Tree Genetics & Genomes (2013) 9:253–264
DOI 10.1007/s11295-012-0551-2

ORIGINAL PAPER

Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in *Populus trichocarpa*

Ting Lan · Jie Gao · Qing-Yin Zeng

Received: 19 April 2012 / Revised: 18 June 2012 / Accepted: 1 July 2012 / Published online: 29 July 2012

Abstract Late embryogenesis abundant (LEA) proteins are a large and highly diverse group of polypeptides that are believed to function in desiccation and freezing tolerance in plants. This report presents a genome-wide analysis of LEA proteins and their encoding genes in *Populus trichocarpa*. Fifty-three *LEA* genes were identified from the *Populus* genome and divided into eight groups. The LEA4 and LEA5 groups were found in green algae and all land plants, whereas the other six groups existed only in land plants, indicating that the *LEA* family underwent rapid expansion during the early evolution of land plants. A majority of *Populus* LEA proteins contained repeated motifs that were often specific to a LEA group. Except for *PtLEA2-1* and *PtLEA2-3*, all *Populus* LEA proteins were highly hydrophilic. Examination of the chromosomal locations of *Populus LEA* genes revealed that 30 % were arranged in tandem repeats, indicating that tandem duplications significantly contributed to the expansion of this gene family in *Populus*. Expression patterns of all *Populus LEA* genes under normal growth conditions and abiotic stress (salinity and drought) were investigated by reverse transcription polymerase chain reaction. Twelve of 53 *Populus LEA* genes were selectively expressed in a specific tissue and/or in response to a specific treatment. *LEA* genes also showed extensive divergence in expression patterns, even among those from the same group or gene cluster. The expression profiles revealed that the *Populus LEA* gene family was not involved systematically in the same regulatory pathway.

Keywords Late embryogenesis abundant · Gene family · Abiotic stresses · Gene expression

Communicated by P. Ingvarsson

Electronic supplementary material The online version of this article (doi:10.1007/s11295-012-0551-2) contains supplementary material, which is available to authorized users.

T. Lan · J. Gao · Q.-Y. Zeng (✉)
State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
e-mail: qingyin.zeng@ibcas.ac.cn

Introduction

Late embryogenesis abundant (LEA) proteins were found to accumulate at high levels in the latter stages of seed maturation and to disappear following germination (Galau et al. 1986). LEA proteins have been detected in angiosperms, gymnosperms, lower plants and algae (Velten and Oliver 2001; Campos et al. 2006), suggesting that they may be ubiquitous in the plant kingdom. These proteins are also found in anhydrobiotic invertebrates, fungi, protists and even prokaryotes. LEA proteins are highly hydrophilic, contain a high percentage of glycine or other small amino acids (alanine, serine and threonine) and lack or contain small amounts of tryptophan and cysteine residues (Campos et al. 2006). Plant *LEA* genes are considered to play important roles in protecting cells from abiotic stress and in normal plant growth and development.

LEA gene expression has been observed in vegetative tissues exposed to dehydration or osmotic, cold, salt or abscisic acid stress treatment (Skriver and Mundy 1990; Close 1997; Steponkus et al. 1998; Grelet et al. 2005). Overexpression of *LEA* genes in transgenic plants and yeast has been used to elucidate the contributions of corresponding proteins to stress tolerance. The freezing tolerance of strawberry leaves was enhanced by the expression of a wheat *LEA* gene (*WCOR410*) (Houde et al. 2004). The overexpression of multiple *LEA* genes enhanced tolerance to freezing stress in *Arabidopsis* (Puhakainen et al. 2004), while overexpression of a barley *LEA* gene (*HVA1*) in rice and wheat conferred

increased drought and salt tolerance to transgenic plants (Xu et al. 1996; Sivamani et al. 2000). LEA proteins have been considered to act as protectors of macromolecules and/or some cellular structures during water deficit. Functional mechanisms of LEA proteins as protectors may interact with available water molecules, providing a hydration shell that protects target integrity and function (Bray 1997; Garay-Arroyo et al. 2000; Hoekstra et al. 2001). LEA proteins from citrus and barley could stabilise lactate dehydrogenase and malate dehydrogenase during freezing and/or drying (Hara et al. 2001; Sanchez-Ballesta et al. 2004), and recently, a mitochondrial LEA protein was found to stabilise model membranes in the dry state (Tolleter et al. 2010).

The heterologous expression of two cold-induced LEA proteins from spinach had no profound influence on stress tolerance in transgenic tobacco (Kaye et al. 1998), while the overexpression of *Arabidopsis* LEA proteins in *Escherichia coli* inhibited bacterial growth (Campos et al. 2006). In *Arabidopsis*, mutant seeds with a knocked-out *LEA* gene (*ATEM6*) displayed premature seed dehydration and maturation at the distal ends of siliquae, demonstrating that this protein was required for normal seed development. These data indicate that some LEA proteins may play important roles in various aspects of normal plant growth and development.

To date, comprehensive analyses of the *LEA* gene family have been performed only in *Arabidopsis thaliana*. The *Arabidopsis* genome contains 51 *LEA* genes, which were divided into nine groups on the basis of amino acid sequence similarity and specific motifs (Hundertmark and Hincha 2008; Bies-Etheve et al. 2008). The majority of *Arabidopsis* LEA proteins were predicted to be highly hydrophilic. Wide ranges of sequence diversity, intracellular localisations and expression patterns were observed in *Arabidopsis* (Hundertmark and Hincha 2008; Bies-Etheve et al. 2008). In this study, we conducted a genome-wide analysis of LEA proteins and their encoding genes in *Populus trichocarpa*. The genome sequence of *Populus* provided a model system for tree genomics and eudicot species that diverged from *Arabidopsis* approximately 120 million years ago. We identified 53 full-length *LEA* genes from *Populus*. Integrative analyses of sequence similarity, genomic organisation, biochemical characteristics and gene expression patterns provided a comprehensive view of this enigmatic group of proteins.

Methods

Identification of *LEA* genes in the *Populus* genome and other plant species

To obtain the complete *Populus LEA* gene family, the *P. trichocarpa* genome sequence (ver. 1.0; http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) was queried by 51 *Arabidopsis* LEA protein sequences using the TBLASTN search programme. All *Populus* LEA candidates were analysed first using the protein families database (Pfam) to confirm the presence of LEA-conserved domains in their protein structures. Proteins without LEA-conserved domains were further validated by joint phylogenetic analysis with all *Arabidopsis* LEAs. Proteins grouped with *Arabidopsis* LEAs were 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. For notational convenience, the SMP and dehydrin groups were renamed LEA7 and LEA8, respectively, in this study. A univocal name was assigned to each *Populus LEA* gene consisting of two italic letters denoting the source organism, the family name, subfamily numeral and a progressive number for each gene (e.g. *PtLEA1-1*).

To trace the evolutionary origin of the plant *LEA* gene family, we also searched the genomes of two green algae (*Chlamydomonas reinhardtii* and *Volvox carteri*), a moss (*Physcomitrella patens*), a lycophyte (*Selaginella moellendorffi*) and *Oryza sativa* using the strategy described above. No whole-genome sequence of a gymnosperm was available. The most comprehensive genomic resource currently available is the expressed sequence tag (EST) collection of *Pinus taeda* (328,662 ESTs in the National Center for Biotechnology Information database: http://www.ncbi.nlm.nih.gov/). Thus, in this study, we identified all *LEA* genes from the *P. taeda* EST database.

Phylogenetic analysis

Full-length amino acid sequences were aligned using ClustalX 1.83 software (Thompson et al. 1997) and adjusted manually with BioEdit (Hall 1999). Phylogenetic analysis was carried out by the maximum parsimony method with 1,000 bootstrap replicates using MEGA 4 software (Tamura et al. 2007).

Bioinformatic analysis of the *Populus LEA* gene family

To investigate the characteristics of LEA proteins, the grand average of hydropathicity index (GRAVY), theoretical isoelectric point (pI) and molecular weight were obtained using the ProtParam Tool (http://web.expasy.org/protparam/). Sequence identity within each LEA group was calculated with BioEdit (Hall 1999). Conserved motifs for each LEA protein were investigated using the Multiple Expectation Maximization for Motif Elucidation (MEME) system (version 3.0 http://meme.sdsc.edu/meme/cgi-bin/meme.cgi).

Plant growth conditions and RT-PCR analysis

Seedlings of *P. trichocarpa* were cultivated in potting soil for 2 months. Five replicate seedlings were then irrigated

and sprayed with 150 mM NaCl solution for 1 week (Ding et al. 2010). Drought stress was conducted by withholding water for 2 weeks (Caruso et al. 2002). After stress treatments, total RNA was isolated from leaf, shoot, bud, phloem and root tissues of each seedling 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 reverse transcribed into cDNA using an RNA PCR Kit (AMV) version 3.0 (TaKaRa, Otsu, Japan). Based on the multiple sequence alignment of *Populus LEA* gene sequences, specific primers were designed for RT-PCR analysis (Supplemental Table 1). Optimised PCR conditions were: initial denaturation for 3 min at 95 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 60 °C and 30 s at 72 °C and a final extension of 5 min at 72 °C. In all RT-PCR analyses, the *actin* gene was used as an internal control. PCR products from each sample were analysed using 1 % agarose gel and validated by DNA sequencing.

Results

Populus composed a large *LEA* gene family

The genome sequence of *P. trichocarpa* was searched using the TBLASTN programme with *A. thaliana* LEA protein sequences. In total, 53 full-length genes encoding putative LEA proteins were identified (Table 1). Domain analysis using Pfam showed that 40 of the 53 candidate genes had LEA-conserved domains. The remaining 13 genes were further validated by joint phylogenetic analysis with all *Arabidopsis* LEAs (Fig. 1). All of these genes were grouped with *Arabidopsis* LEAs, confirming that they belonged to the *LEA* family. Thus, these 13 *LEA* genes were included in the phylogenetic and experimental analyses.

Using the Pfam nomenclature, the *Populus LEA* gene family was divided into eight groups: LEA1–6, SMP and dehydrin. For notational convenience, the SMP and dehydrin groups were renamed LEA7 and LEA8, respectively. To facilitate reference to LEA proteins described in previous publications, we compared the LEA nomenclature with two recently published systems proposed by Hundertmark and Hincha (2008) and Bies-Etheve et al. (2008) (Table 2).

Based on the Pfam and phylogenetic analysis (Fig. 1), all *Populus LEA* genes could be divided into eight groups: LEA1–8. The LEA4 and LEA8 groups were largest, with 26 and 10 members, respectively. The LEA3 group had five members, LEA2 had four and LEA1 had three members. The LEA6 and LEA7 groups had two members, and LEA5 contained only one gene. The *Arabidopsis LEA* gene family had a species-specific group (AtM) that was absent in the *Populus* genome.

LEA gene family in plants

To understand the evolutionary path of the *LEA* gene family in the plant kingdom, we also identified all *LEA* genes from two green algae (*C. reinhardtii* and *V. carteri*), a moss (*P. patens*), a lycophyte (*S. moellendorffi*) and an angiosperm (*O. sativa*) (Supplemental Table 2). No whole-genome sequence is currently available for any gymnosperm species; the most comprehensive genomic resource is the EST collection for loblolly pine (*P. taeda*). In the database of 328,662 *P. taeda* ESTs, 43 *LEA* genes were identified using TBLASTN searches. Copy numbers of *LEA* family in above land plant species were at least 42 (Table 3). However, the green algae *V. carteri* and *C. reinhardtii* only contained five and three copies, respectively, indicating that the *LEA* family underwent rapid expansion during the early evolution of land plants. The LEA4 and LEA5 groups were found in all green algae and land plants, whereas six groups (LEA1–3, LEA6–8) existed only in land plants.

Genomic organisation and gene duplication of the *Populus LEA* gene family

Thirty-seven of 53 *LEA* genes were localised on 12 of 19 *Populus* chromosomes, whereas the remaining 16 genes were localised on 14 as yet unattributed scaffold fragments (Table 1). The distributions of *LEA* genes in the *Populus* genome were obviously heterogeneous (Fig. 2). Two clusters (clusters I and II) with high densities of *LEA* genes were observed on chromosomes 4 and 13. Two *LEA* genes (*PtLEA4-1/2*) were arranged in tandem repeats on chromosome 1. The remaining *LEA* genes showed a disperse distribution on chromosomes 2–5, 7, 9, 10, 14, 16 and 19. Chromosomes 6, 8, 11, 12, 15, 17 and 18 did not reside any *LEA* genes. Three genes (*PtLEA4-19/20/21*) were arranged in tandem repeats on a scaffold fragment (Table 1).

Previous analysis of the *Populus* genome identified paralogous segments created by a whole-genome duplication event in the Salicaceae (salicoid duplication) that occurred 60–65 million years ago (Tuskan et al. 2006). Twenty-two *LEA* genes were located in duplicate blocks (Fig. 2). However, nine of these genes (*PtLEA1-1*, *PtLEA2-2*, *PtLEA4-3/4/5/6*, *PtLEA5-1*, *PtLEA6-2* and *PtLEA7-1*) lacked corresponding duplicates; they may have been created after the salicoid duplication event, or the corresponding homologues may have been lost after the duplication event. Four duplicate pairs (*PtLEA2-1/3*, *PtLEA3-1/4*, *PtLEA8-1/3* and *PtLEA8-2/4*) were each located in a pair of paralogous blocks and can be considered to be direct results of the salicoid duplication event (Fig. 2). Four LEA8 genes (*PtLEA8-5/6/7/8*) in cluster II and *PtLEA8-9* were each located in a pair of paralogous blocks. Phylogenetic analysis showed these five genes (*PtLEA8-5/6/7/8/9*) were grouped

Table 1 Full-length *LEA* genes identified from the *Populus trichocarpa* genome

Class	Genes	Map position (bp)	Length of proteins (aa)	Number of intron
LEA1	PtLEA1-1	LG_XVI:2886739-2887662	175	1
	PtLEA1-2	scaffold_29:1973814-1974314	124	1
	PtLEA1-3	scaffold_118:246670-247170	162	1
LEA2	PtLEA2-1	LG_II:12611862-12612660	151	1
	PtLEA2-2	LG_VII:83225-84688	314	1
	PtLEA2-3	LG_XIV:2910617-2911946	149	1
	PtLEA2-4	scaffold_170:87597-88143	163	1
LEA3	PtLEA3-1	LG_II:15184538-15184813	91	0
	PtLEA3-2	LG_III:6737025-6737424	90	1
	PtLEA3-3	LG_X:2055108-2055591	107	0
	PtLEA3-4	LG_XIV:5692999-5693277	92	0
	PtLEA3-5	scaffold_77:1559377-1564578	82	1
LEA4	PtLEA4-1	LG_I:9324682-9325441	104	2
	PtLEA4-2	LG_I:9327372-9327881	140	1
	PtLEA4-3	LG_II:10311004-10317108	241	6
	PtLEA4-4	LG_II:11079838-11082351	178	2
	PtLEA4-5	LG_II:23327574-23328217	145	2
	PtLEA4-6	LG_IV:4760309-4761109	232	1
	PtLEA4-7	LG_IV:8885858-8886430	67	2
	PtLEA4-8	LG_IV:8902320-8902887	67	2
	PtLEA4-9	LG_IV:8947735-8948307	67	2
	PtLEA4-10	LG_IV:8955157-8955717	67	2
	PtLEA4-11	LG_IV:8963282-8963844	67	2
	PtLEA4-12	LG_IV:8972685-8973248	67	2
	PtLEA4-13	LG_IV:8984367-8984943	67	2
	PtLEA4-14	LG_VII:10864766-10867398	415	2
	PtLEA4-15	LG_X:242931-244007	309	1
	PtLEA4-16	LG_XIX:9904957-9906653	469	1
	PtLEA4-17	scaffold_41:379834-382030	628	1
	PtLEA4-18	scaffold_57:1466253-1468795	408	2
	PtLEA4-19	scaffold_123:221562-222004	67	2
	PtLEA4-20	scaffold_123:225314-225755	67	2
	PtLEA4-21	scaffold_123:228731-228291	67	2
	PtLEA4-22	scaffold_124:533912-534766	252	1
	PtLEA4-23	scaffold_129:618634-620114	437	2
	PtLEA4-24	scaffold_152:337204-342597	213	2
	PtLEA4-25	scaffold_4063:1331-1989	91	3
	PtLEA4-26	scaffold_4478:406-969	67	2
LEA5	PtLEA5-1	LG_X:16724397-16724836	93	1
LEA6	PtLEA6-1	LG_II:326728-327003	91	0
	PtLEA6-2	LG_V:5912121-5912369	82	0
LEA7	PtLEA7-1	LG_X:9565886-9566558	164	2
	PtLEA7-2	scaffold_70:746150-747195	263	2
LEA8	PtLEA8-1	LG_II:763648-765706	619	2
	PtLEA8-2	LG_IV:14781031-14781725	133	1
	PtLEA8-3	LG_V:17250724-17251513	225	1
	PtLEA8-4	LG_IX:2866181-2866987	183	1
	PtLEA8-5	LG_XIII:4591347-4591796	149	0
	PtLEA8-6	LG_XIII:4595053-4595762	171	1
	PtLEA8-7	LG_XIII:4601646-4602303	152	1
	PtLEA8-8	LG_XIII:4610249-4610906	152	1
	PtLEA8-9	LG_XIX:5525884-5526174	96	0
	PtLEA8-10	scaffold_1432:11124-11837	237	0

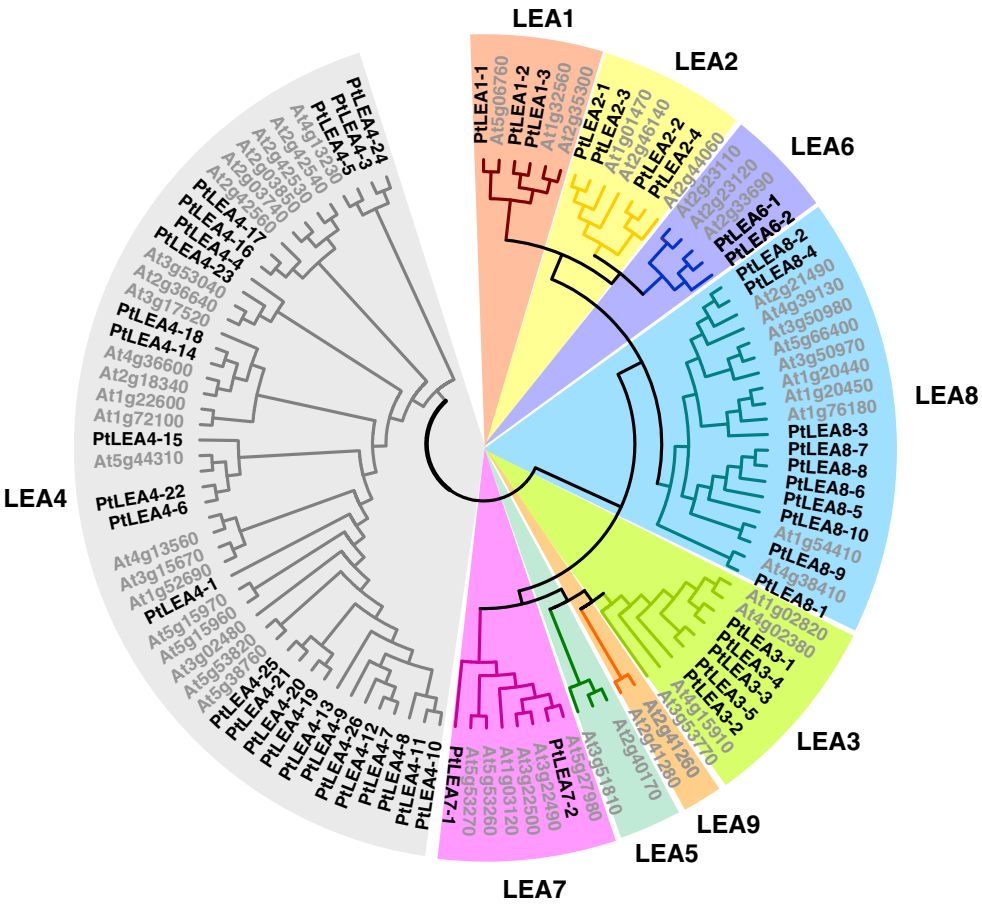

Fig. 1 Phylogenetic relationships of *Populus* and *Arabidopsis LEA* genes. *LEA* groups are distinguished by *colour*

together (Figs. 3a and 4a), indicating that they had descended from a common ancestor. Based on the gene tree and the positions of these five genes, we reconstructed their expansion history (Fig. 3b). An ancestral gene was probably duplicated during the whole-genome duplication event, followed by three rounds of tandem duplication in cluster II.

Characteristics of *Populus* LEA proteins

To better understand the characteristic features of the *Populus* LEA groups, we have summarised group-specific characteristics in Table 4 and Fig. 4. Except for *PtLEA2-1/3*, the GRAVY scores of all *Populus* LEA proteins were negative (Supplemental Table 3), indicating that they are hydrophilic. Other characteristics, such as gene structures, protein motifs and biochemical traits, were not conserved among LEA groups, but were conserved within each group.

LEA1 group

The *Populus* LEA1 group contains three genes (*PtLEA1-1/2/3*), each of which has two exons (Fig. 4c). The molecular weights of the PtLEA1-1/2/3 proteins are 17.80, 13.40 and 17.86, respectively. MEME analysis showed that the three proteins have two group-specific motifs (motifs 1 and 2) in the N-terminal region. The theoretical pI values of the PtLEA1-1/2/3 proteins are 8.04, 9.88 and 8.37, respectively, indicating that they are alkaline.

LEA2 group

The *Populus* LEA2 group contains four genes (*PtLEA2-1/2/3/4*), each with two exons. Sequence identities among the four proteins are 8–77 %. The PtLEA2-1/3/4 proteins contain 151, 149 and 163 amino acids, respectively. However, the PtLEA2-2 protein has 314 amino acids and is thus nearly twofold longer than the other three proteins. MEME analysis showed the all *Populus* LEA2 proteins have three motifs (motifs 3–5), which are repeated twice in the PtLEA2-2 protein. The theoretical pI values of the four LEA2 proteins range from 4.64 to 5.07, indicating that they are acidic. The GRAVY scores of the PtLEA2-2/4 proteins were negative, indicating that they are hydrophilic. In contrast, the GRAVY scores of PtLEA2-1/3 proteins were positive, indicating they are hydrophobic. Thus, the biochemical characteristics show divergence between the PtLEA2-2/4 and PtLEA2-1/3 proteins.

Table 2 The nomenclature of *LEA* groups in the Pfam database and according to Hundertmark and Hincha (2008) and Bies-Etheve et al. (2008)

In this study	Pfam	Hundertmark et al. (2008)	Bies-Etheve et al. (2008)
LEA1	LEA1	LEA1	LEA4
LEA2	LEA2	LEA2	LEA7
LEA3	LEA3	LEA3	LEA6
LEA4	LEA4	LEA4	LEA3
LEA5	LEA5	LEA5	LEA1
LEA6	LEA6	PvLEA18	LEA8
LEA7	SMP	SMP	LEA5
LEA8	Dehydrin	Dehydrin	LEA2
–	–	AtM	LEA9

LEA3 group

The *Populus* LEA3 group contains five members, and the sequence identities among proteins are 12–84 %. The theoretical pI values of PtLEA3-1/2/3/4 range from 8.06 to 9.89, indicating that they are alkaline. However, the PtLEA3-5 protein is acidic (pI=6.17). MEME analysis showed that PtLEA3-1/3/4 share similar motif elements (motifs 6–8), whereas PtLEA3-2 lacks motif 7 and PtLEA3-5 has only motif 6. Variation in motifs in this group indicates functional divergence.

The *PtLEA3-2/3/5* genes each have two exons/one intron structures (Fig. 4c), whereas *PtLEA3-1/4* lack introns. A duplicate gene pair (*PtLEA3-1/4*) was created by the salicoid whole-genome duplication event (Fig. 2). Before duplication, the most recent ancestral gene of this pair may have been created by retroposition. The retrogene is generally believed to have lacked an intron and possessed poly(A) stretches at the 3′ end and short direct repeats at both ends. The *PtLEA3-1* gene exhibits these features (Fig. 5), supporting the identification of *PtLEA3-1/4* as retrogenes.

LEA4 group

The *Populus* LEA4 group is the largest, being represented by 26 genes with protein lengths ranging from 67 to 628 amino acids and sequence identities of 4–100 %. The theoretical pI values of the LEA4 proteins are variable, ranging from 4.7 to 10.3. Although all *Populus* LEA4 proteins are hydrophilic (negative GRAVY scores), the GRAVY scores showed much variation, ranging from −1.4 to −0.49 (Supplemental Table 3). The variation in biochemical characteristics among the 26 LEA4 proteins indicates functional divergence.

The gene structures of 11 LEA4 genes (*PtLEA4-19/20/21/26* and seven genes in cluster I) are highly conserved (Fig. 4c). These genes have similar extron lengths and positions. In contrast, the other 15 genes in this group show a low degree of gene structure conservation. MEME analysis revealed three motifs (motifs 9–11) in this group. Except for PtLEA4-3/4/16/24, which have only motif 10, all proteins in this group share motif 9. The PtLEA4-5/6/17/22 proteins contain all three motif elements, whereas the PtLEA4-14/15/18/23 proteins have only motifs 9 and 10. Fourteen LEA4 proteins (PtLEA4-1/2/19/20/21/25/26 and seven proteins in cluster I) have motifs 9 and 11. Note that some motifs repeat several times in some genes (Fig. 4d). For example, PtLEA4-14 has four motif 9 repetitions and three motif 10 repetitions, and PtLEA4-17 has nine motif 10 repetitions.

LEA8 group

LEA8 is the second largest group in the *Populus LEA* gene family. This group contains 10 genes with highly variable protein sequences (sequence identity, 2–99 %) and theoretical pI values ranging from 5.01 to 10.34. The pI values of PtLEA8-2/3/4 are 5.01, 5.13 and 6.45, respectively, indicating that they are acidic. The other seven proteins are alkaline (pI>8.43). *PtLEA8-1* has three exons and two introns (Fig. 4c), whereas *PtLEA8-5/9/10* lack introns and the other

Table 3 Comparison of LEA group sizes in *Populus*, *Arabidopsis*, *Oryza*, *Pinus*, *Selaginella*, moss and green algae

Organisms	LEA groups									Total LEAs
	LEA1	LEA2	LEA3	LEA4	LEA5	LEA6	LEA7	LEA8	AtM	
Populus trichocarpa	3	4	5	26	1	2	2	10	–	53
Arabidopsis thaliana	3	3	4	18	2	3	6	10	2	51
Oryza sativa	4	5	6	9	2	2	9	8	–	45
Pinus taeda	3	1	6	16	3	1	3	10	–	43
Selaginella moellendorffii	2	10	5	16	6	–	8	–	–	47
Physcomitrella patens	1	4	1	27	4	–	3	2	–	42
Chlamydomonas reinhardtii	–	–	–	2	1	–	–	–	–	3
Volvox carteri	–	–	–	3	2	–	–	–	–	5

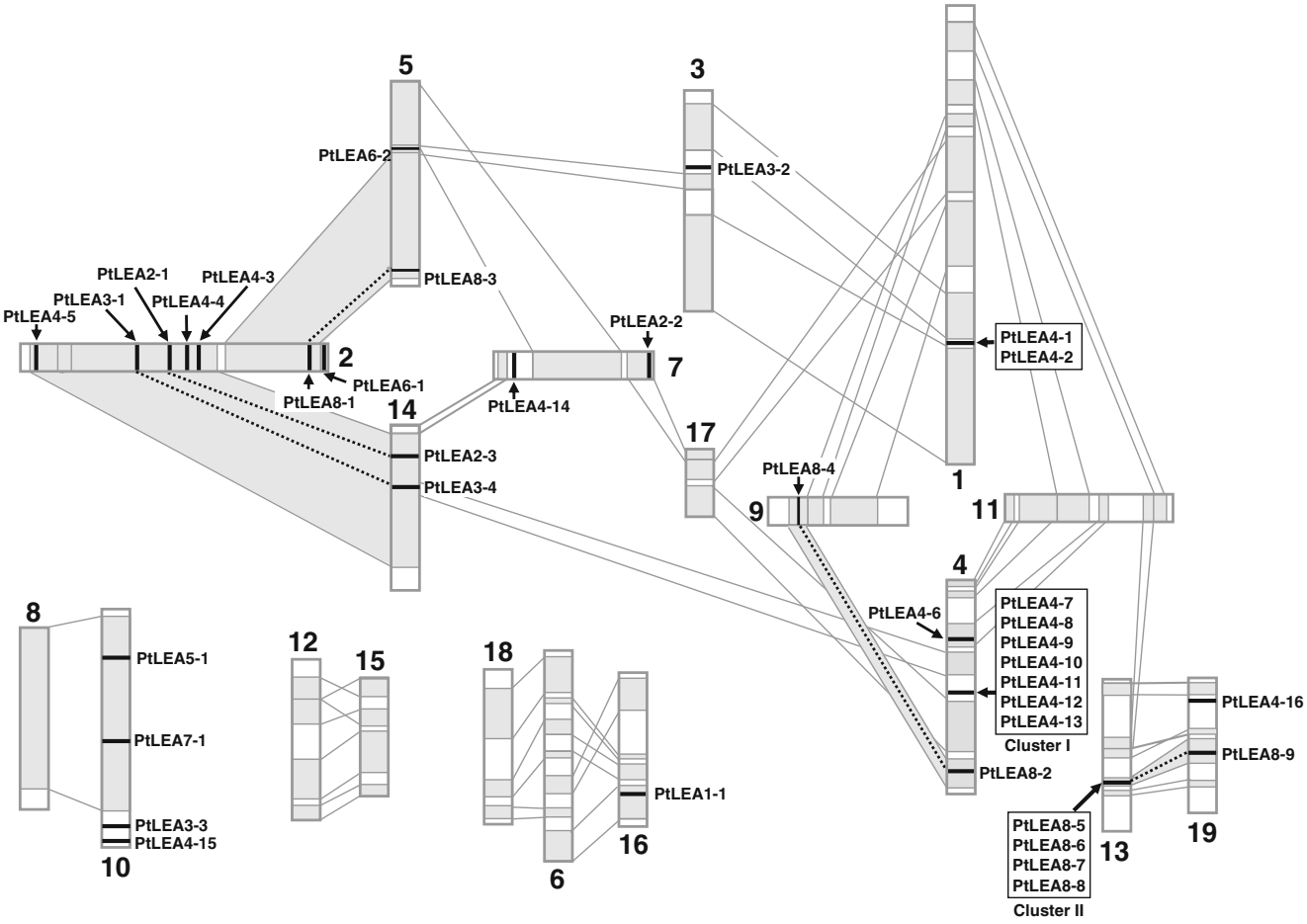

Fig. 2 Genomic localisation of *Populus LEA* genes. Schematic view of chromosome reorganisation by the most recent whole-genome duplication in *Populus* (adapted from Tuskan et al. 2006). Regions that are assumed to correspond to homologous genome blocks are *shaded in grey* and connected with lines. Paralogous *LEA* genes are indicated by *dashed lines* within the *grey-shaded trapezoids*

six genes contain only one intron. MEME analysis showed that all proteins in this group share motif 21. Except for PtLEA8-5/6/10, all proteins share motif 22. The PtLEA8-5/6/7/8/9/10 proteins have motif 23. Similar to the LEA4 group, some motifs in this group repeat several times in some proteins. In particular, motif 21 is repeated up to 12 times in PtLEA8-1 (Fig. 4d).

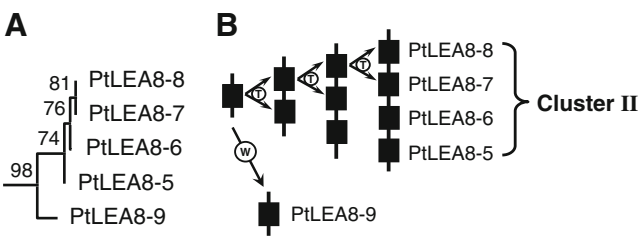

Fig. 3 Phylogenetic relationships (**a**) and hypothetical evolutionary histories (**b**) of the *LEA* genes in cluster II. The letters *T* and *W* in the schematic diagram of hypothetical origins of *LEA* genes indicate putative tandem duplication and whole-genome duplication, respectively

Intragenic evolution

The presence of repeated motifs is a distinguishing feature of *Populus* LEA proteins. Using sequence comparison and phylogenetic analysis of motifs in individual proteins, we attempted to obtain information on the possible history of these repetitions. The PtLEA8-1 protein contains the greatest number of repetitions (12 repetitions of motif 21), with highly conserved sequences (Fig. 6a). Based on the phylogenetic tree and the positions of the motif, we reconstructed the expansion history (Fig. 6b, c). At least 10 rounds of tandem duplications likely created the 12 repetitions of motif 21 in the PtLEA8-1 protein.

Expression of *LEA* genes under normal growth condition and abiotic stress

The expression patterns of *Populus LEA* genes under normal growth condition and abiotic stress (salinity and drought) were investigated using RT-PCR analysis. We analysed the

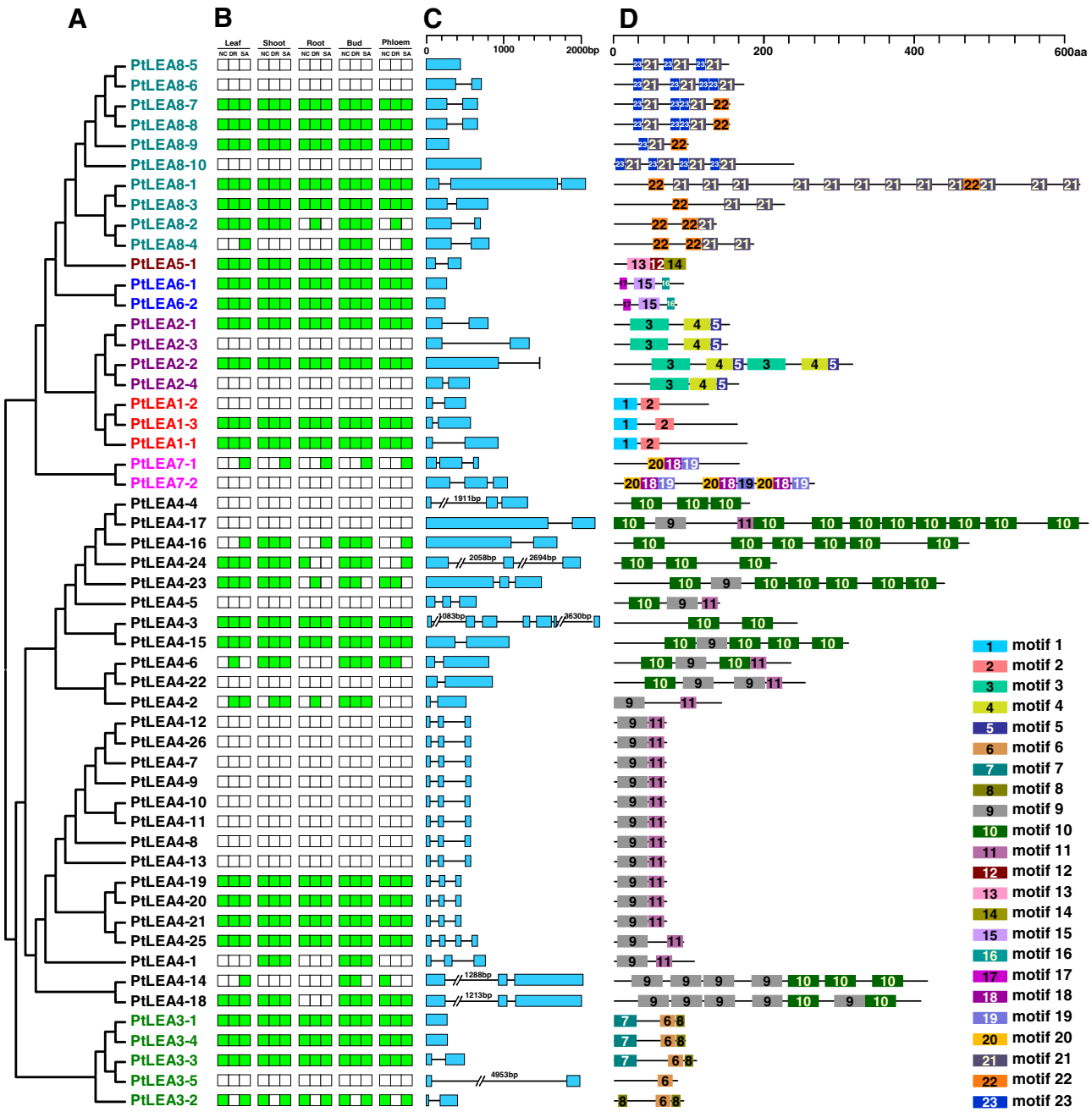

Fig. 4 Phylogenetic relationships (**a**), expression patterns (**b**), gene structure (**c**) and motif structure (**d**) of *Populus* LEAs. In **b**, the *green box* indicates positive detection of gene expression in the corresponding tissue under normal growth condition (*NC*) and following drought (*DR*) and salt (*SA*) stress treatments. In **d**, *boxes* labelled with numbers are protein motifs; the motif sequences are provided in the supplemental Figure 1

expressions of all 53 *LEA* genes in five tissues (leaf, shoot, root, bud and phloem; Fig. 4b). Among the 53 *Populus LEA* genes, 21 (39.6 %) genes were expressed in all tissues under all growth conditions, whereas 20 (37.7 %) genes were not expressed in any tissue or in response to any treatment applied in this study. Thus, these 20 *LEA* genes are expressed at subdetectable levels, or they are only induced in response to treatments and/or in tissues not examined in our study, or they are pseudogenes. The other 12 *LEA* genes were expressed selectively in a specific tissue and/or in response to a specific treatment. Nine of these genes (*PtLEA4-2/6/14/16/23/24*, *PtLEA7-1*, and *PtLEA8-2/4*) were not expressed in a specific tissue under normal growth condition but were expressed under salt or drought stress, suggesting that they play specific roles under stress conditions. The expressions of *LEA* genes are generally considered to restrict

Table 4 Group-specific characteristics of the *Populus* LEA proteins

Group	Number of genes	Protein sequence identities (%)		Molecular weight (kDa)		GRAVY		Theoretical pI	
		Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
LEA1	3	22	98	13.40	17.86	−0.861	−0.509	8.04	9.88
LEA2	4	8	77	16.24	34.94	−0.439	0.148	4.64	5.07
LEA3	5	12	84	8.65	11.32	−0.774	−0.146	6.17	9.89
LEA4	26	4	100	6.89	68.01	−1.415	−0.485	4.66	10.31
LEA5	1	–	–	–	–	–	–	–	–
LEA6	2	46		9.02	9.87	−1.421	−1.075	5.32	5.59
LEA7	2	22		17.88	27.65	−0.487	−0.418	5.26	9.12
LEA8	10	2	99	10.68	69.53	−1.995	−0.479	5.01	10.34

seed or stress tolerance. However, 32 of the 53 (60 %) *Populus LEA* genes examined in this study were expressed in vegetative tissues and in the absence of stress, indicating that *LEA* genes also play important roles in normal plant growth and development.

Among four tandem-arrayed LEA8 genes in cluster II, *PtLEA8-7/8* were expressed in all test samples but *PtLEA8-5/6* were not, suggesting functional divergence of this gene cluster. Seven LEA4 genes in cluster I were not expressed in any tissue under normal growth condition or abiotic stress. In this study, we identified four duplicate gene pairs created by a whole-genome duplication event. Two duplicate gene pairs (*PtLEA3-1/4* and *PtLEA8-1/3*) showed similar expression patterns: all duplicate genes were expressed in all tissues under normal growth condition and abiotic stress. Two duplicate genes in the *PtLEA8-2/4* pair showed different tissue-specific expression patterns. In the duplicate gene pair *PtLEA2-1/3*, *PtLEA2-1* was expressed in all test samples but *PtLEA2-3* was not, indicating functional divergence.

Discussion

In this study, we identified 53 *LEA* genes from the *Populus* genome; 30 % were arranged in tandem repeats and 15 % were created by the salicoid whole-genome duplication event. In *Arabidopsis*, 33 % of *LEA* genes were grouped in clusters and 24 % were the result of whole-genome duplication (Hundertmark and Hincha 2008; Bies-Etheve et al. 2008). These data indicate that tandem duplications contributed significantly to the expansion of this gene family. In addition, compared with the green algae *V. carteri* and *C. reinhardtii*, we found the rapid expansion of the *LEA* gene family in moss and other land plants. Extensive evidence has shown that *LEA* genes play an important role in cellular protection during abiotic stress tolerance, especially in drought stress (Babu et al. 2004; Bahieldin et al. 2005; Bahrndorff et al. 2009; Tolleter et al. 2010). The initial evolution of vegetative desiccation tolerance was a crucial step in the colonisation of the land by primitive plants from an origin in fresh water (Oliver et al. 2000). Synthetic phylogenetic analyses suggested that vegetative desiccation tolerance was primitively present in bryophytes, the lowest living clades of land plants (Oliver et al. 2000). This rapid expansion of the *LEA* gene family in land plants might have adaptive significance for the establishment of desiccation tolerance.

Genes that respond to stress generally contain fewer introns. Notably, 51 of the 53 (96 %) *Populus LEA* genes had fewer than two introns. Low intron numbers have also been observed in other gene families that respond to stress, e.g. tau glutathione transferase (one intron), the trehalose-6-phosphate synthase gene family (two introns) and the class III peroxidase gene family (three introns). Various hypotheses have been proposed to explain why highly expressed genes tend to be compact in a wide range of species. The transcriptional efficiency model is based on the observation that transcription is a slow and expensive process (Castillo-Davis et al. 2002). Introns have deleterious effects on gene expression, namely delayed transcript production caused by splicing and the added length of the nascent transcript, and an additional energetic cost due to increased transcript length (Jeffares et al. 2008).

AAGTTTCTCGAAGAAGAAACAAACAAGTTTCTACGGCAGTTGAAATATA
TAAAA**GATCCAT**CATCTCTTGTTTTCGTAAAACTTCTTTCACAAAGTTT
GAATCAAATCACACACTGTATTTGTAGA**ATGGCTCGTTCTTTCTCAAAC**
GCCAAGGTCATCTCTGGCCTGATCAGCGAGGCAATCAACGGCAGAGGAT
TCTCAGCTGTTGCATCCCAAGGAGCTGCTGTGTCCAAGGCAAGAAGCGG
TGCTGCTGTAATGAAGAAAACAGGGGAGGAGGTTACCAAGACCACCGAG
AAGATTTCCTGGGTTCCAGATCCTCGTACTGGATTCTACAGACCAGAGA
ATGTTGCTCAGGAAATCGATGCGGCTGAATTACGTGCTACTCTCTTGAA
GAAGCATTGAAGAAATTACTAATTCATGAGATTAATAAAATCTGATTAC
TACTACATCATGTTCTATGAT**AAAAAAAAAAAAAAAA**GG**GATCGAT**GGA
TATATGTGGGTTGTGGGGCTCGTGGGCTATCGGCTTGTGCTGTGACATG

Fig. 5 DNA sequence of the *PtLEA3-1* retrogene. The cDNA sequence is *underlined* and in *boldface* type. A poly(A) tail is *shaded in grey*. Two possible direct repeats are marked in *boxes*

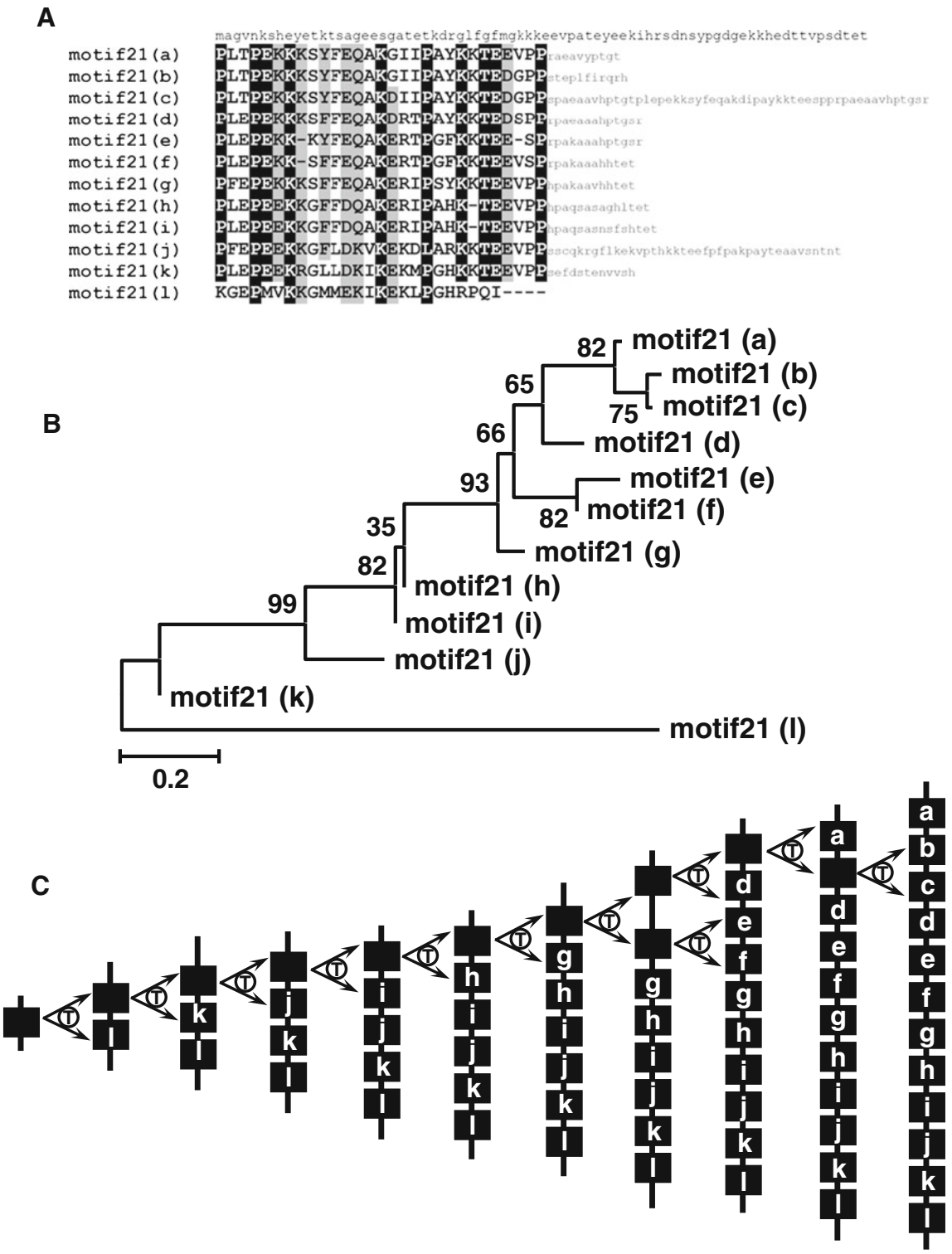

Fig. 6 Sequence comparison (**a**), phylogenetic relationships (**b**) and hypothetical evolutionary histories (**c**) of conserved motif 21 in the PtLEA8-1 protein. The letter *T* in (**c**) indicates putative tandem duplication

LEA4 is the largest group in the *Populus* and *Arabidopsis* genomes; it is represented by 26 and 16 genes, respectively, indicating rapid expansion of the *LEA4* gene in *Populus* (Bies-Etheve et al. 2008). Twelve of 26 (46 %) *Populus* LEA4 genes were grouped in clusters, whereas only two *LEA4* genes were arranged in tandem repeats in the *Arabidopsis*

genome. These findings indicate that tandem duplications contributed significantly to the expansion of LEA4 genes in *Populus*. Among the 26 *Populus LEA4* genes, 12 (46 %) genes were not expressed in any tissue or in response to any treatment applied in this study. However, all 16 LEA4 genes in the *Arabidopsis* genome were expressed: eight genes in seeds and eight in buds or in response to stress, including salinity or freezing (Hundertmark and Hincha 2008; Uemura et al. 1996). Thus, based on the expression data for the *Arabidopsis LEA4* group, we predict that the 12 *Populus* LEA4 genes (not expressed in this study) might be expressed in seeds or in response to freezing. In addition, 6 of 26 (23 %) *Populus LEA4* genes were expressed in all test tissues, and eight (31 %) were selectively expressed in a specific tissue and/or in response to a specific treatment. The divergent patterns of the *Populus LEA4* genes indicate functional divergence.

The *Arabidopsis LEA7* group has six members (Bies-Etheve et al. 2008). Four *Arabidopsis LEA7* genes were expressed in seed, one in bud and one in response to salt stress (Hundertmark and Hincha 2008). Analysis of transgenic *Arabidopsis* plants showed that one *LEA7* gene (*At1g03120*) played an important role in the ion cell balance during late embryogenesis and germination (Borrell et al. 2002). Expression patterns and transgenic *Arabidopsis* data suggest that this group may have divergent functions. The *Populus* LEA7 group contains only two genes (*PtLEA7-1/2*). *PtLEA7-1* was not expressed in the five tissues sampled under normal growth conditions or drought stress, but it was expressed under salt stress, indicating that this gene might play an important role in salt tolerance. The *PtLEA7-2* gene was not expressed in any test sample. Based on expression information for the *Arabidopsis LEA7* group, the *PtLEA7-2* gene may be expressed in seeds and play an important biological function during seed maturation.

Acknowledgments This study was supported by grants from the National Basic Research Program of China (2009CB119104), the Natural Science Foundation of China (NSFC 30800873) and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-J-1).

References

Babu RC, Zhang JX, Blum A, Ho THD, Wu R, Nguyen HT (2004) *HVA1*, a LEA gene from barley confers dehydration tolerance in transgenic rice (*Oryza sativa* L.) via cell membrane protection. Plant Sci 166:855–862

Bahieldin A, Mahfouz HT, Eissa HF, Saleh OM, Ramadan AM, Ahmed IA, Dyer WE, El-Itriby HA, Madkour MA (2005) Field evaluation of transgenic wheat plants stably expressing the *HVA1* gene for drought tolerance. Physiol Plant 123:421–427

Bahrndorff S, Tunnacliffe A, Wise MJ, McGee B, Holmstrup M, Loeschcke V (2009) Bioinformatics and protein expression analyses implicate LEA proteins in the drought response of *Collembola*. J Insect Physiol 55:210–217

Bies-Etheve 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 Mol Biol 67:107–124

Borrell A, Cutanda MC, Lumbreras V, Pujal J, Goday A, Culianez-Macia FA, Pages M (2002) *Arabidopsis thaliana* Atrab28: a nuclear targeted protein related to germination and toxic cation tolerance. Plant Mol Biol 50:249–259

Bray EA (1997) Plant responses to water deficit. Trends Plant Sci 2:48–54

Campos F, Zamudio F, Covarrubias AA (2006) Two different late embryogenesis abundant proteins from *Arabidopsis thaliana* contain specific domains that inhibit *Escherichia coli* growth. Biochem Biophys Res Commun 342:406–413

Caruso A, Morabito D, Delmotte F, Kahlem G, Carpin S (2002) Dehydrin induction during drought and osmotic stress in *Populus*. Plant Physiol Biochem 40:1033–1042

Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA (2002) Selection for short introns in highly expressed genes. Nat Genet 31:415–418

Close TJ (1997) Dehydrins: a commonality in the response of plants to dehydration and low temperature. Physiol Plant 100:291–296

Ding M, Hou P, Shen X, Wang M, Deng S, Sun J, Xiao F, Wang R, Zhou X, Lu C, Zhang D, Zheng X, Hu Z, Chen S (2010) Salt-induced expression of genes related to Na(+)/K(+) and ROS homeostasis in leaves of salt-resistant and salt-sensitive poplar species. Plant Mol Biol 73:251–269

Galau GA, Hughes DW, Dure L (1986) Abscisic-acid induction of cloned cotton late embryogenesis-abundant (LEA) messenger-RNAs. Plant Mol Biol 7:155–170

Garay-Arroyo A, Colmenero-Flores JM, Garciarrubio A, Covarrubias AA (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. J Biol Chem 275:5668–5674

Grelet J, Benamar A, Teyssier E, Avelange-Macherel MH, Grunwald D, Macherel D (2005) Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. Plant Physiol 137:157–167

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hara M, Terashima S, Kuboi T (2001) Characterization and cryoprotective activity of cold-responsive dehydrin from *Citrus unshiu*. J Plant Physiol 158:1333–1339

Hoekstra FA, Golovina EA, Tetteroo FA, Wolkers WF (2001) Induction of desiccation tolerance in plant somatic embryos: how exclusive is the protective role of sugars? Cryobiology 43:140–150

Houde M, Dallaire S, N'Dong D, Sarhan F (2004) Overexpression of the acidic dehydrin *WCOR410* improves freezing tolerance in transgenic strawberry leaves. Plant Biotechnol J 2:381–387

Hundertmark M, Hincha DK (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics 9:e118

Jeffares DC, Penkett CJ, Bahler J (2008) Rapidly regulated genes are intron poor. Trends Genet 24:375–378

Kaye C, Neven L, Hofig A, Li QB, Haskell D, Guy C (1998) Characterization of a gene for spinach CAP160 and expression of two spinach cold-acclimation proteins in tobacco. Plant Physiol 116:1367–1377

Oliver MJ, Tuba Z, Mishler BD (2000) The evolution of vegetative desiccation tolerance in land plants. Plant Ecol 151:85–100

Puhakainen T, Hess MW, Makela P, Svensson J, Heino P, Palva ET (2004) Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in *Arabidopsis*. Plant Mol Biol 54: 743–753

Sanchez-Ballesta MT, Rodrigo MJ, Lafuente MT, Granell A, Zacarias L (2004) Dehydrin from *Citrus*, which confers in vitro dehydration tolerance and freezing protection activity, is constitutive and highly expressed in the flavedo of fruit but responsive to cold and water stress in leaves. J Agric Food Chem 52:1950–1957

Sivamani E, Bahieldin A, Wraith JM, Al-Niemi T, Dyer WE, Ho TD, Qu R (2000) Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA*1 gene. Plant Sci 155:1–9

Skriver K, Mundy J (1990) Gene expression in response to abscisic acid and osmotic stress. Plant Cell 2:503–512

Steponkus PL, Uemura M, Joseph RA, Gilmour SJ, Thomashow MF (1998) Mode of action of the *COR15a* gene on the freezing tolerance of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 95:14570–14575

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Tolleter D, Hincha DK, Macherel D (2010) A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. Biochim Biophys Acta 1798:1926–1933

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Uemura M, Cilmour SJ, Thomashow MF, Steponkus PL (1996) Effects of COR6.6 and COR15am polypeptides encoded by *COR* (cold-regulated) genes of *Arabidopsis thaliana* on the freeze-induced fusion and leakage of liposomes. Plant Physiol 111:313–327

Velten J, Oliver MJ (2001) Tr288, a rehydrin with a dehydrin twist. Plant Mol Biol 45(6):713–722

Xu D, Duan X, Wang B, Hong B, Ho T, Wu R (1996) 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