

SCIENTIFIC REPORTS



OPEN

Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabuliformis* (Pinaceae) in *Escherichia coli*

Received: 23 September 2015

Accepted: 14 December 2015

Published: 19 January 2016

Jie Gao¹ & Ting Lan²

Late embryogenesis abundant (LEA) proteins are a large and highly diverse gene family present in a wide range of plant species. LEAs are proposed to play a role in various stress tolerance responses. Our study represents the first-ever survey of LEA proteins and their encoding genes in a widely distributed pine (*Pinus tabuliformis*) in China. Twenty-three LEA genes were identified from the *P. tabuliformis* belonging to seven groups. Proteins with repeated motifs are an important feature specific to LEA groups. Ten of 23 pine LEA genes were selectively expressed in specific tissues, and showed expression divergence within each group. In addition, we selected 13 genes representing each group and introduced these genes into *Escherichia coli* to assess the protective function of PtaLEA under heat and salt stresses. Compared with control cells, the *E. coli* cells expressing PtaLEA fusion protein exhibited enhanced salt and heat resistance and viability, indicating the protein may play a protective role in cells under stress conditions. Furthermore, among these enhanced tolerance genes, a certain extent of function divergence appeared within a gene group as well as between gene groups, suggesting potential functional diversity of this gene family in conifers.

Abiotic stresses, such as drought, salinity, and extreme temperature are limiting factors for normal plant growth and development. High salinity upsets homeostasis in water potential and ion distribution at both cellular and whole plant levels¹. It results in molecular damages, arrest of development and even mortality². High temperature can cause irreversible damages including protein denaturation, aggregation and degradation³, as well as protein synthesis inhibition and enzymes inactivation. Extreme temperature can also increase fluidity of membrane lipids⁴ which could lead to loss of membrane integrity^{5–7}. Higher plants have developed several genetic responses to these environmental factors⁸. An important group is the late embryogenesis abundant (LEA) protein gene family which functions in cellular protection during abiotic stress tolerance^{9,10}. Since the first discovery of the LEA proteins in cotton and wheat^{11,12}, the LEA proteins have been described throughout the entire plant kingdom, as well as in other organisms, from invertebrate to prokaryotes^{13,14,15}.

The LEA genes are mainly expressed in seeds and, as its name suggests, are accumulated during late embryo development stage, comprising up to 4% of cellular proteins¹⁶. They have also been found in seedlings, roots and other organs throughout the whole developmental stage. After suffering environmental stresses, plants accumulate high levels of LEA proteins in response^{17,18}. They have been proposed to have various functions, including protection of cellular structures from the effects of water loss and desiccation¹⁹, protection of proteins from stress-induced damage^{19,20,21}, sequestration of ions²⁰, and folding of denatured proteins²². LEA proteins can also act as chaperone proteins to resist cellular damage^{18,23}.

The LEA genes often exist as a large gene family in higher plants. For example, *Arabidopsis* and *Populus* contain 51 and 53 members respectively, which are divided into eight groups based on the amino acid sequence homology

¹Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan, China. ²State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 10093, China. Correspondence and requests for materials should be addressed to J.G. (email: gaojie@xtbg.org.cn)

and specific motifs²⁴. To date, comprehensive analysis of LEA genes have only been made in angiosperms like *Arabidopsis thaliana*²⁵, *Oryza sativa*²⁶, *Hordeum vulgare*²⁷, *Prunus mume*²⁸, *Populus*²⁴, *Solanum*^{29,30}, *Malus domestica*³¹ and legumes³². In contrast, considerably fewer LEA genes have been identified in gymnosperms, mostly due to limited genome information. Only a few LEAs and LEA-like proteins have been examined in *Picea glauca*³³, *Pseudotsuga menziesii*³⁴ and *Pinus pinaster*³⁵.

Pinus tabuliformis, a dominant conifer species is widely distributed over mountainous areas from northern to central China. It grows from 0–2700 m above sea level. Most of the region has heterogeneous habitats, including a wide range of temperatures, altitudes, elevated ozone and UV levels³⁶. Long generation time and large effective population sizes of *P. tabuliformis* suggest that this species has adapted to variable environmental conditions. Gymnosperms, especially conifers represent a large group of plants with a long evolutionary history. However, long generation time makes it a challenge to investigate transgenic functioning in conifers. The accumulation of LEAs during environmental stresses also makes the study of a single LEA gene difficult. Therefore, an alternative approach is to express the PtaLEA proteins *in vivo* in yeast or *E. coli* for functional expression screening, which has been successful in previous studies^{37–41}. By integrating phylogenetic analysis, protein motif structure, gene expression and heterologous expression analysis in *E. coli* of 23 full-length LEA genes, we examine the evolutionary and functional diversity of the LEA gene family in *P. tabuliformis*.

Materials and Methods

Identification of LEA genes from the *Pinus taeda* EST database. We identified LEA genes from *Pinus taeda* by searching 53 *Arabidopsis* LEA protein sequences in the *P. taeda* EST database (328,662 ESTs in the National Center for Biotechnology Information database: <http://www.ncbi.nlm.nih.gov/>) using TBLASTN. All the identified *P. taeda* candidates were analyzed using the protein family database (Pfam; <http://pfam.sanger.ac.uk/>) to confirm the presence of LEA conserved domains in their protein structure. We retrieved 43 whole LEA genes from *P. taeda*, allowing us to design 43 primers based on these gene sequences and use them to clone the LEA genes from *P. tabuliformis*.

Molecular cloning, phylogenetic analysis and bioinformatic analysis of *Pinus tabuliformis* LEA genes.

Total RNA isolation from seeds of *P. tabuliformis* was performed using an Aurum Total RNA Kit (Bio-Rad, Hercules, CA, USA). Total RNA was treated with RNase-free DNase I (Promega, Madison, WI, USA) and the first strand cDNA was synthesized using an RNA PCR Kit (AMV) version 3.0 (TaKaRa, Otsu, Japan). PCR conditions consisted of an initial denaturation of 3 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 40 s at 55 °C and 60 s at 72 °C, and a final extension of 5 min at 72 °C. PCR products were cloned into pEASY-T3 (TransGen, Beijing, China), and sequenced in both directions to verify the gene sequence. We confirmed that the sequences obtained had LEA conserved domains using Pfam.

We cloned 23 full-length LEA genes from *P. tabuliformis*. The multiple sequence alignment of the 23 PtaLEA genes was performed using MUSCLE⁴² (<http://www.ebi.ac.uk/Tools/msa/muscle/>) with default parameters and manually adjusted by BioEdit⁴³. Phylogenetic analysis was carried out by the maximum parsimony method with 1000 bootstrap replicates using MEGA v.5 software⁴⁴.

Conserved motifs for each LEA proteins were investigated using the Multiple Expectation maximization for Motif Elucidation (MEME) system version 4.10.24⁴⁵ (<http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>), any number of repetitions are expect to be distributed in sequences, with the minimum width of 6, and maximum width of 50. The signal peptide analysis was done by the TargetP algorithm⁴⁶ (TargetP: <http://www.cbs.dtu.dk/services/TargetP/>).

Transcription abundance of PtaLEA genes in *P. tabuliformis* tissues. To determine the expression pattern of each *P. tabuliformis* LEA gene, specific primers of the 23 LEA genes were designed (Table S-2) and used in RT-PCR, based on the multiple sequence alignment of LEA gene sequences. Total RNA was isolated from the top bud, needle, phloem in the stem, root of more than 10-year old *P. tabuliformis* and also two weeks germinants. Total RNA was treated as in the previously described method. Amplification was performed in a volume of 25 µL containing 10 pmol of each primer, 2.5 µL TaKaRa 10 × PCR buffer, 0.125 µL TaKaRa ExTaq (5 U/µL), 2 µL dNTP (2.5 mM each). PCR conditions consisted of an initial denaturation of 3 min at 94 °C, followed by 25 cycles of 30 s at 94 °C, 40 s at 60 °C and 1 min at 72 °C with a final extension of 3 min at 72 °C. In all PCR analysis, the *Actin* gene was used as an internal control. The PCR products were analyzed by 1.5% agarose gel electrophoresis. PCR products from each sample were validated by DNA sequencing. For each biological sample three repetitions were performed.

Construction of *E. coli* strains expressing LEA proteins. First, we added a NotI site upstream from the start codon ATG with the 5' end primer, and a SpeI site after the stop codon TAA with the 3' end primer. Then, the LEA genes we cloned from the first step were used to do the PCR amplification. The PCR products were ligated into T-vector (Promega Madison, USA). The vector was then digested using a NotI/SpeI double digestion, and the resulting DNA was gel-purified and subcloned into the TWIN1 vector (New England BioLabs, Beijing, China), linearized by a double digestion with the same restriction enzymes to construct TWIN-LEA plasmid. The PtaLEA coding region and the junction sequences were confirmed by DNA sequencing. The TWIN-LEA plasmid was transformed into applicable *E. coli* host strain BL21 (DE3). Transformed *E. coli* cells were grown in Luria-Bertani (LB) medium containing 100 mg/ml of carbenicillin and kept at 37 °C overnight. The overnight cultures were diluted 1,000-fold using fresh LB medium and incubation continued at 37 °C until mid-log phase (3–4 h, OD₆₀₀ = 0.6). Isopropyl b-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM, and incubation was continued at 37 °C for 3 h. Cells were harvested by centrifugation and resuspended in Buffer 1

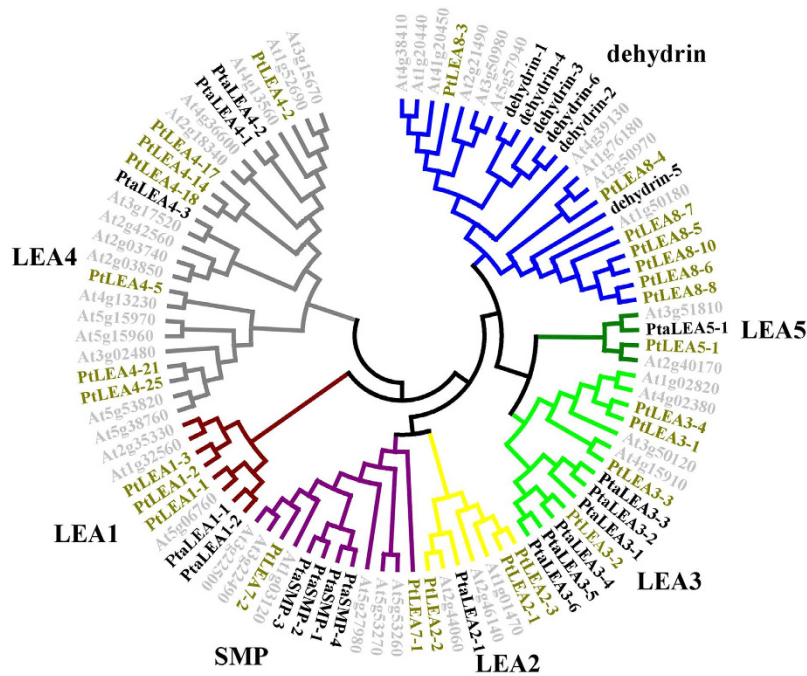


Figure 1. Phylogenetic relationships of *Pinus tabuliformis*, *Populus* and *Arabidopsis* LEA genes. LEA groups are distinguished by color. The light grey represented the LEA genes from *Arabidopsis*, the bold names represented the LEA genes from *P. tabuliformis*, and the green names are genes from *Populus*.

[20 mM Tris–HCl (pH 8.5), 500 mM NaCl, 1 mM EDTA] and lysed by sonication. The lysate was centrifuged at 12,000 rpm for 10 min to collect clarified cell extract for SDS–PAGE analysis.

Heat and salt treatment with transformed *E. coli* cells. Cell cultures were grown as described above. IPTG was added to mid-log phase cultures ($OD_{600} = 0.6$) to a final concentration of 1 mM, and incubation was continued at 37 °C for 3 h. After IPTG induction, the cultures were diluted and transferred to 50 °C. Samples (10 μ l) were taken at 0, 30, 60, 120, and 180 min, and serial dilutions were plated onto LB plus carbenicillin plates. We have detected the values of OD_{600} from each sample and make them to be the same value. Cell viability was estimated by counting the number of colony-forming units after incubation of the plate over-night at 37 °C. For salt treatment, after IPTG induction, ten microliters of each sample was spotted onto the LB plates (and we also controlled the OD_{600} for each of these samples), containing 171 mM NaCl and the LB plates with 300 mM NaCl, 400 mM NaCl, 500 mM NaCl, 600 mM NaCl, 700 mM NaCl, respectively. Plates were then incubated at 37 °C over night.

The colony number on each plate was recorded using the software of Cellprofiler⁴⁷. Viability ratio of the transformants under salt and heat conditions was calculated according to the following formula:

$$\text{Viability ratio} = \left(\frac{\text{mean of colony number on stressed Plate}}{\text{mean of colony number on control LB plate}} \right) \times 100\%.$$

For both treatments (heat and salt), the means of three experiments were determined from three independent transformants.

Results

Sequence and structure characteristics of the PtaLEA gene. Forty-three full-length genes encoding putative LEA proteins in *P. taeda* were identified from the *Pinus taeda* EST database. Due to the lack of *Pinus* whole-genome sequence information, we designed the primers based on the sequences of *P. taeda*. The *P. tabuliformis* may have the same or even more members of LEA proteins but we successfully cloned 23 full length LEA genes from the *P. tabuliformis*. All of these genes had LEA conserved domains which was confirmed by the Pfam domain analysis. A phylogenetic analysis based on amino acid sequences of the predicted proteins revealed that the 23 PtaLEA genes grouped into seven groups together with the *Arabidopsis* and *Populus* LEAs, confirming that they belonged to the LEA family (Fig. 1). We followed the LEA gene family nomenclature in Hundertmark and Hinch (2008) and the Pfam domain name. We found that group LEA3 and dehydrin, each have six genes. The LEA1 group has two members, LEA4 has three members, and SMP has four members, whereas groups LEA2 and LEA5 each contained only one gene. We did not obtain any members from LEA6 group.

These genes and their families possessed several group-specific characteristics (Table 1; Fig. 2). A majority of *P. tabuliformis* LEA proteins contained repeated motifs that were often specific to a LEA group. We analyze the group-specific characteristics with the comparison to the conserved motifs identified in *Arabidopsis*⁴⁸ and other species^{49,50}. The LEA4 group had several repeats of conserved motif 1: three times in PtaLEA4-1, up

Groups	Pfam groups	Proteins	Protein length(aa)	Molecular weights(kD)	Predicted subcellular localization
LEA1	LEA1	PtaLEA1-1	114	12.55	other
		PtaLEA1-2	117	12.51	other
LEA2	LEA2	PtaLEA2-1	190	21.40	other
LEA3	LEA3	PtaLEA3-1	113	12.22	Mitochondrion
		PtaLEA3-2	96	10.72	Mitochondrion
		PtaLEA3-3	106	11.80	Mitochondrion
		PtaLEA3-4	104	11.07	Mitochondrion
		PtaLEA3-5	103	11.31	Mitochondrion
		PtaLEA3-6	89	9.81	Mitochondrion
LEA4	LEA4	PtaLEA4-1	153	16.02	other
		PtaLEA4-2	98	10.44	other
		PtaLEA4-3	314	33.99	other
LEA5	LEA5	PtaLEA5-1	246	26.96	Chloroplast
SMP	SMP	PtaSMP-1	270	27.63	other
		PtaSMP-2	262	27.46	other
		PtaSMP-3	282	29.46	other
		PtaSMP-4	261	26.60	other
dehydrin	Dehydrin	dehydrin-1	177	18.88	other
		dehydrin-2	165	18.10	other
		dehydrin-3	238	25.85	other
		dehydrin-4	135	14.62	other
		dehydrin-5	93	10.32	other
		dehydrin-6	161	17.68	other

Table 1. The characteristics of the *P. tabuliformis* LEA proteins.

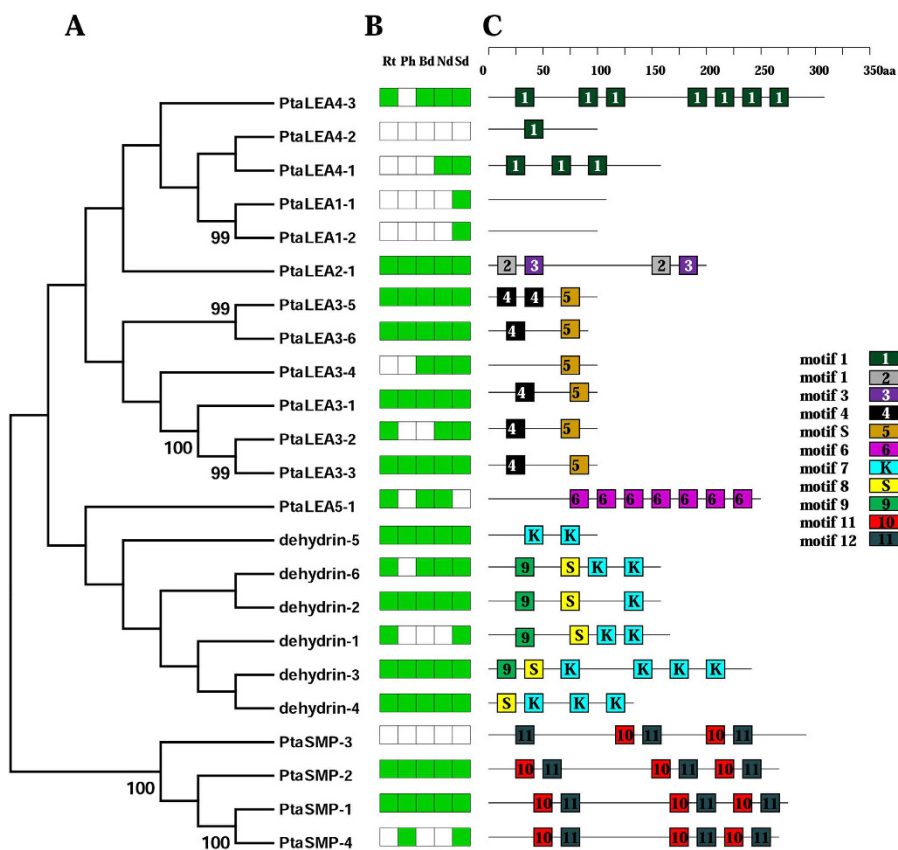


Figure 2. Phylogenetic relationships (A), expression patterns (B) and motif structure of *P. tabuliformis* LEAs. In B, the green box indicates positive detection of gene expression in the corresponding tissue under root (Rt), phloem (Ph), bud (Bd), needle (Nd), two weeks germinants (Ge). In C, boxes labeled with numbers are protein motifs; the motif sequences are provided in the supplemental Fig. 1.

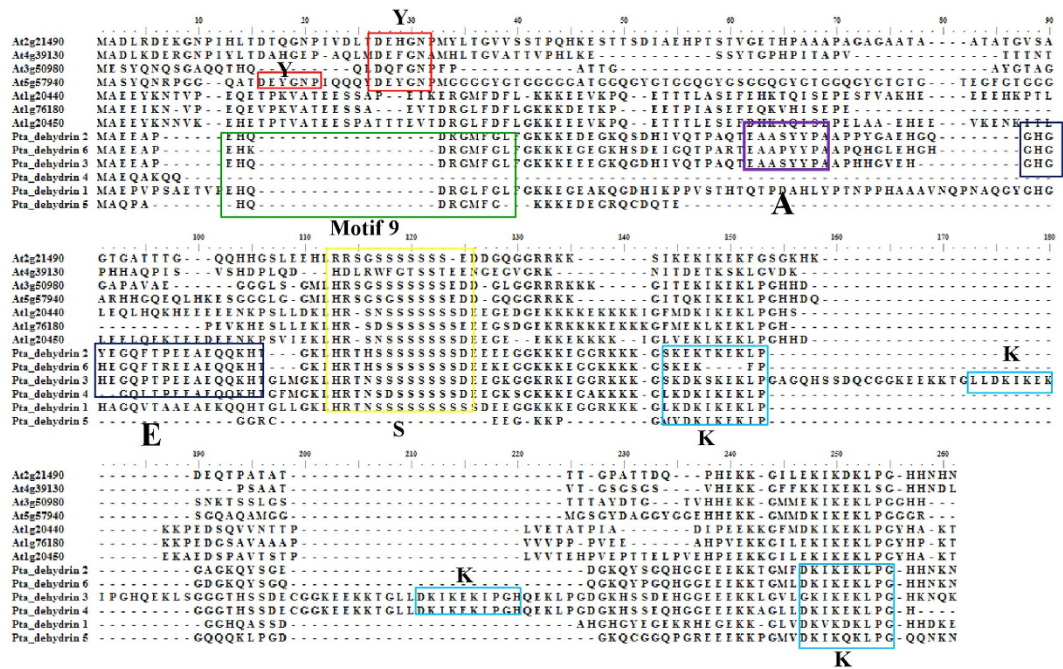


Figure 3. Alignment of the amino-acid sequences of *P. tabuliformis* dehydrins and previously known *dehydrin* genes from *Arabidopsis*. Alignments were performed using the CLUSTAL v program. Dashes indicate where a sequence has been expanded to allow optimal sequence alignment. The Y-, S-, K-, A- and E-segments are indicated by color boxes.

to seven times in *PtaLEA4-3* but only once in *PtaLEA4-2*. The *PtaLEA5-1* protein from group *LEA5* had the repeated group-specific motif 6 with seven repetitions. Group *dehydrin* is the second largest group in *LEA* family of which most gene members have been cloned in *P. tabuliformis* (6 members: *dehydrin-1/2/3/4/5/6*). MEME analysis showed that the motif structure was variable in this group. All of the six genes share the similar motifs S and K which are the conserved motif characteristic to *dehydrin* proteins, but we failed to find the Y motif in *P. tabuliformis* which could be found in *Arabidopsis* (see the alignment in Fig. 3). Apart from that, the repetitions of the motif K are variable in different *dehydrin* proteins: twice in *dehydrin-5/6/1*, three times in *dehydrin-4*, and up to four times in *dehydrin-3*, but only appeared once in *dehydrin-2*. The variation in motif structure indicates functional divergence.

Computational prediction of the subcellular distribution of the *LEA* proteins using targetP indicates further differences between the *LEA* groups. Whereas the members of the most groups are localized in the cytosol (“other” in Table 1 indicates that no signal peptide was detected), the *LEA3* proteins are exclusively targeted to mitochondria, and the *LEA5* is probably targeted to chloroplasts.

Transcription abundance of *LEA* genes in *P. tabuliformis*. In order to get a better understanding of the expression pattern of *P. tabuliformis* *LEA* genes, five different tissues/organs (root, phloem, needle, bud and two weeks germinants; Fig. 1B) were collected and analyzed using RT-PCR to detect the transcription abundance. Among the 23 *PtaLEA* genes, 11 (47.8%) genes were expressed in all tissues/organs examined under normal condition, indicating that *LEA* gene family has a wide distribution and expression in *P. tabuliformis*, and are undoubtedly involved in normal plant growth and development. All the genes (except *PtaLEA4-2*, *PtaLEA5-1* and *PtaSMP-3*) are expressed in the analyzed germinants, indicating that *LEA* genes have played an indispensable role in the seedling development process of *P. tabuliformis*. Two genes (*PtaLEA4-2* and *PtaSMP-3*) were not expressed in any tissue we studied, possibly because they are expressed at sub-detectable levels, or they are only induced in response to treatments and/or in tissues not including in our study (like seed). The other 10 *LEA* genes were expressed selectively in specific tissues and showed expression divergence within each group. For example, in the *dehydrin* group, the genes *PtaLEA 8-2/3/4/5* were expressed in all tissues, while *dehydrin-6* was expressed in all tissues except the phloem, and *dehydrin-1* was only expressed in the root and germinants. Within *LEA3* group, the *PtaLEA3-1/2/5/6* were expressed in all tissues, while *PtaLEA3-3* was expressed in root, needle and germinants, and the *PtaLEA3-4* was expressed in bud, needle and germinants. In group *LEA4*, the *PtaLEA4-1* was expressed in needle and germinants, while the *PtaLEA4-3* was expressed in all tissues except phloem, however the *PtaLEA4-2* did not have detectable expression in any organs. Generally, the different expression pattern of *LEA* family probably implies the divergence of function between genes within groups.

Enhancement of heat and salt-tolerance of recombinant *E. coli* cells with *PtaLEA* genes. In order to determine the function of expressed *LEA* fusion protein in stress conditions, we selected 2–3 genes from

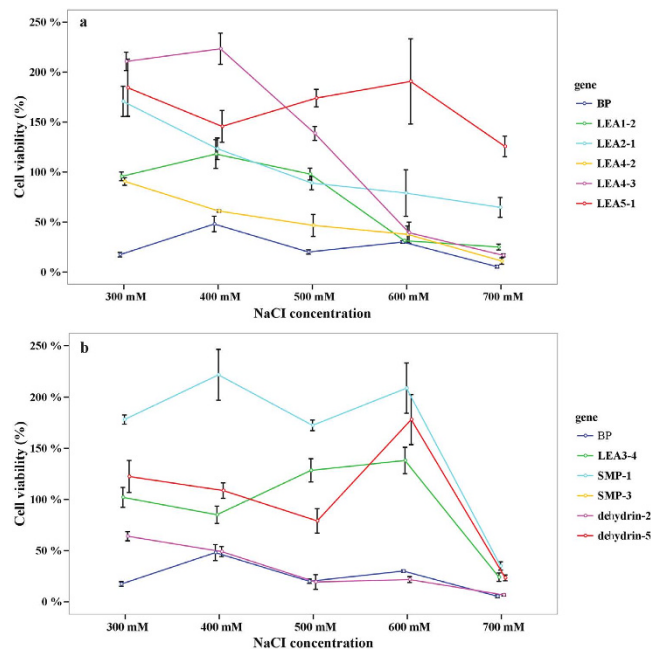


Figure 4. Cell Viability ratio of *E. coli* transformed with TWIN-LEA and TWIN1 constructs under salt treatment. The values are the mean \pm SE from three samples.

each group except group LEA 2 and LEA5 which we only have one gene in these group (13 genes in total) to construct the recombinant BL/LEA to do the heat and salt treatment.

For the salt treatment, we considered the gene as improving salt tolerance when the viability ratio was higher than the control strain BL/PTWN. The results showed that ten of the thirteen *PtaLEA* genes had mean viability ratios 2–7 fold higher than those of the control strain BL/PTWN under different concentration of salt stress (Fig. 4a,b). These ten genes belong to different groups. We selected *PtaLEA3-3/4*, *PtaLEA4-2/3* and *PtaSMP-1/3* from each group. All these genes, except *PtaLEA3-3*, have shown an increased salt tolerance. For the group *dehydrin*, we detected three genes *dehydrin-4/5/2*, two of which (*dehydrin-5/2*) have shown an increased salt tolerance with a mean viability ratio 4.3 and 1.3 fold higher than the control strain respectively. LEA2 and LEA5 have only one gene in each group, and LEA1 has two members. We chose all members in these three groups for experiments. The results showed that all of them increased salt tolerance except *PtaLEA1-1*. The genes *PtaSMP-1* (6.8 fold), *PtaLEA2-1* (3 fold), *PtaLEA4-3* (5.3 fold), *PtaLEA5-1* (6.8 fold), *dehydrin-5* (4.3 fold) had significantly higher survival ratios than other genes under all concentrations. Genes within group showed different patterns to salt stress, i.e. the viability ratios of *PtaLEA4-3* (5.3 fold), *dehydrin-5* (4.3 fold) and *PtaSMP-1* (6.8 fold) were higher than their group members *PtaLEA4-2* (2.0 fold), *dehydrin-2* (1.3 fold) and *PtaSMP-3* (1.6 fold) respectively.

Results in the heat treatment demonstrated that seven of the thirteen genes showed signs of increased heat tolerance, with viability ratios higher than the control strain BL/PTWN (Fig. 5a,b). Each group of *LEA1*, *LEA5* and *SMP* had one gene and group *LEA3* and *dehydrin* had two members. After 50 °C induction for 60 minutes, the viability ratio increased significantly which indicated that the genes need a certain induction time. The *PtaLEA3-4* (2.0 fold), *dehydrin-4/2* (1.5/2.4 fold) and *PtaSMP-3* (2.4 fold) showed the highest viability ratio among these genes. Within *LEA3* group, the viability ratio of *PtaLEA3-4* (2.0 fold) is much higher than *PtaLEA3-2* (1.3 fold).

In conclusion, the expressed *P. tabuliformis* LEA fusion polypeptide conferred salt and heat tolerance to the host cells. Some of the groups showed an increased resistance to heat and salt, but not significantly, like *dehydrin-2* that increased the viability by 1.3 fold under salt stress and *PtaLEA3-2* that increased the viability by 1.3 fold under heat stress. Some of the groups showed responses to both heat and salt stresses, but genes within group showed different patterns. In particular, within group *dehydrin*, the *dehydrin-5* (4.3 fold) showed higher viability under salt treatment, *dehydrin-4* (1.5 fold) showed higher viability under heat treatment, while *dehydrin-2* responded to both heat (2.4 fold) and salt (1.3 fold) treatment. All these results suggested that function divergence appeared not only between groups but also between genes within groups.

Discussion

The *LEA* genes play crucial roles during embryonic development and response to abiotic stresses throughout the plant kingdom as well as in many other organisms. Our molecular evolution analysis detected 43 gene members in the *Pinus taeda* genome, fewer than in *Populus*²⁴ which has 53 members. In this study, we have cloned 23 genes from the *P. tabuliformis* genome belonging to seven groups of the *LEA* gene family, the protein sequence identities among groups is very lower than the values within each group which exhibited a high degree of divergence among the protein groups.

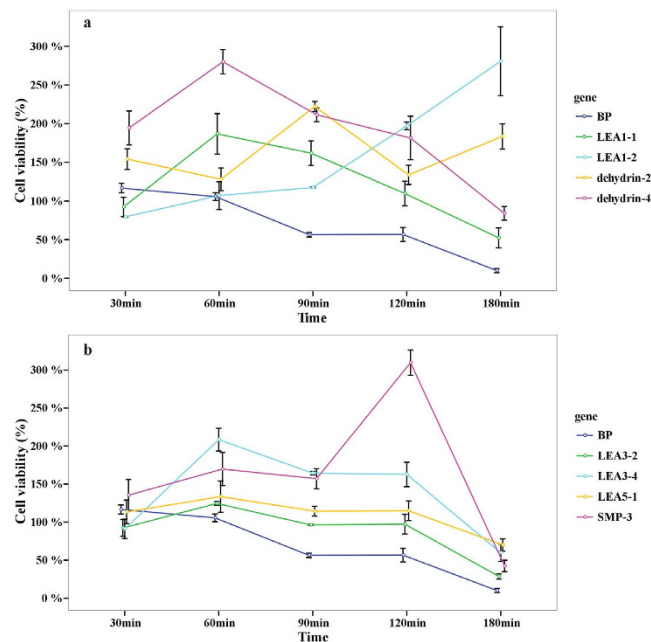


Figure 5. Cell viability ratio of *E. coli* transformed with TWIN-LEA and TWIN1 constructs under heat treatment at the temperature of 50°C. The values are the mean \pm SE from three samples.

Previously, the functional expression screening of LEA proteins *in vivo* in yeast or *E. coli* under abiotic stresses have been investigated successfully in various flowering plant species. Wang *et al.* (2008) investigated an *LEA* gene from *Tamarix* expressed in yeast which showed enhanced tolerance to high temperature, NaHCO_3 , ultraviolet radiation, salt, drought and freezing. Reportedly, one of *LEA3* proteins (*HVA1*) displayed increased tolerance to salt stress and water deficit in both yeast expression system and transgenic rice plants^{51,52}. Similarly, one *LEA4* gene from *Brassica napus* is able to enhance the cellular tolerance to temperature and salt stresses of *E. coli* cells⁵³. More recently a SMP protein from tea has reported function as a chaperone enhanced tolerance to *E. coli* against stresses⁴⁰. However, only a few LEA proteins increasing the tolerance to abiotic stresses had been studied and were primarily from the major group *LEA3*, *LEA4* and *dehydrin*. No gymnosperm *LEA* genes had been screened prior to this study. Furthermore, there was no study about the molecular mechanism of *P. tabuliformis* adapting to stressful environments. The *P. tabuliformis* has a very large distribution area including high temperate, arid and semiarid regions³⁶. The gradual salinization of land and the increased summer forest fires have been identified as the major threats to the pine forest in these regions^{54–56}. A study showed that *Pinus pinea* seed germination is sensitive to drought, salinity and heat influence⁵⁷. And another report also confirmed that the reduction of *dehydrin* proteins in seed have reduced seed longevity in *Arabidopsis thaliana*⁵⁸. This might be the same in the *P. tabuliformis* seed germination. The LEAs are considered playing a pivotal role in the late embryonic development during the seed germination¹⁶. In this study, 70% and 50% of the PtaLEAs analyzed increased the tolerance to salt and heat stresses, respectively. These genes come from almost all groups of the *LEA* gene family, which indicates the whole gene family has a function in stress response. Thus, for the first time we have confirmed that the *P. tabuliformis* *LEA* proteins from all *LEA* subgroups enhance tolerance to elevated salt and heat conditions. It indicates that the *LEA* genes have a wider repertoire of roles in stress responses and may play a common role in plant acclimation to stress conditions. This could be due to their highly conserved sequence motifs and high content of hydrophilic amino acids²¹, and we will discuss it later. Furthermore, the *LEA* subgroups represented diversified adaptation to heat and salt stress. *LEA1* increased the resistance to heat and salt, but not significantly, its viability ratio showed 2.3 and 3.1 fold higher than the control strain for heat and salt treatments respectively; while *SMP*, *LEA4* and *LEA5* induced a higher tolerance to salt stress (more than fivefold higher than the control strain), and *LEA3*, *dehydrin* induced a higher tolerance to heat treatment (more than twofold than the control strain). For large gene families with multiple subgroups, genes from different groups usually show functional diversification⁵⁹. Another functional analysis performed using the *E. coli* heterologous expression system, including three genes from group *LEA1*, *LEA3* and *dehydrin* respectively, also supported the hypothesis that genes from *LEA1* and *LEA3* increased the tolerance to salt and low temperature, while genes from *dehydrin* did not induce obvious growth improvement⁶⁰. While a recent study has investigated functions of six groups of *LEA* proteins from *Arabidopsis thaliana* which were expressed in *Saccharomyces cerevisiae* under desiccation stress. None of proteins from *LEA1*, *LEA5* and *AtM* showed protection but proteins from *LEA2*, *LEA4* and *dehydrin* has enhanced tolerance to desiccation stress⁴¹.

In addition to the expected functional divergence between gene groups, we observed substantial divergence between genes within a single group. The phenomenon was especially apparent in group *LEA4* and group *dehydrin*. *LEA4* is the largest group in *LEA* gene family, with the *Populus* and *Arabidopsis* genomes containing 26 and 16 members, respectively^{24,25}. It is reported that this group is very heterogeneous and the proteins differ greatly in

size and GRAVY²⁵. *LEA4* is also the largest group (16 genes) in *P. taeda*. In *P. tabuliformis* we cloned three members and investigated two of them under stress conditions. Within group *LEA4*, the gene *PtaLEA4-3* induced a higher salt tolerance (5.3 fold) than the gene *PtaLEA4-2* (2.0 fold). Clear divergence in expression patterns and motif structure was observed between the two genes. *PtaLEA4-3* had the group-specific motif 1 repeated up to seven times, and its expression was observed in all of the tissues except phloem. The *PtaLEA4-2* had only one repetitions of motif 1 that did not express in any tissue we studied. The enhanced salt-tolerance of *PtaLEA4-3* with more motif repetitions was consistent with a previous study which investigated the repeat region (a highly conserved 22-mer motif) in the PM2 (*LEA3*) protein in *E. coli*. They have found that the recombinant strains with more repeat motifs had much higher viability ratios than the recombinant strains with shortened repeat motifs⁶¹. This evidence proved the fact that the conserved repeat motifs in LEA protein have played important function in salt tolerance.

The *dehydrin* group, also referred to as *dehydrins*, is the second largest group in the *LEA* family. Within this group, the *dehydrin-5* induced a higher viability ratio to salt treatment, and *dehydrin-4* increased the heat tolerance, while *dehydrin-2* responded to both treatments. The *P. tabuliformis* proteins share the highly conserved motif elements of K-segment and S-segment, but they lack the Y-segment [(V/T) DEYGNP], which consistent with the report that this segment only been described in angiosperm⁶². The K-segments are found located near the C-terminal region and supposed to be involved in the formation of class A2 amphipathic α -helix⁶³. The helix is considered to establish hydrophobic interactions with other proteins in order to stabilize cell membranes⁶⁴. Thus, the role of the K-segment may be hydrophobic interaction with partially denatured proteins and protect the cell membranes, especially under the stressed condition^{65,66}. Another important S-segment (a tract of Ser-residues) is the motif S in our analysis and it has been demonstrated that Ser-residues in the segment can be phosphorylated and leads to calcium binding⁶⁷. The phosphorylation is also related to the binding of nuclear localization signal peptides to nuclear transport⁶⁵. These motifs repetitions are variable in *dehydrin* proteins, especially in gymnosperms compared to the angiosperms. The *dehydrin-5* lack the motif S, *dehydrin-4* has three repetition of motif K, while *dehydrin-2* has only one motif K. A recent study also reported novel conserved segments are associated with differential expression patterns in *Pinus pinaster*⁶⁸. In that study the discovered novel conserved A - segment (EAASYYP) and E - segment (GHGYEGQFTPEAEQKH) which also could be found in *dehydrin -2/3/4/6* in our study but could not be found in *dehydrin-5*. These diversified motif structures presumably related to the divergent tolerance abilities observed but the detailed underlying mechanisms remain to be explored.

References

- Serrano, R. & Rodriguez-Navarro, A. Ion homeostasis during salt stress in plants. *Curr. Opin. Cell Biol.* **13**, 399–404 (2001).
- Zhu, J. K. Plant salt tolerance. *Trends Plant Sci.* **6**, 66–71 (2001).
- Heremans, K. & Smeller, L. Protein structure and dynamics at high pressure. *BBA-Protein struct. M.* **1386**, 353–370 (1998).
- Hazel, J. R. Thermal adaptation in biological membranes: is homeoviscous adaptation the explanation? *Annu. Rev. Physiol.* **57**, 19–42 (1995).
- Berry, J. & Bjorkman, O. Photosynthetic response and adaptation to temperature in higher plants. *Annu. Rev. Physiol.* **31**, 491–543 (1980).
- Howarth, C. J. *Genetic improvements of tolerance to high temperature.* (Howarth Press, 2005).
- Steponkus, P. L. Responses to extreme temperatures. Cellular and sub-cellular bases. *Encyclopedia of Plant Physiology* **12**, 371–402 (1981).
- Wang, W. X., Vinocur, B. & Altman, A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* **218**, 1–14 (2003).
- Tolter, D., Hinch, D. K. & Macherel, D. A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. *BBA-Gen. subjects* **1798**, 1926–1933 (2010).
- Hinch, D. K. & Thalhammer, A. LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem. Soc. T.* **40**, 1000–1003 (2012).
- Cuming, A. C. *LEA proteins.* (Kluwer press, 1999).
- Galau, G. A., Hughes, D. W. & Dure, L. Abscisic acid induction of cloned cotton late embryogenesis-abundant (Lea) mRNAs. *Plant Mol. Biol.* **7**, 155–169 (1986).
- Goyal, K., Browne, J. A., Burnell, A. M. & Tunnacliffe, A. Dehydration-induced *tps* gene transcripts from an anhydrobiotic nematode contain novel spliced leaders and encode atypical GT-20 family proteins. *Biochimie* **87**, 565–574 (2005).
- Battista, J. R., Park, M. J. & Mclemore, A. E. Inactivation of two homologues of proteins presumed to be involved in the desiccation tolerance of plants sensitizes *Deinococcus radiodurans* R1 to desiccation. *Cryobiology* **43**, 133–139 (2001).
- Hand, S. C., Jones, D., Menze, M. W. & Witt, T. L. Life without water: expression of plant LEA genes by an anhydrobiotic arthropod. *J. Exp. Zool.* **305**, 1–5 (2006).
- Roberts, J. K., DeSimone, N. A., Lingle, W. L. & Dure, L. Cellular concentrations and uniformity of cell-type accumulation of two LEA proteins in cotton embryos. *Plant Cell* **5**, 769–781 (1993).
- Silveira, V. *et al.* Endogenous abscisic acid and protein contents during seed development of *Araucaia angustifolia*. *Biol. Plantarum* **52**, 101–104 (2008).
- Wise, M. J. LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles. *BMC Bioinformatics* **4**, 52 (2003).
- Serrano, R. & Montesinos, C. Molecular bases of desiccation tolerance in plant cells and potential applications in food dehydration. *Food Sci. Technol. Int.* **9**, 157–161 (2003).
- Grelet, J. B. A. *et al.* Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiol.* **137**, 157–167 (2005).
- Shao, H. B., Liang, Z. S. & Shao, M. A. LEA proteins in higher plants: structure, function, gene expression and regulation. *Colloid. Surface B.* **45**, 131–135 (2005).
- Bray, E. A. Molecular responses to water deficit. *Plant Physiol.* **103**, 1035–1040 (1993).
- Umezawa, T. *et al.* Engineering drought tolerance in plants: discovering and tailoring genes unlock the future. *Curr. Opin. Biotech.* **17**, 113–122 (2006).
- Ting, L., Jie, G. & Qing-Yin, Z. Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in *Populus trichocarpa*. *Tree Genet. Genomes* **9**, 253–264 (2013).
- Hundertmark, M. & Hinch, D. K. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**, 118–139 (2008).

26. Wang, X. S. *et al.* Genome-scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). *Plant Sci.* **172**, 414–420 (2007).
27. Choi, D. W., Zhu, B. & Close, T. J. The barley (*Hordeum vulgare* L.) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv Dicktoo. *Theor. Appl. Genet.* **98**, 1234–1247 (2007).
28. Du, D. L. *et al.* Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*. *Mol. Biol. Rep.* **40**, 1937–1946 (2013).
29. Cao, J. & Li, X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). *Planta* **241**, 757–772 (2015).
30. Charfeddine, S. *et al.* Genome-wide identification and expression profiling of the late embryogenesis abundant genes in potato with emphasis on dehydrins. *Mol. Biol. Rep.* **42**, 1163–1174 (2015).
31. Liang, D., Xia, H., Wu, S. & Ma, F. W. Genome-wide identification and expression profiling of dehydrin gene family in *Malus domestica*. *Mol. Biol. Rep.* **39**, 10759–10768 (2012).
32. Battaglia, M. & Covarrubias, A. A. Late embryogenesis abundant (LEA) proteins in legumes. *Front. Plant Sci.* **4** 190 doi: 10.3389/fpls.2013.00190 (2013).
33. Jin-Zhuo, D. & David, I. D. Cloning and characterization of six embryogenesis-associated cDNAs from somatic embryos of *Picea glauca* and their comparative expression during zygotic embryogenesis. *Plant Mol. Biol.* **39** 859–864 (1998).
34. Jarvis, S. B., Taylor, M. A., Macleod, M. R. & David, H. V. Cloning and characterization of the cDNA clones of three genes that are differentially expressed during dormancy-breakage in the seeds of Douglas fir (*Pseudotsuga menziesii*). *J. Plant Physiol.* **147**, 559–566 (1996).
35. Perdiguero, P., Collada, C. & Soto, A. Novel dehydrins lacking complete K-segments in Pinaceae, the exception rather than the rule. *Front. Plant Sci.* **5**, 682. doi: 10.3389/fpls.2014.00682 (2014).
36. Mao, J. F. & Wang, X. R. Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan plateau. *Am. Nat.* **177**, 424–439 (2011).
37. Wang, B. F. *et al.* Verification of the resistance of a LEA gene from *Tamarix* expression in *Saccharomyces* to abiotic stresses. *Journal of Forestry Research* **19**, 58–62 (2008).
38. Lei, Z., Akinori, O., Masamichi, T. & Ryozi, I. Expression of plant group 2 and group 3 lea genes in *Saccharomyces cerevisiae* revealed functional divergence among LEA proteins. *J. Biochem.* **127**, 611–616 (2000).
39. Honjoh, K. *et al.* Introduction of the hiC6 gene, which encodes a homologue of a late embryogenesis abundant (LEA) protein, enhances freezing tolerance of yeast. *J. Plant Physiol.* **155**, 509–512 (1999).
40. Paul, A., Singh, S., Sharma, S. & Kumar, S. A stress-responsive late embryogenesis abundant protein 7 (CsLEA7) of tea [*Camellia sinensis* (L.) O. Kuntze] encodes for a chaperone that imparts tolerance to *Escherichia coli* against stresses. *Mol. Biol. Rep.* **41**, 7191–7200 (2014).
41. Dang, N. X., Popova, A. V., Hundertmark, M. & Hinch, D. K. Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and *in vitro*. *Planta* **240**, 325–336 (2014).
42. Thompson, J. D. *et al.* The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
43. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 4 (1999).
44. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2738 (2011).
45. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
46. Emanuelsson, O., Nielsen, H., Brunak, S. & Von, H. G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
47. Martha, S. V. & Anne, E. C. Using cellprofiler for automatic identification and measurement of biological objects in images. *Current Protocols in Molecular Biology* **109**, 1–13 (2008).
48. Bies-Etheve, N. *et al.* Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* **67**, 107–124 (2008).
49. Du, D. L. *et al.* Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*. *Mol. Biol. Rep.* **40**, 1937–1946 (2013).
50. Liang, D., Xia, H., Wu, S. & Ma, F. W. Genome-wide identification and expression profiling of *dehydrin* gene family in *Malus domestica*. *Mol. Biol. Rep.* **39**, 10759–10768 (2012).
51. Xu, D. P., Duan, X. L. & Wang, B. Y. Expression of a late embryogenesis abundant protein gene HVA1, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.* **110**, 249–257 (1996).
52. Zhang, L., Ohta, A., Takagi, M. & Imai, R. Expression of plant group 2 and group 3 lea genes in *Saccharomyces cerevisiae* revealed functional divergence among LEA proteins. *J. Biochem.* **127**, 611–616 (2000).
53. Dalal, M., Tayal, D., Chinnusamy, V. & Bansal, K. C. Abiotic stress and ABA-inducible group 4 LEA from *Brassica napus* plays a key role in salt and drought tolerance. *J. Biotechnol.* **139**, 137–145 (2009).
54. Tang, J. & Lin, N. F. Some problems of ecological environmental geology in arid and semiarid areas of China. *Environ. Geol.* **26**, 64–67 (1995).
55. Tang, J. & Lin, N. F. Study on the environment evolution and the analysis of causes to land salinization and desertification in songnen plain. *Quaternary Sciences* **04**, 474–483 (2005).
56. Li, X. Y. *et al.* Assessment for salinized wasteland expansion and land use change using GIS and remote sensing in the west part of northeast China. *Environ. Monit. Assess* **131**, 421–437 (2007).
57. Sidari, M., Mallamaci, C. & Muscolo, A. Drought, salinity and heat differently affect seed germination of *Pinus pinea*. *J. Forest Res.* **13**, 326–330 (2008).
58. Hundertmark, M., Buitink, J., Leprince, O. & Hinch, D. K. The reduction of seed-specific dehydrins reduces seed longevity in *Arabidopsis thaliana*. *Seed Sci. Res.* **21**, 165–173 (2011).
59. Ting, L. *et al.* Extensive functional diversification of the *Populus* Glutathione S-Transferase supergene family. *Plant Cell* **21**, 3749–3766 (2009).
60. Ying, L., Dan, C. & Yi-Zhi, Z. Expression in *Escherichia coli* of three different soybean late embryogenesis abundant genes to investigate enhanced stress tolerance. *J. Integr. Plant Biol.* **47**, 613–621 (2005).
61. Liu, Y. & Zheng, Y. Z. PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. *Biochem. Biophys. Res. Co.* **331**, 325–332 (2005).
62. Perdiguero, P. *et al.* Novel dehydrins lacking complete K-segments in Pinaceae. The exception rather than the rule. *Front. Plant Sci.* **5**, 682 (2014).
63. Baker, J., Dennis-Steele, C. & Dure, L. Sequence and characterization of 6 LeA proteins and their genes from cotton. *Plant Mol. Biol.* **11**, 277–291 (1988).
64. Koag, M. C., Fenton, R. D., Wilkens, S. & Close, T. J. The binding of Maize DHN1 to lipid vesicles gain of structure and lipid specificity. *Plant Physiol.* **131**, 309–316 (2003).
65. Close, T. J. Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiology Plant* **97**, 795–803 (1996).

66. Koag, M. C. *et al.* The K-segment of maize DHN1 mediates binding to anionic phospholipid vesicles and concomitant structural changes. *Plant Physiol.* **150**, 1503–1514 (2009).
67. Alsheikh, M. K., Svensson, J. T. & Randall, S. K. Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. *Plant cell Environ.* **28**, 1114–1122 (2005).
68. Perdiguero, P. & Barbero, M. C. Novel conserved segments are associated with differential expression patterns for *Pinaceae* dehydrins. *Planta* **236**, 1863–1874 (2012).

Acknowledgements

We would like to thank Prof. Chuck Canon from Texas Tech University USA and Dr. Joeri S. Strijk for valuable suggestions and English revision of the manuscript, Dr. Zhao Wei helping use the Cellprofiler software, Zhao Li and Yang Xue help with the experiment. This study was supported by grants from the National Science Foundation of China (NSFC, 31300560).

Author Contributions

J.G. and T.L. conceived and designed research. J.G. conducted experiments, analyzed data and wrote the manuscript. All authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Gao, J. and Lan, T. Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabulaeformis* (Pinaceae) in *Escherichia coli*. *Sci. Rep.* **6**, 19467; doi: 10.1038/srep19467 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>