12e-58
PasLEA1-1 1.27e-96
Motif Symbol Motif Consensus
LEA_2
Name p-value Motif Locations
PasLEA2-17 2.15e-26
PasLEA2-30 6.58e-66
PasLEA2-9 5.34e-23
PasLEA2-12 2.13e-24
PasLEA2-18 1.43e-27
PasLEA2-24 1.89e-27
PasLEA2-21 4.03e-24
PasLEA2-8 4.81e-10
PasLEA2-3 6.68e-28
PasLEA2-25 4.93e-55
PasLEA2-16 7.39e-25
PasLEA2-15 1.90e-28
PasLEA2-27 1.38e-32
PasLEA2-10 5.05e-18
PasLEA2-11 2.25e-17
PasLEA2-1 4.83e-40
PasLEA2-13 2.50e-29
Motif Symbol Motif Consensus
LEA_3
Name p-value Motif Locations
PasLEA3-3 1.45e-72
PasLEA3-2 5.54e-63
PasLEA3-1 3.37e-62
Motif Symbol Motif Consensus
LEA_4
Name p-value Motif Locations
PasLEA4-13 1.22e-200
PasLEA4-23 8.57e-196
Motif Symbol Motif Consensus
LEA_5
Name p-value Motif Locations
PasLEA5-1 2.88e-100
PasLEA5-2 6.59e-119
PasLEA5-3 1.84e-120
Motif Symbol Motif Consensus
LEA_6
Name p-value Motif Locations
PasLEA6-2 1.86e-84
PasLEA6-1 8.33e-76
Motif Symbol Motif Consensus
Dehydrin
Name p-value Motif Locations
Pasdehydrin-3 6.23e-85
Pasdehydrin-2 2.95e-87
Pasdehydrin-5 4.11e-88
Pasdehydrin-1 3.66e-71
Pasdehydrin-4 2.27e-72
Motif Symbol Motif Consensus
SMP
Name p-value Motif Locations
PasSMP-1 6.30e-160
PasSMP-2 5.89e-153
PasSMP-3 4.62e-162
Motif Symbol Motif Consensus
Motif-pfam03760
Motif-pfam03168
Motif-pfam03242
Motif-pfam02987
Motif-pfam00477
Motif-pfam10714
Motif-pfam04927
Motif-pfam00257

Figure S4. Cis-acting elements analysis in the promoter of *PasLEA* genes. The numbers of MBS (CAACTG), ARE (AAACCA), MRE (AACCTAA), ABRE (ACGTG), and other *cis*-acting regulatory elements in the promoter region of *P. armeniaca* × *P. sibirica PasLEA* genes from each *LEA* family. The cis-acting elements were searched in the 2-kb upstream of the translation initiation site using the PLACE database.

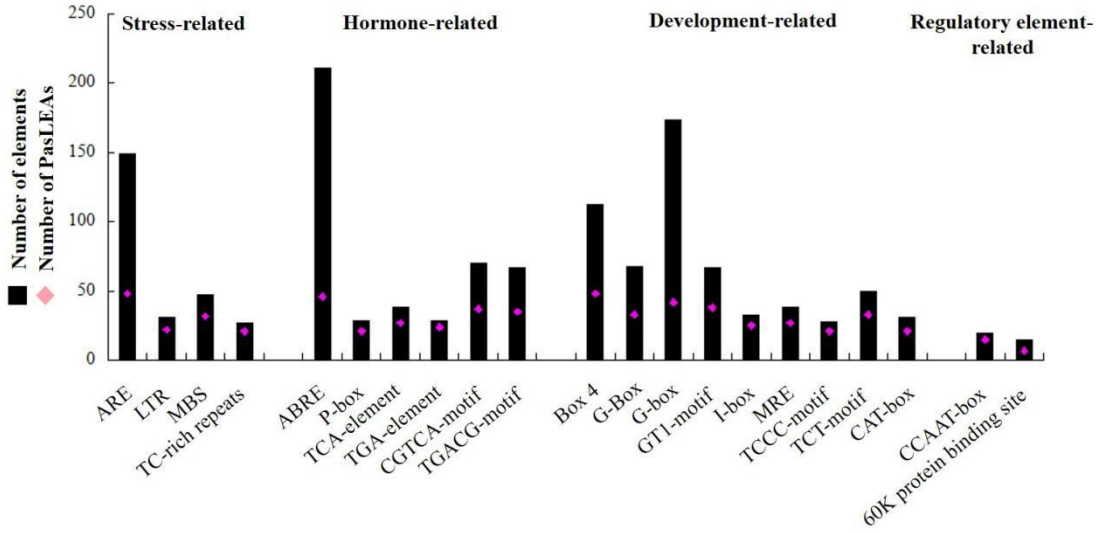


Figure S5. Hierarchical clustering of the expression profiles of *PasLEA* family at three developmental stages. ED, SP, and GS indicate that the experimental materials were collected during ecological dormancy, the sprouting period, and germination stage of *P. armeniaca* × *P. sibirica*, respectively. The FPKM value are recorded in the middle of each box.

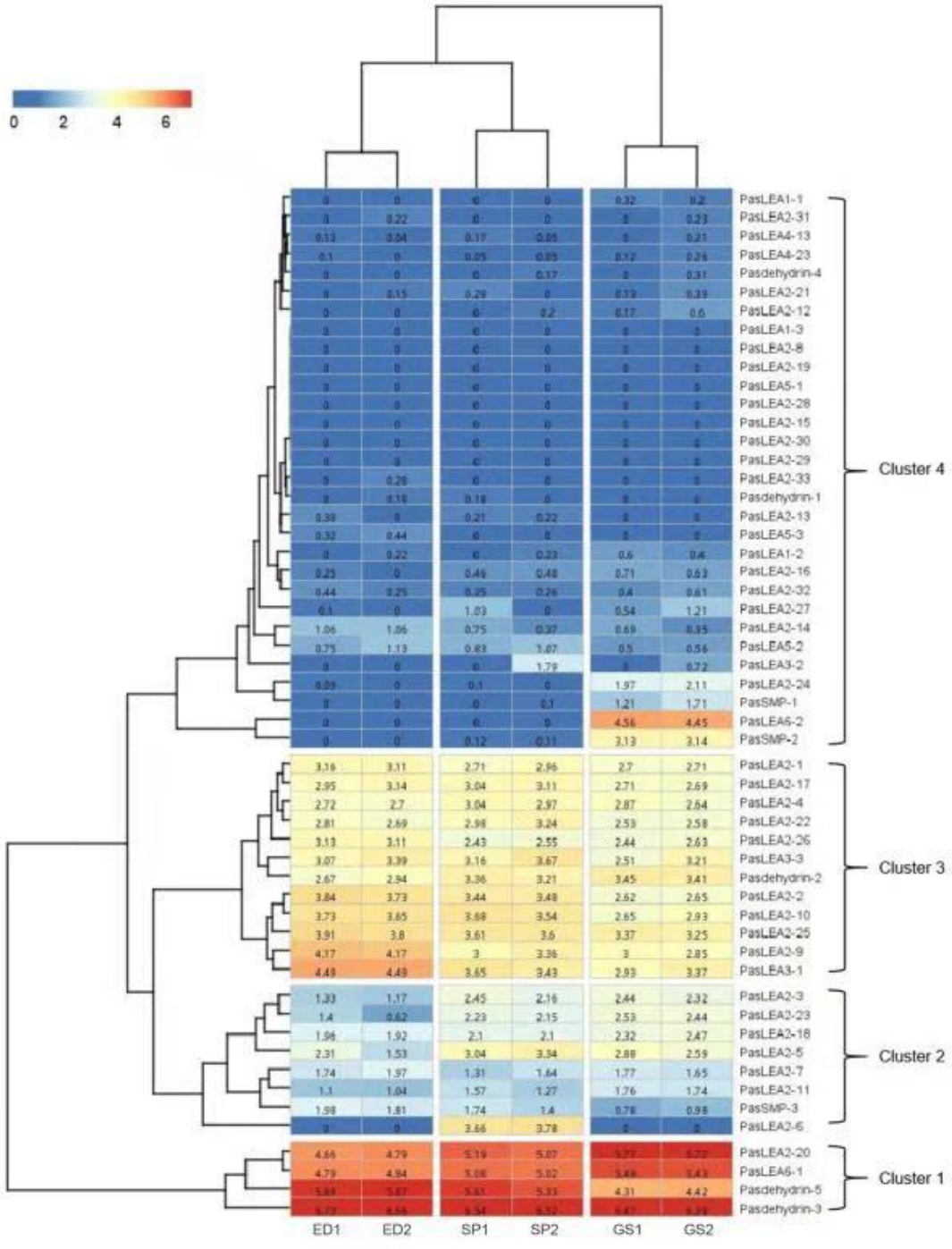
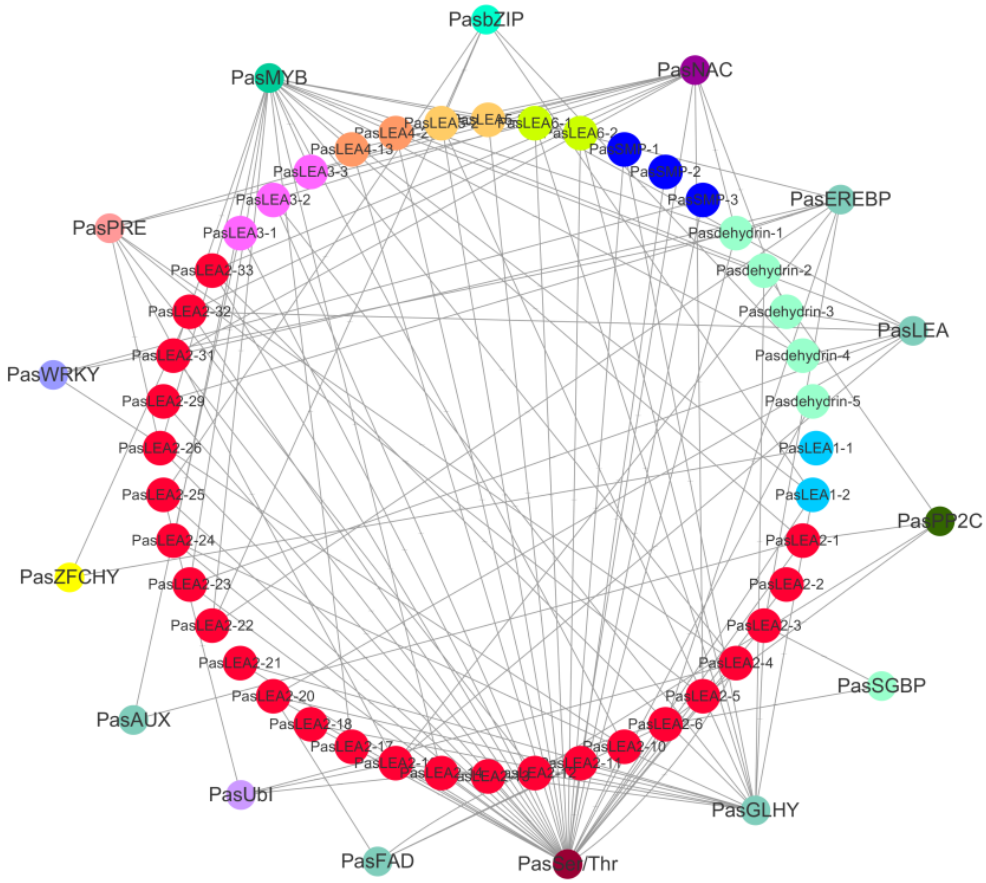


Figure S6. Coexpression network of *PasLEAs*. (a) A total of 2350 coexpression pairs were coexpressed with *PasLEAs*, among which 15 genes are involved in cold tolerance based on previous research. (b) Transcript levels of 15 genes involved in cold resistance in stem and flower buds during the PD period, ED period, SP period, and GS period of *P. armeniaca*×*P. sibirica*. *PasLEA* gene expression levels were normalized to the reference gene of *PaElf*. Each bar represents the mean ± SD of three replicates. The a, b, c, d letters standard for significant differences at $P < 0.05$ (one-way ANOVA).

(a)


PasbZIP
PasMYB
PasNAC
PasLEA4-2
PasLEA3-5
PasLEA6-1
PasLEA6-2
PasLEA4-13
PasLEA3-3
PasLEA3-2
PasEREBP
PasPRE
PasLEA3-1
Pasdehydrin-1
PasLEA2-33
Pasdehydrin-2
PasLEA2-32
Pasdehydrin-3
PasLEA
PasLEA2-31
Pasdehydrin-4
PasWRKY
PasLEA2-29
Pasdehydrin-5
PasLEA2-26
PasLEA1-1
PasLEA2-25
PasLEA1-2
PasLEA2-24
PasLEA2-1
PasZFCHY
PasLEA2-23
PasLEA2-2
PasLEA2-22
PasLEA2-3
PasLEA2-21
PasLEA2-4
PasSGBP
PasLEA2-20
PasLEA2-5
PasAUX
PasLEA2-18
PasLEA2-6
PasLEA2-17
PasLEA2-10
PasUbl
PasGLHY
PasFAD

(b)

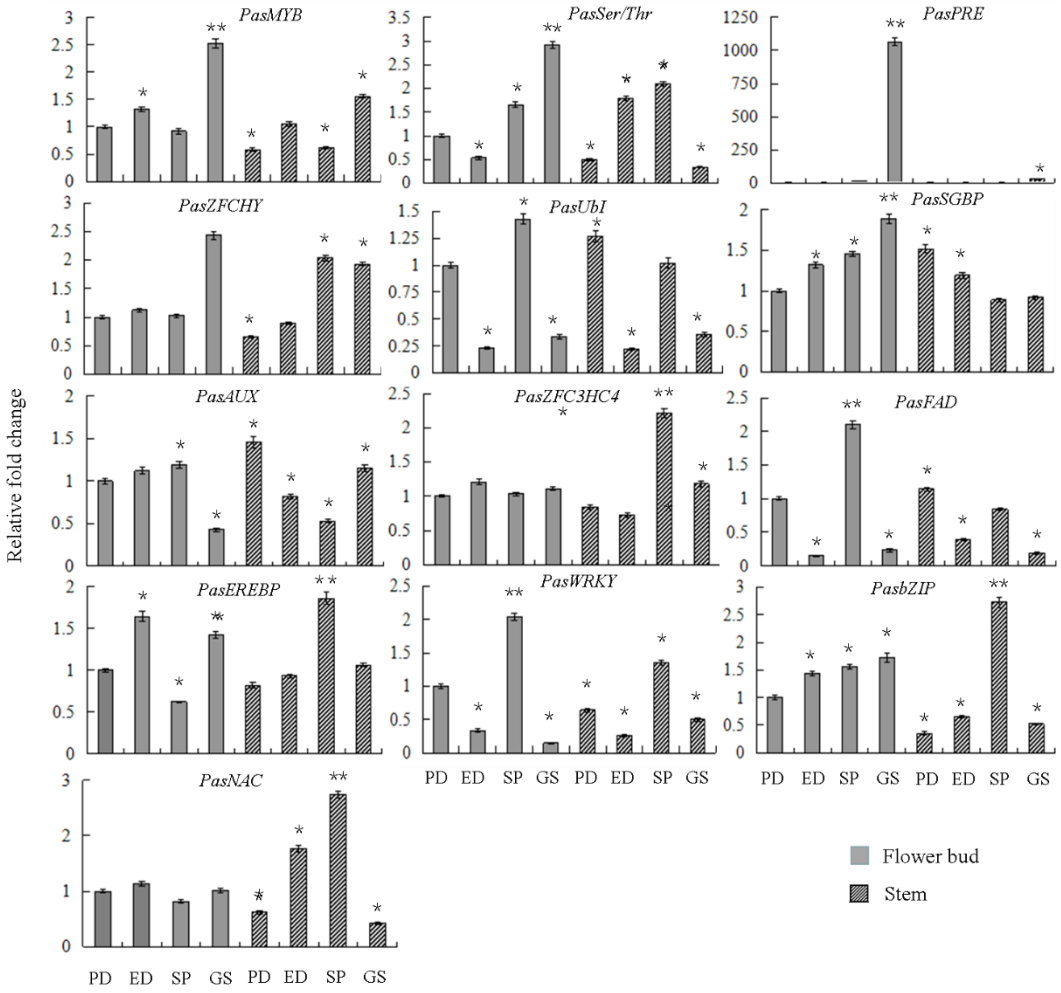
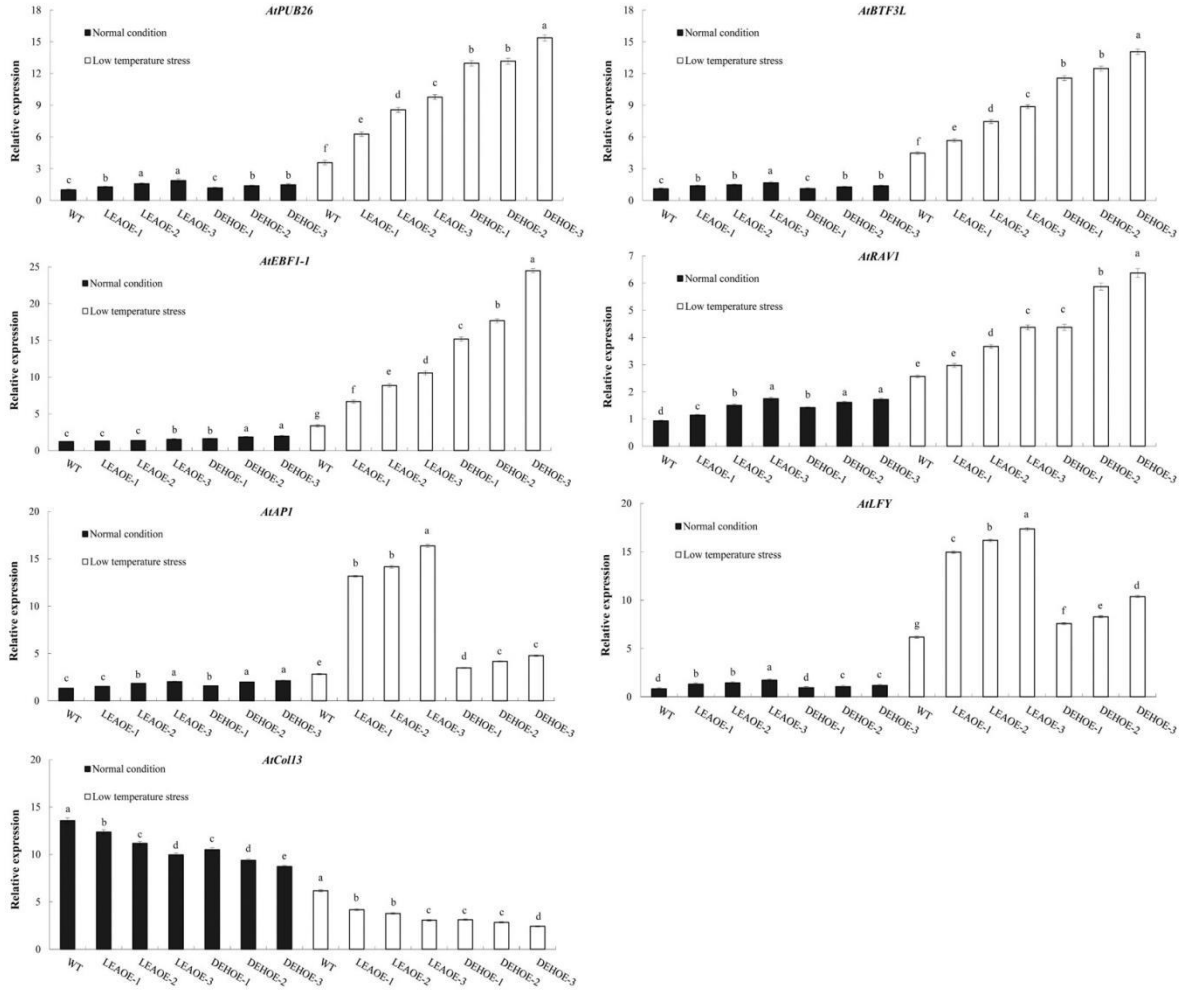
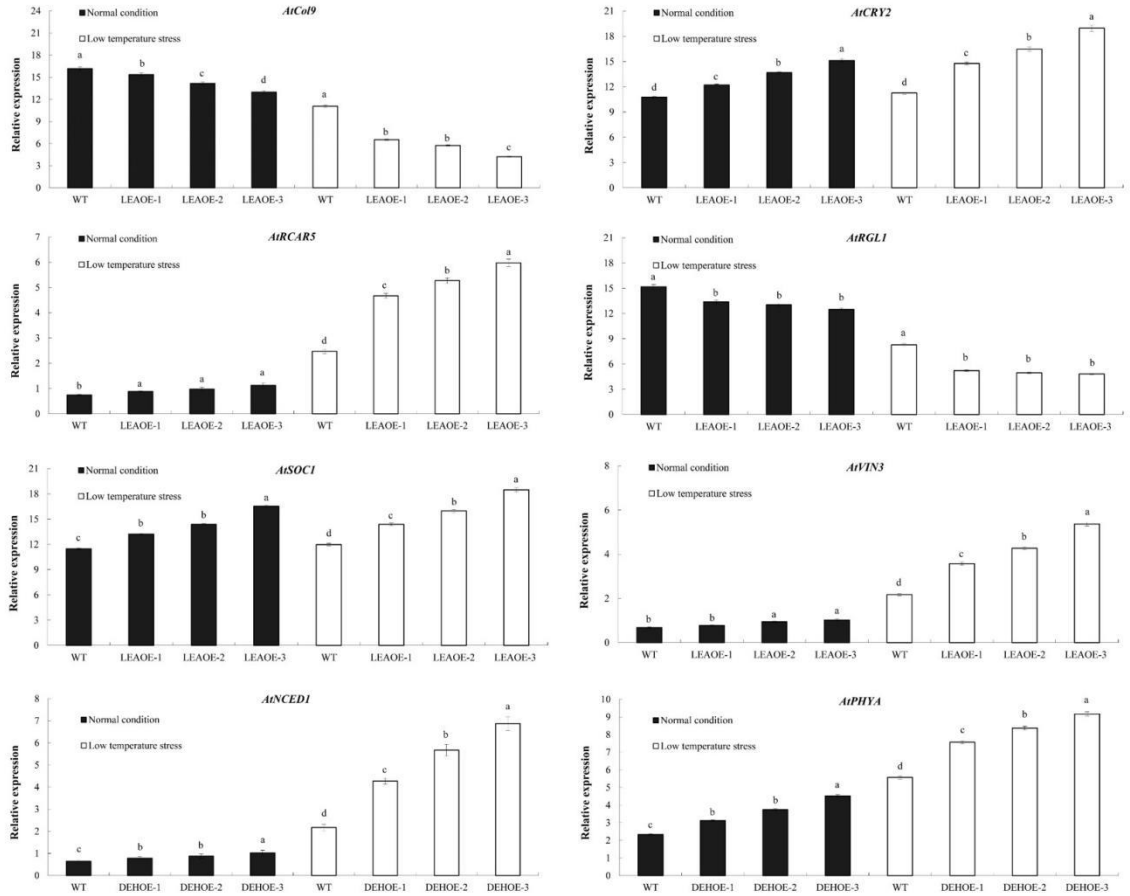


Figure S7. The expression levels of flowering related genes was analyzed in *PasLEA3-2* and *Pasdehydrin-3* overexpressed *Arabidopsis* and wild-type plants under low temperature stress. (a) *AtAP1*, *AtBTF3L*, *AtEBF1-1, AtLFY, AtPUB26*, *AtGl* and *AtRAV1* genes were upregulated both in *PasLEA3-2* (LEA1, LEA2, and LEA3) and *Pasdehydrin-3* overexpressing *Arabidopsis* plants. (b) *AtCRY2*, *AtRCAR5*, *AtSOC1* and *AtVIN3* were upregulated in *PasLEA3-2* overexpressing *Arabidopsis* lines, and *AtNCED1* and *AtPHYA* were upregulated in *Pasdehydrin-3* overexpressing lines. Also we detected *AtCol9* and *AtRGL1* were downregulated in *PasLEA3-2* overexpressing plants. *Actin* was used as an internal control.The expression level of each gene in the wild type was set to 1. Error bars represent SD of three biological repetitions.

(a)


AtPUB26
AtBTF3L
AtEBF1-1
AtRAV1
AtAP1
AtLFY
AtCol13
Normal condition
Low temperature stress
Relative expression
WT
LEAOE-1
LEAOE-2
LEAOE-3
DEHOE-1
DEHOE-2
DEHOE-3

(b)



Author contributions

S.L and Z.T conceived and designed the research. S.L, X.X, and F.W performed all the experiments. P.W, S.W, and G.Z contributed reagents/materials/ analysis tools. J.Z, Y.Z, and H.T analysed the data. S.L wrote the paper. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare that they have no conflicts of interests.

Funding source

This work was supported by the Fundamental Research Funds of CAF (CAFYBB2018MA003 and CAFYBB2020QD001), the National Natural Science Foundation of China (31770705 and 31400570).

References

Amara I, Odena A, Oliveira E, Moreno A, Masmoudi K, Pagès M, Goday A. 2012. Insights into Maize LEA Proteins: From Proteomics to Functional Approaches. *Plant Cell Physiol* **53:** 312–329.

Bateman A. 2010. The Pfam protein families database. *Nucleic Acids Res* **38:** D211-D222.

Bies-Ethève N, Gaubier-Comella P, Debures A, Lasserre E, Jobet E, Raynal M, Cooke R, Delseny M. 2008. Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in Arabidopsis thaliana. *Plant Molecular Biology* **67(1-2):** p107-124.

Boursiac Y, Prak S, Boudet J, Postaire O, Luu DT, Tournaire-Roux C, et al. 2008. The response of *Arabidopsis* root water transport to a challenging environment implicates reactive oxygen species- and phosphorylation dependent internalization of aquaporins. *Plant Signal Behav* **3:** 1096-1098.

Cai G, Xu Y, Zhang S, Chen T, Liu G, Li Z, Zhu Y, Wang G. 2022. A tomato chloroplast-targeted DnaJ protein, SlDnaJ20 maintains the stability of photosystem I/II under chilling stress. *Plant Signal Behav* **17(1):** 2139116.

Candat A, Paszkiewicz G, Neveu M, Gautier R, Logan DC, Avelange-Macherel M, Macherel D. 2014. The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in arabidopsis offers tailored protection against abiotic stress. *The Plant Cell* **26:** 3148-3166.

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. 2018. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* **13(8):** 1194-1202.

Chen, M, Tan, Q, Sun, M, Li, D, Fu, X, Chen, X, et al. 2016. Genome-wide identification of WRKY family genes in peach and analysis of *WRKY* expression during bud dormancy. *Mol Genet Genomics* **291:** 1319-1332.

Choi D, Lee Y, Cho HT, Kende H. 2003. Regulation of expansin gene expression affects growth and development in transgenic rice plants. *Plant Cell* **15(6):** 1386-1398.

Deokar AA, Bunyamin T. 2016. Genome-wide analysis of the aquaporin gene family in Chickpea (*Cicer arietinum* L.). *Front Plant Sci* **7:** 1802.

Ding Y, Jia Y, Shi Y, Zhang X, Song C, Gong Z, Yang S. 2018. OST1-mediated BTF3L phosphorylation positively regulates CBFs during plant cold responses. *The EMBO Journal* **37(8):** e98228.

Dong J, Jiang Y, Chen R, Xu Z, Gao X. 2011. Isolation of a novel xyloglucan endotransglucosylase (*OsXET9*) gene from rice and analysis of the response of this gene to abiotic stresses. *African Journal of Biotechnology* **10(76):** 17424-17434.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Finn RD, Clements J, Eddy SR. 2011. HMMER Web server: interactive sequence similarity searching. *Nucleic Acids Res* **39:** W29-37.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I , Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* **31:** 3784-3788.

Han Y, Han S, Ban Q, He Y, Jin M, Rao J. 2017. Overexpression of persimmon *DkXTH1* enhanced tolerance to abiotic stress and delayed fruit softening in transgenic plants. *Plant Cell Rep* **36:** 583-596.

Hernández-Sánchez IE, Maruri-López I, Ferrando A, Carbonell J, Graether SP and Jiménez-Bremont JF. 2015. Nuclear localization of the dehydrin OpsDHN1 is determined by histidine-rich motif. *Front Plant Sci* **6:** 702.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35:** W585-W587.

Hu G, Wang K, Huang B, Mila I, Frasse P, Maza E, Djari A, Hernould M, Zouine M, Li Z, Bouzayen M. 2022. The auxin-responsive transcription factor SlDOF9 regulates inflorescence and flower development in tomato. *Nature Plants* **(4):8,** 419-433.

Huang L, Ye Z, Bell RW, Dell B. 2005. Boron nutrition and chilling tolerance of warm climate crop species. *Annals of botany* **96:** 755-767.

Kratsch HA, Wise RR. 2000. The ultrastructure of chilling stress. *Plant, Cell & Environment* **23:** 337-350.

Kumar S, Stecher G, Tamura K. 2016. MEGA7, Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33:** 1870.

Lan T, Gao J, Zeng QY. 2013. Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in Populus trichocarpa. *Tree Genetics & Genomes* **9(1):** 253-264.

Lee TH, Tang H, Wang X, Paterson AH. 2012. PGDD: a database of gene and genome duplication in plants, *Nucleic Acids Res* **41:** 1152-1158.

Leigh A, Sevanto S, Close JD, Nicotra AB. 2017. The influence of leaf size and shape on leaf thermal dynamics: does theory hold up under natural conditions? *Plant, Cell & Environment* **40:** 237-248.

Letunic I, Doerks T, and Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* **40:** 302-305.

Leubner-Metzger G, Petruzzelli L, Waldvogel R, Vögeli-Lange R, Meins F. 1998. Ethylene-responsive element binding protein (EREBP) expression and the transcriptional regulation of class I β-1,3-glucanase during tobacco seed germination. *Plant Molecular Biology* **38(5):** 785-795.

Liu J, Meng Q, Xiang H, Shi F, Ma L, Li Y, Liu C, Liu Y, Su B. 2021. Genome-wide analysis of Dof transcription factors and their response to cold stress in rice (Oryza sativa L.). *BMC Genomics* **22(1):** 800.

Liu Y, Wang L, Xing X, Sun L, Pan J, Kong X, Zhang M, LiD. 2013. ZmLEA3, a multifunctional group 3 LEA protein from maize (Zea mays L.), is involved in biotic and abiotic stresses. *Plant & Cell Physiology* **6:** 944-959.

Liu Y, Xie L, Liang X, Zhang S. 2015. CpLEA5, the Late Embryogenesis Abundant Protein Gene from Chimonanthus praecox, Possesses Low Temperature and Osmotic Resistances in Prokaryote and Eukaryotes. *Int J Mol Sci* **16:** 26978-26990.

Lurlaro A, De Caroli M, Sabella E, De Pascali M, Rampino P, De Bellis L, Perrotta C, Dalessandro G, Piro G, Fry SC, Lenucci MS. 2016. Drought and heat differentially affect XTH expression and XET activity and action in 3-Day-Old seedlings of durum wheat cultivars with different stress susceptibility. *Front Plant Sci* **7:** 1686.

Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, Zhou Z, Zhang Z, Salih H, Wang K, Liu F. 2018. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *BMC Genetics* **19(1):** 6.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43:** D222-D226.

Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. *Plant physiol* **171:** 2294-2316.

Pedrosa AM , Martins C, Gonçalves LP, Costa M. 2015. 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* **10(12):** e0145785.

Ramegowda V, Gill US, SivalingamPN, Gupta A, Gupta C, Govind G, Nataraja KN, Pereira A, Udayakumar M, Mysore KS, Senthil-Kumar M. 2017. GBF3 transcription factor imparts drought tolerance in Arabidopsis thaliana. *Scientific Reports* **22;7(1):** 9148.

Saito T, Tuan PA, Katsumi-Horigane A, Bai S, Ito A, Sekiyama Y, Ono H, Moriguchi T. 2015. Development of flower buds in the Japanese pear (*Pyrus pyrifolia*) from late autumn to early spring. *Tree Physiology* **35(6):** 653.

Song S-w, Lei Y-l, Huang X-m, Su W, Chen R-y, Hao Y-w. 2019. Crosstalk of cold and gibberellin effects on bolting and flowering in flowering Chinese cabbage. *Journal of Integrative Agriculture* **18:** 992-1000.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* **320:** 486-488.

Tang H, Bowers JE, Wang X, Paterson AH. 2009. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci U.S.A.* **107(1):** 472-477.

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res* **18:** 1944–1954.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673-4680.

Tiwari M, Kumar R, Subramanian S, Doherty CJ, Jagadish SVK. 2023. Auxin–cytokinin interplay shapes root functionality under low-temperature stress. *Trends in plant science* **28:** 447-459.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science* **313:** 1596-1604.

Wang Y, Cui Y, Hu G, Wang X, Chen H, Shi Q, Xiang J, Zhang Y, Zhu D, Zhang Y. 2018. Reduced bioactive gibberellin content in rice seeds under low temperature leads to decreased sugar consumption and low seed germination rates. *Plant Physiology and Biochemistry* **133:** 1-10.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40:** e49.

Wu J, Jin Y, Liu C, Vonapartis E, Liang J, Wu W, Gazzarrini S, He J, Yi M. 2018. GhNAC83 inhibits corm dormancy release by regulating ABA signaling and CK biosynthesis in *Gladiolus hybridus*. *Journal of Experimental Botany* **70(4):** 1221-1237.

Wu T, Yang C, Ding B, Feng Z, Wang Q, He J, Tong J, Xiao L, Jiang L, Wan J. 2016. Microarray-based gene expression analysis of strong seed dormancy in rice cv. N22 and less dormant mutant derivatives. *Plant Physiology & Biochemistry* **99:** 27-38.

Xu M, Tong Q, Wang Y, Wang Z, Xu G, Elias G, Li S, Liang Z. 2020. Transcriptomic Analysis of the Grapevine LEAGene Family in Response to Osmotic and Cold Stress Reveals a Key Role for *VamDHN3*. *Plant Cell Physiol* **61(4):** 775-786.

Zhao Y, Yun JF, Shi FM, Wang JJ, Yang QC, Chao YH. 2010. Molecular cloning and characterization of a group 3 LEA gene from agropyron mongolicum keng. *African Journal of Biotechnology* **9(37):** 6040-6048.