Plant Mol Biol Rep (2012) 30:848–859
DOI 10.1007/s11105-011-0395-1

ORIGINAL PAPER

Genome-wide Identification and Characterization of a Dehydrin Gene Family in Poplar (*Populus trichocarpa*)

Chang-Cai Liu · Chun-Ming Li · Bao-Guang Liu · Su-Jie Ge · Xiu-Mei Dong · Wei Li · Hang-Yong Zhu · Bai-Chen Wang · Chuan-Ping Yang

Published online: 3 January 2012

Abstract Dehydrins (DHNs) define a complex group of stress inducible proteins characterized by the presence of one or more lysine-rich motifs. *DHNs* are present in multiple copies in the genome of plant species. Although genome-wide analysis of DHNs composition and chromosomal distribution has been conducted in herbaceous species, it remains unexplored in woody plants. Here, we report on the identification of ten genes encoding eleven putative DHN polypeptides in *Populus*. We document that *DHN* genes occur as duplicated blocks distributed over seven of the 19 poplar chromosomes likely as a result of segmental and tandem duplication events. Based on conserved motifs, poplar DHNs were assigned to four subgroups with the K_n subgroup being the most frequent. One putative DHN polypeptide (PtrDHN-10) with a SKS arrangement could originate from a recombination between SK_n and K_nS genes. In silico analysis of microarray data showed that in unstressed poplar, *DHN* genes are expressed in all vegetative tissues except for mature leaves. This exhaustive survey of *DHN* genes in poplar provides important information that will assist future studies on their functional role in poplar.

Electronic supplementary material The online version of this article (doi:10.1007/s11105-011-0395-1) contains supplementary material, which is available to authorized users.

C.-C. Liu · S.-J. Ge · X.-M. Dong · H.-Y. Zhu · B.-C. Wang ·
C.-P. Yang (✉)
State Key Laboratory of Tree Genetics and Breeding,
Northeast Forestry University,
26 Hexing Road,
Harbin 150040, People's Republic of China
e-mail: yangcp5516@gmail.com

C.-C. Liu
Laboratory for Chemical Defense and Microscale Analysis,
P.O. Box 3, Zhijiang 443200, People's Republic of China

C.-M. Li
Forestry Research Institution of Heilongjiang Province,
Harbin 150081, People's Republic of China

B.-G. Liu
Forestry College, Beihua University,
Jilin 132013, People's Republic of China

W. Li
School of Forestry, Northeast Forestry University,
26 Hexing Road,
Harbin 150040, People's Republic of China

H.-Y. Zhu
Bureau of Garden and Park,
26 Shanhu Road,
Qitaihe 154600, People's Republic of China

Keywords *Populus* · LEA · Dehydrins · Cold stress · Woody plants · Cellular dehydration

Abbreviations

LEA	Late embryogenesis abundant
DHN	Dehydrin
LGs	Linkage groups

Introduction

Dehydrins (DHNs) are Group II (D-11 family), late embryogenesis abundant (LEA) proteins that accumulate during seed desiccation and in response to water deficit induced by drought, low temperature or salinity in vegetative tissues or reproductive tissues (Close 1996; Allagulova et al. 2003; Kosova et al. 2007; Tunnacliffe and Wise 2007). A vital role in bud dormancy and cold acclimation of trees has been attributed to their certain DHN proteins (Rinne et al. 2010; HongXia et al. 2009; Rohde et al. 2007; Rorat 2006). DHNs are widely distributed in various organisms of plant kingdom including all seed plants, nonvascular plants and seedless vascular plants, where they accumulated in different cell compartments but mostly in the cytoplasm and nucleus (Battaglia et al. 2008; Tunnacliffe and Wise 2007; Allagulova et al. 2003).

The distinctive sequence feature of all DHN proteins is a conserved, Lys-rich 15-residue motif, EKKGIMDKIKEKLPG, named the K-segment often found in one to 11 copies within a single protein. Other optionally additional motifs in DHNs are the Y-segment ([V/T]D[E/Q]YGNP) usually found in one to 35 tandem copies in the N-terminus; the S-segment containing a track of Ser residues; the less conserved Φ-segment rich in polar amino acids and lay interspersed between K-segment (Close 1996; Allagulova et al. 2003). The presence and arrangement of these different conserved motifs in a single protein allow the classification of DHN proteins into five subgroups: Y_nSK_2, K_n, SK_n, K_nS, and Y_2K_n (Rorat 2006; Allagulova et al. 2003). In addition, some DHNs could only be assigned to certain intermediate forms instead of the five subgroups, such as SK_3S arrangement in one DHN protein of chickweed (Z21500; Close 1996). These considerable research efforts have been employed in exploring DHNs structure and function for herb model plants, such as *Arabidopsis*, maize and barley, but such in-depth study has not yet been directed towards woody trees.

The genes encoding DHN are a multigene family (Hundertmark and Hincha 2008). Recent studies, together with the release of complete genome sequences for different organisms, have led to the identification of *DHN*s in single plant genome. In previous published reports, 12, 9, and 10 *DHN* genes had successively been identified using different methods in *Arabidopsis* (Hundertmark and Hincha 2008; Tunnacliffe and Wise 2007; Alsheikh et al. 2005), 13 in barley (Choi et al. 1999; Rodriguez et al. 2005), 8 in rice (Wang et al. 2007). In addition, so far, only one SK_2 type *DHN* in different poplars was successively identified and their response to various stresses was confirmed (Bae et al. 2009; Caruso et al. 2002). Even though genes encoding DHNs have been identified in several plant species, to date, there is still no comprehensive and systematic study characterizing all *DHN* genes in a single woody plant genome. In order to explore all genes encoding DHN proteins in poplar, complete *Populus trichocarpa* genome was investigated using the method of domain search. Here, we exhibit an identification and analysis of DHN proteins and their respective genes in *P. trichocarpa*. As we know, this is the first systematic characterization of all genes encoding DHN proteins in a single woody plant genome, and represents the basis for future studies on the in vivo each poplar DHN function.

Methods

Identification and chromosomal location of poplar DHN genes

The complete protein sequence database was downloaded from *P. trichocarpa* v1.1 (www.jgi.doe.gov/poplar). Hidden Markov Model (HMM) profile file (dehydrin.hmm) of the Pfam Dehydrin domain (PF00257) was downloaded from the Pfam database (http://pfam.sanger.ac.uk/). The dehydrin.hmm file was exploited as a query to identify the *DHN* genes in the poplar protein database using the hmmer search command of the HMMER (v 3.0) software, which was widely applied for identification of homologues of an interested protein family (Finn et al. 2010; Eddy 2009). All nonredundant (Nr) hits with expected values less than 0.1 were collected, and then were respectively searched applying BLASTP program across REFseq Nr protein database in NCBI (http://www.ncbi.nlm.nih.gov/). The expressed sequence tags (EST) were retrieved by BLASTN the corresponding transcript/CDS from *P. trichocarpa* v1.1 (www.jgi.doe.gov/poplar) as query sequence online search against all of the *Populus* EST sequences in NCBI. Matches above 95% identity and over an alignment of at least 100 bp were considered as corresponding sequences of the dehydrin genes. Multiple sequences alignments of these sequences with their individual transcript/CDS sequence were performed using ClustalW program in BioEdit software under the default parameters settings (Hall 1999). Sequence alignments were manually adjusted to get maximum matching.

The 11 identified *DHN* genes were located in the genome of *P. trichocarpa* using NCBI map viewer (http://www.ncbi.nlm.nih.gov/projects/mapview/). Identification of duplicated regions between chromosomes was completed as described in Tuskan et al. (2006). The tandem gene duplication in poplar was determined according to the criteria that five or fewer gene loci occurred within a range of 100 kb distance (Hu et al. 2010; Finn et al. 2006).

Identification of conserved motifs

Extraction of motifs from 34 DHN protein sequences in poplar, *Arabidopsis* and barley, are performed using the software of MEME online-version 4.6.1 (Multiple Expectation Maximization for motif Elicitation), which is one of the most widely used tools for observation of new sequence patterns in biological sequences and analysis of their significance (Bailey and Elkan 1994; Bailey et al. 2006). MEME program is run with the following parameters: the optimum number for each motif is between 2 and 120, distribution of motif occurrences is any number of repetitions, maximum number of motifs is 15, and the optimum motif widths were restricted between 8 and 16 residues.

Phylogenetic analysis and in silico microarray analysis

Multiple sequences alignments of the full-length protein sequences were performed using ClustalW program in BioEdit software with default parameters (Hall 1999). Based on these aligned sequences, the unrooted phylogenetic trees were constructed using MEGA 5.0 software (Tamura et al. 2011), by both Neighbor-joining method (Saitou and Nei 1987) and Minimum Evolution method with the parameters (p-distance and completed deletion). The reliability of the phylogenetic tree was estimated using bootstrap value with 1000 replicates. Probe sets corresponding to individual poplar *DHN* gene were retrieved using an online probe match tool available at NetAffx^{TM} Analysis center (http://www.affymetrix.com/analysis/index.affx). The transcript relative abundance values of all poplar *DHN* genes from various tissues were obtained from the poplar transcript abundances datasets (Wilkins et al. 2009) in the website of the *Populus* electronic fluorescent pictograph browser (Poplar eFP browser; http://bar.utoronto.ca/efppop/cgi-bin/efpWeb.cgi), whose data originated from the NCBI Gene Expression Omnibus (accession number: GSE13990). For genes with more than one probe set, the mean expression values were considered. When several genes have the same probe set, then they are considered as the same level of transcript abundance. Dendrogram and heat map for display expression pattern were obtained using the Cluster 3.0 (de Hoon et al. 2004) for normalizing and hierarchical clustering with average linkage based on Pearson coefficients, and then Java Tree-View 1.1 program (Saldanha 2004) for visualizing the analyzing datasets.

Results and Discussion

Identification and characterization of *DHN* gene family in *Populus*

To identify *DHN* genes and their putative encoded polypeptides present in *Populus* genome, initially, keyword search of "dehydrin" against *P. trichocarpa* genome database was performed (www.jgi.doe.gov/poplar). It was found that nine members had been annotated as *DHN* genes displayed in Nos. 1–9 of Table 1; Subsequently, aim to confirm this reliability of these identified genes, HMM profile file (dehydrin.hmm) of the Pfam Dehydrin domain (PF00257) was exploited as query file for search across *P. trichocarpa* genome (www.jgi.doe.gov/poplar). A total of 10 non-redundant putative *DHN* family genes were identified as significantly encoding dehydrin domain, of which eight (No. 1–8 of Table 1) were included and two (No. 10 and 11, 817,405 and 276,757) were not. The detailed information of *DHN* family genes in poplar was listed in Table 1. In addition, to provide a simplified nomenclature for each identified gene, all the genes (and corresponding proteins) were denominated as *PtrDHN* (Table 1), and the followed digit represents the gene number within the group.

Table 1 All identified dehydrin genes and putative encoded poplypeptides present in *Populus trichocarpa* genome

NO.	JGI protein ID	Gene and transcript products		Chromosome location	Protein products	
		Novel simplified gene nomenclature	Transcript ID		Protein ID	Novel simplified nomenclature
1[a,b]	550802	*PtrDHN-1.1*	XM_002300629.1	LG_II:763543–765920 (+)	XP_002300665.1	PtrDHN-1.1
		PtrDHN-1.2	AJ300525.4	LG_II:763543–765920 (+)	CAC18724.4	PtrDHN-1.2
2[a,b]	649369	*PtrDHN-2*	XM_002313741.1	LG_IX:2865959–2867055 (−)	XP_002313777.1	PtrDHN-2
3[a,b]	818850	*PtrDHN-3*	XM_002307732.1	LG_V:17250674–17251640 (−)	XP_002307768.1	PtrDHN-3
4[a,b]	571249	*PtrDHN-4*	XM_002319676.1	LG_XIII:4591285–4595783 (−)	XP_002319712.1	PtrDHN-4
5[a,b]	582807	*PtrDHN-5*	XM_002334420.1	scaffold_1432:11124–12141 (+)	XP_002334457.1	PtrDHN-5
6[a,b]	571250	*PtrDHN-6*	XM_002319677.1	LG_XIII:4601199–4602375 (−)	XP_002319713.1	PtrDHN-6
7[a,b]	663123	*PtrDHN-7*	XM_002319678.1	LG_XIII:4609807–4610954 (−)	XP_002319714.1	PtrDHN-7
8[a,b]	195568	*PtrDHN-8*	NA	LG_IV:14780975–14782085 (+)	NA	PtrDHN-8
9[a]	665494	*PtrDHN-9*	NA	LG_XIX:5513433–5526222 (−)	NA	PtrDHN-9
10[b]	817405	*PtrDHN-10*	XM_002303606.1	LG_III:12860185–12860889 (+)	XP_002303642.1	PtrDHN-10
11[b]	276757	*PtrDHN-11*	NA	scaffold_18688:3–1113 (+)	NA	NA

NA denotes not available, while *LG* represents Linkage Group

[a] Represents these identified DHNs gained by keyword search of "dehydrin" against *P. trichocarpa* genome database

[b] Represents these identified DHNs gained by Dehydrin domain (PF00257) search of *P. trichocarpa* genome database

The gene (*PtrDHN-9*, 665494) described as "dehydrin" in *P. trichocarpa* v1.1 appears to be incorrect annotation because of an absence of significant DHN domain throughout its encoding polypeptide. However, our sequence analysis indicated that its encoding protein had high sequence similarity with three KS-type of DHN proteins documented previously in *Arabidopsis* (Hundertmark and Hincha 2008), soybean (Alsheikh et al. 2005) and barley (Rodriguez et al. 2005), especially with their K-, S-, and Φ-segments (Fig. 1). But, it is especially note worthy that, consecutive deletion for four amino acids of "KIKD" were discovered in its K-segment (Fig. 1), being able to explain why it does not match the DHN domain (PF00257) in our domain search. Due to the few deletions in K-segment and high sequence identity with other plant DHN proteins, the *PtrDHN-9* gene was also defined as *DHN* gene.

Thus, in our study, a total of 11 *DHN* genes were finally identified in *P. trichocarpa* genome by the genome-wide survey (Table 1). The number of *DHN* genes in *P. trichocarpa* is roughly equal to that of *Arabidopsis*, which is not in agreement with the ratio of 1.4~1.6 putative *Populus* homologues to each *Arabidopsis* gene according to comparative genomics studies (Tuskan et al. 2006). In contrast, the expansion, often present on a large number of *Populus* multigene families (Tuskan et al. 2006), seems not to occur in *Populus DHN* gene family. It could be speculated that the presence of similar number of *DHN* genes in *Populus* genome might reflect the analogous needs for these genes involving in their specific stress-related function.

Revising of DHN gene-encoding proteins as well as discovering of alternative splicing present in poplar *DHN* genes

Given the current draft nature of the *Populus* genome (www.jgi.doe.gov/poplar), where a first-draft reference set of 45,555 protein-coding gene loci was tentatively identified, the gene set in *Populus* will need to be refined gradually (Tuskan et al. 2006). To calibrate our preliminary identification of the eleven *DHN* genes from JGI poplar database, their encoding proteins were further compared by a BLASTP search against NCBI Reference sequence (RefSeq) database, which provides a non-redundant and validated collection of sequences representing genomic data, transcripts and proteins (Pruitt et al. 2006, 2005). As a result, among them, the three poplar DHN proteins (PtrDHN-8, PtrDHN-11, and PtrDHN-9) without counterparts in NCBI RefSeq database (Table 1), may represent truncated or incorrect proteins. Their corresponding EST were retrieved by BLASTN online search to obtain support and mend them for further analysis. These ESTs from NCBI perfectly matched CDS sequences, particularly for the nucleotide acid sequences encoding amino acid sequences of K-segment, were selected for alignment with their individual transcript/CDS from *P. trichocarpa* v1.1 (Electronic Supplementary Material (ESM) Fig. S1–3). As for the transcript of *PtrDHN-9* (665494), a large number of EST support "ATG" at position 49~51 as translation start codon, "TAA" at position 337~339 as translation stop codon (ESM Fig. S1). According to this, the encoded amino acids after the "TAA" were removed from the original *PtrDHN-9* encoding protein sequence (ESM Fig. S1 and ESM Table S1). The absence of translation start codon "ATG" lead to the incomplete N-terminus of *PtrDHN-8* (195568) protein, our EST sequence alignment and comparative analyses clearly demonstrated that upstream of the first three nucleotides "GCC" from *PtrDHN-8* transcript should be extended by the "ATG" encoding Met as initiation codon as well as the followed "GCT" encoding Ala (ESM Fig. S2 and ESM Table S2). Moreover, "TAG" at position 394~396 was strongly supported by ESTs as its translation terminator codon (ESM Fig. S2). The revised CDS and encoding protein sequence of *PtrDHN-8* were displayed in ESM Table S2; The gene *PtrDHN-11* (276757) had no significant EST match, and "TAG" at position 196~198 of transcript as stop codon caused the early translation termination (ESM Fig. S3 and ESM Table S3). Based on this revised CDS sequence, its encoding amino acid sequence in the front of the stop codon was determined not to match any DHN domain (ESM Fig. 3 and ESM Table S3). Therefore, it was

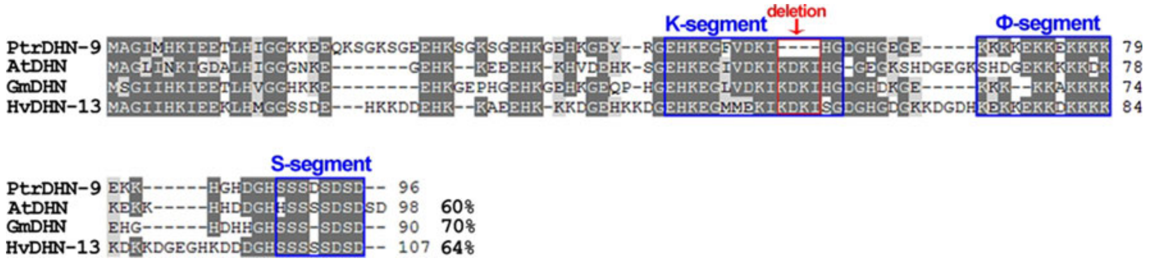

K-segment
deletion
Φ-segment
PtrDHN-9 MAGIMHKIEETLHIGGKKEEQKSGKSGEEHKSGKSGEHKGEHKGEY--RGEHKEGFVDKI----HGDGHGEGE-----KKKKEKKEKKKK 79
AtDHN MAGLINKIGDALHIGGGNKE-------GEHK--KEEEHK-KHVDEHK-SGEHKEGIVDKIKDKIHG-GEGKSHDGEGKSHDGEKKKKKKDK 78
GmDHN MSGIIHKIEETLHVGGHKKE--------EHKGEPHGEHKGEHKGEQP-HGEHKEGLVDKIKDKIHGDGHDKGE-----KK--KKAKKKK 74
HvDHN-13 MAGIIHKIEEKLHMGGSSDE---HKKDDEHK--KAEEHK-KKDGEHKKDGEHKEGMMEKIKDKISGDGHGDGKKDGDHKEKKEKKDKKKK 84
S-segment
PtrDHN-9 EKK------HGHDGHSSSDSDSD-- 96
AtDHN KEKK-----HHDDGHHSSSSSDSDSD 98 60%
GmDHN EHG------HDHHGHSSS-SDSD-- 90 70%
HvDHN-13 KDKKDGEGHKDDDGHSSSSSSDSD-- 107 64%

Fig. 1 Multiple sequence alignment of *Populus* PtrDHN-9 with other plant DHNs. K-, S-, and Φ-segments in our study are marked with *a blue box* under the corresponding description. Consecutive deletion for four amino acids of "KIKD" in K-segment of PtrDHN-9 is displayed with *a red box* under the description of "deletion". *Gray shading* represent 70% identical residues among the sequences. PtrDHN-9 (JGI Protein ID, 665494); AtDHN (At1g54410.1) from *Arabidopsis*; GmDHN (ABO70349.1) from soybean; HvDHN-13 (AAT81473.1) from barley

excluded from the identified 11 *DHN* gene of poplar above mentioned, but identified as putative pseudogene of *DHN* because of its high sequence identity with another *DHN* gene *PtrDHN-1* (550802). In this endeavor, two (*PtrDHN-9* and *PtrDHN-8*) out of the three problematic transcripts were confirmed by EST support with high confidence, and modified into complete protein, whereas the third gene *PtrDHN-11* (276757) was identified as pseudogene of *DHN*.

Processing of alternative transcripts as a mechanism of regulation of gene expression plays a direct role in plant development (Wang and Brendel 2006). Though recent computational studies in *Arabidopsis* and rice have estimated that over 20% of genes are alternatively spliced in both species (Filichkin et al. 2010; Iida et al. 2004; Wang and Brendel 2006), the presence of alternative splicing in *DHN* genes has not been reported. In our study, one cDNA sequence for putative dehydrin (AJ300525.4) from *Populus euramericana*, which has not previously been mapped to poplar genome, has very high identity with CDS of *PtrDHN-1* (550802) gene. Comparison between both of them and the genomic sequence of *PtrDHN-1* gene revealed the presence of alternative splicing in *PtrDHN-1* genes, the cDNA sequence (AJ300525.4) and CDS being its two splicing isoforms (Fig. 2a and Table 1). Their encoding products are PtrDHN-1.2 (CAC18724.4) and PtrDHN-1.1 (XP_002300665.1),

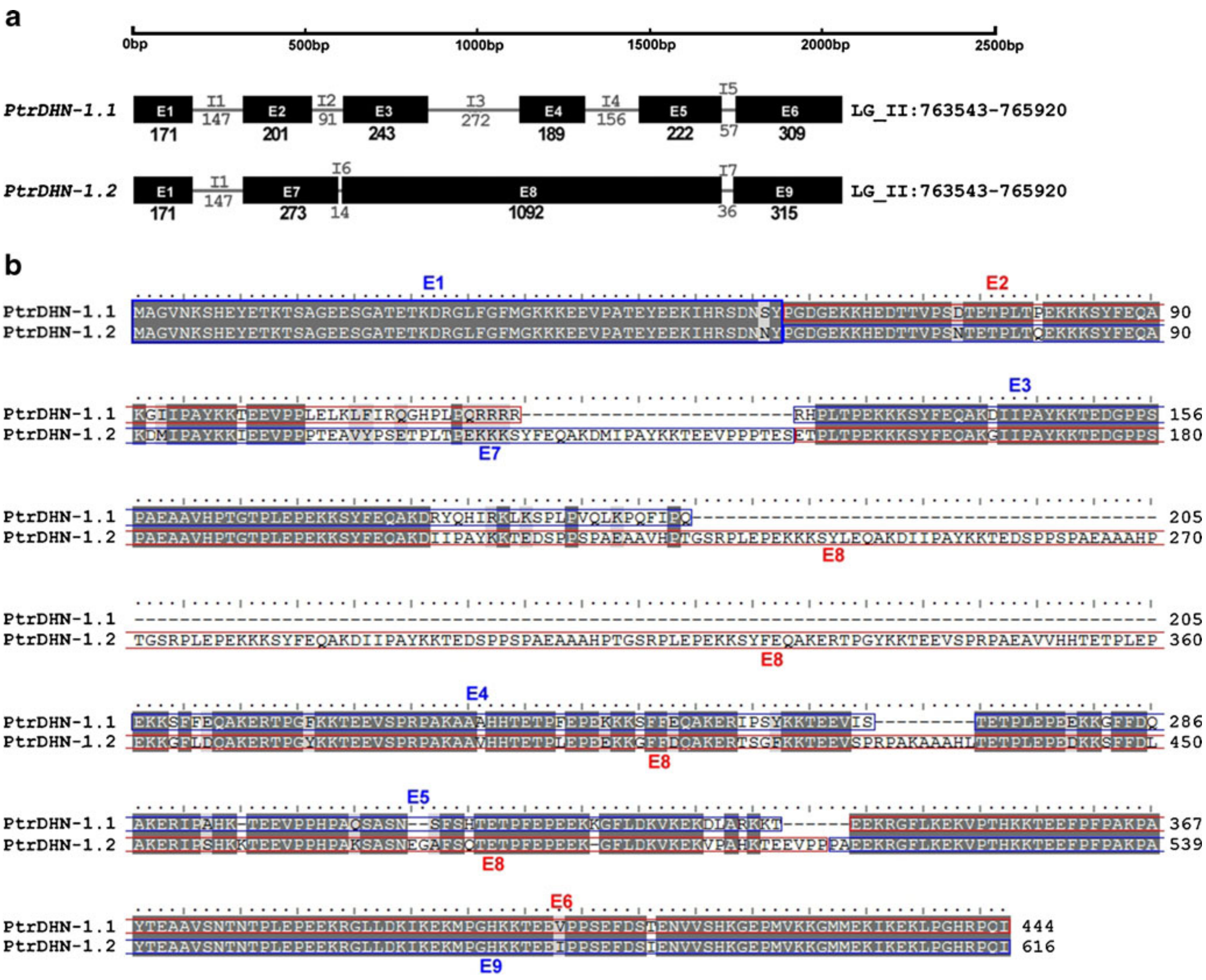


Fig. 2 Schematic representation of the intron/exon structure of the alternatively spliced transcripts and their protein product isoforms for *PtrDHN-1* gene. **a** Display the intron/exon structure of the two alternatively spliced transcripts (*PtrDHN-1.1* and *PtrDHN-1.2*) for *PtrDHN-1* gene. **b** Sequence alignment of protein product isoforms encoded by the two alternatively spliced transcripts of *PtrDHN-1* gene for comparison of their exons encoding peptides. Exons sequences are represented by *black boxes* and numbered E1–E9, while intron sequence are indicated with *gray lines* and are numbered I1–I7. Base pairs length of exons and introns was shown under each region, and also can be estimated by the scale at the top. The names of the alternatively spliced transcripts are shown on the *left*, with their chromosomal location on the *right*. E1–E9 in (**b**) represents each exons encoding peptides

respectively. In order to reveal the nature of the splicing variation, the positions of exons and introns were determined based on their sequence alignment with the genomic sequence, and their intron/exon structure and product isoforms of both the alternatively spliced transcripts were displayed as Fig. 2a and b. The presence of alternative splicing in *PtrDHN-1* (550802) gene resulted in its encoding two splicing isoforms of *PtrDHN-1.1* and *PtrDHN-1.2* (Table 1). Therefore, 10 *DHN* genes encoding 11 DHN proteins were identified in total in our genome-wide investigation of *DHN* genes.

Chromosomal location and duplication of *DHN* gene in *Populus*

In silico mapping of the gene loci showed that, except for the two *DHN* genes of *PtrDHN-5* and *PtrDHN-11* assigned to individual scaffold fragments, the others were distributed across 7 of 19 Linkage Groups (LG; Table 1 and Fig. 3). The distribution of the *DHN* genes among these LGs appears to be uneven: LG II, III, IV, V, IX, and XIX individual have only one *DHN* gene, high density of *DHN* genes was discovered in LG XIII, where three *DHN* genes (*PtrDHN-4*, *-6* and *-7*) were organized in one cluster (Cluster I) within a 20 kb fragment (Fig. 3).

Previous analysis of *Populus* genome has identified the presence of paralogous segments caused by the whole-genome duplication event in the Salicaceae (salicoid duplication), which occurred 65 million years ago and significantly contributed to the amplification of many multi-gene families (Tuskan et al. 2006). To determine the possible relationship between the *DHN* genes and paralogous segments, the *Populus DHN* genes were mapped to the duplicated blocks of *P. trichocarpa* established in the studies of Tuskan et al. (2006). The distribution of *DHN* genes relative to the duplicated blocks is illustrated as Fig. 3. It was found that all the nine mapped *DHN* genes (100%), are located in duplicated blocks. Two duplicated pairs (*PtrDHN-1/3* and *PtrDHN-2/8*) are each located in a pair of paralogous blocks and can be considered as direct results of the segmental duplication event (Fig. 3). Similarly, Cluster I/*PtrDHN-9* also corresponds to a pair of paralogous blocks created by the whole-genome duplication event (Fig. 3). One duplicated pair (*PtrDHN-10*) harbored *DHN* genes on only one of the blocks and lack corresponding duplicates, suggesting that the loss event of its corresponding paralogous genes should have occurred after the segmental duplication events (Fig. 3). The findings support the result that the most abundant genes losses in eukaryotes occur following the whole genome duplication (Abdel-Haleem 2007).

Furthermore, the tandem duplications also contribute to the expansion of *DHN* gene family. In LG XIII, there is one *DHN* cluster (Cluster I) with three genes tandem arranged in the same orientation spanning a 20-kb fragment (Table 1 and Fig. 3). Together with the high sequence identities among them, the three tandem *DHN* genes within Cluster I were considered to be direct results of the tandem duplication events. Their organization in duplicated blocks implied

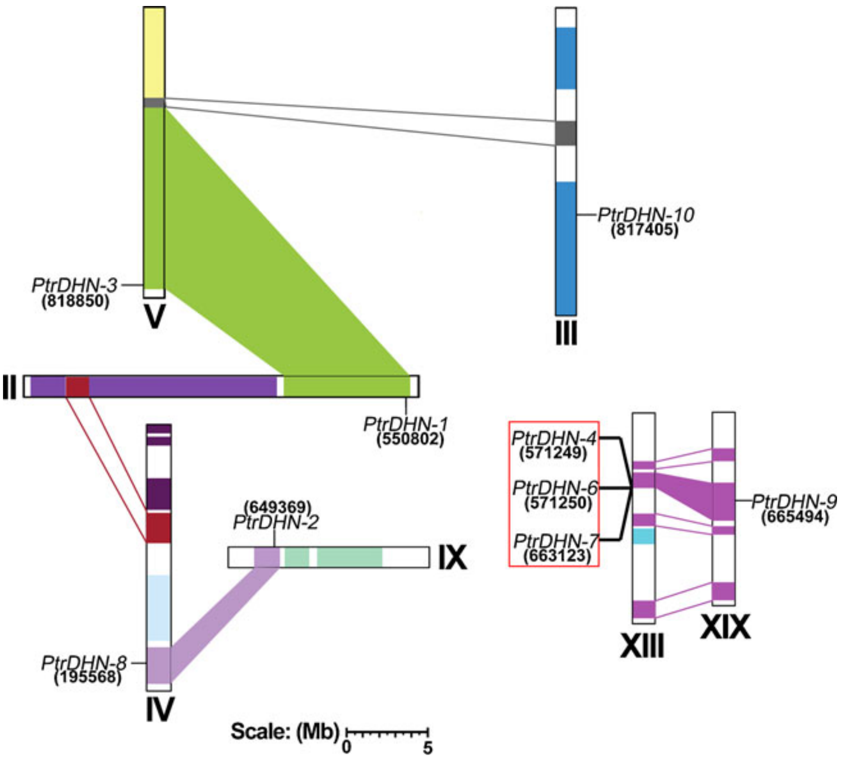


Fig. 3 Chromosomal location of the *Populus DHN* genes. Nine genes are mapped to the 7 of 19 Linkage Groups (LG), while the other two genes located on unassembled scaffolds. The schematic representation of genome-wide chromosome organization arisen from the whole-genome duplication event in *Populus* was obtained from (Tuskan et al. 2006). Segmental duplicated homologous regions are shown with the same color. Only the duplication blocks containing *DHN* genes are connected with lines in shaded colors. Three tandemly duplicated genes within 20 kb displayed with *red box* were organized into one cluster (Cluster I). Scale at the bottom represents a 5-Mb chromosomal distance

that the presence of the segmental duplication events was prior to the tandem duplication. According to the genomic organization of *DHN* genes, segmental duplication as well as tandem duplication events contributed to the expansion of *DHN* gene family in the *Populus* genome. Similarly, the two events had also been shown to contribute to the expansion of *DHN* genes in *Arabidopsis* (Hundertmark and Hincha 2008) and rice (Wang et al. 2007).

In our study, *Populus DHN* gene family has been preferentially retained at a rate of 100%, while in *Populus* genome, about only one-third of putative genes are retained in duplicated blocks resulting from the whole genome duplication events (Tuskan et al. 2006). The high retention rate of duplicated genes had also previously been documented in other *Populus* gene families (Hu et al. 2010; Barakat et al. 2009; Kalluri et al. 2007). In addition, the segmental duplication ratio of *DHN* genes in this study is predominantly higher than that of the tandem duplication, suggesting that the segmental duplication might be main events contributing to the expansion of *Populus DHN* genes.

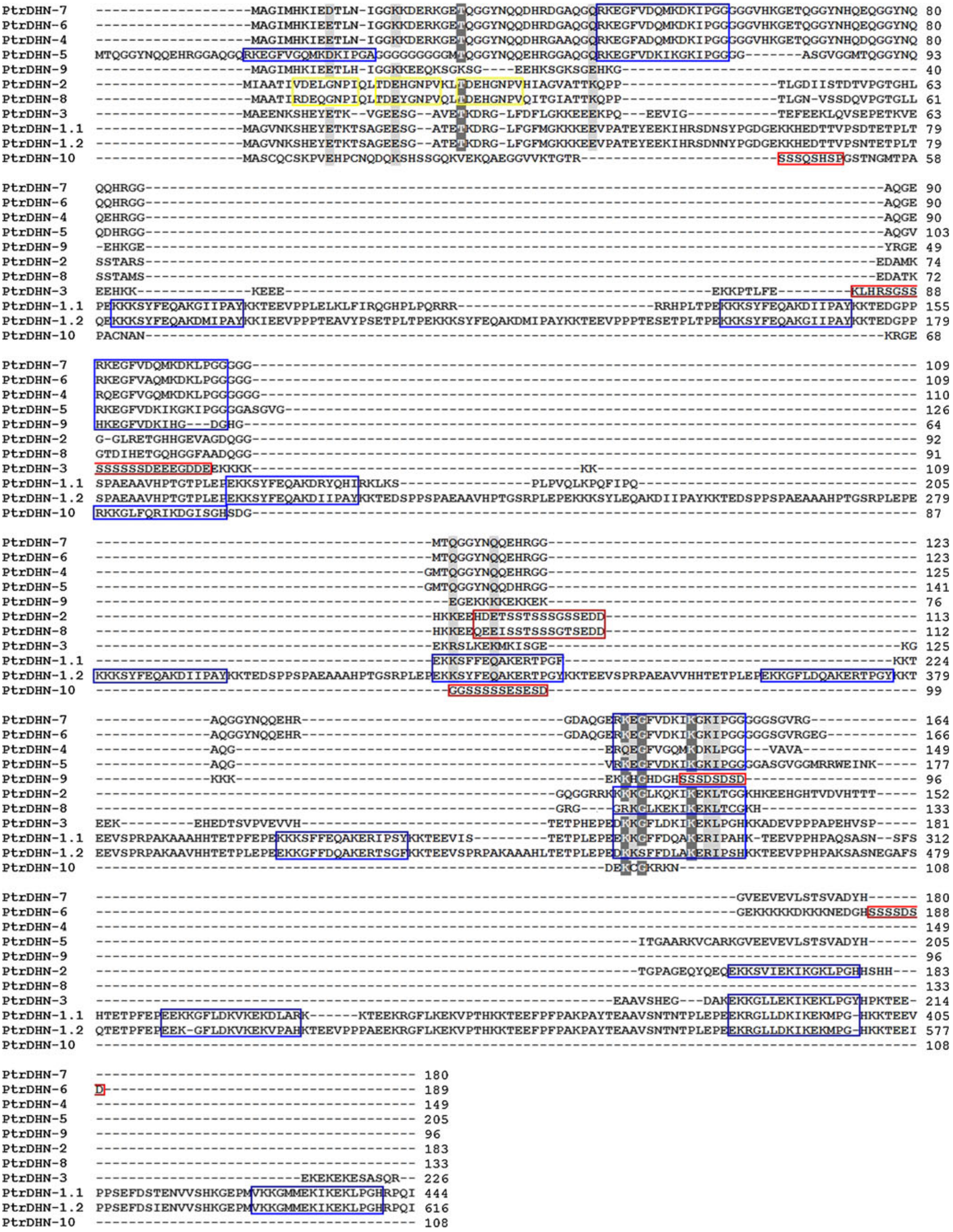
Identification of conserved motif and classification of *Populus* DHN proteins

The conserved K-segment is the most distinctive feature in motifs of all DHNs, while other motifs of S- and Y-segments are also identified as important motifs (Rorat 2006; Close 1996). To reveal these motifs present on *Populus* DHN proteins, extraction of motifs was performed by MEME based on total 34 DHN protein sequences from poplar, *Arabidopsis* (Hundertmark and Hincha 2008), and barley (Choi et al. 1999; Rodriguez et al. 2005; Fig. 4, ESM Table S4, S5 and S6). As a result, six significant motifs were retrieved (ESM Fig. S4a–f), among which motif-1 (ESM Fig. S4a) and motif-4 (ESM Fig. S4d) were respectively identified as K- and Y-segments based on their good identities with previous K- and Y-motifs of DHNs. Both motif-2 (ESM Fig. S4c) and motif-3 (ESM Fig. S4d), characterized by a track of Ser residue, were considered as S-segment, in order to distinguish them, motif-2 with width of eight amino acid residues are designated as “S-segment”, motif-3 with width of 16 residues as “S-segment”. In contrast, motif-5 and motif-6 were identified as novel motifs in DHNs because of no known homologous motifs matched in Pfam and SMART databases (ESM Fig. S4e and 4f). Our additional investigation showed that besides DHNs of *Populus* (4/11; Fig. 5b), the motif-5 still widespread occurred in those of barley (7/13) and rice (4/6), but no occurrence in those of *Arabidopsis* (ESM Fig. S5); instead, the other novel motif-6 was rarely present, except for DHNs of *Populus* (3/11), only one (At1g20440.1, YSK$_3$) in those of *Arabidopsis* (ESM Fig. S5). However, whether the two novel motifs confer unique functional roles to DHNs remains to be further investigated.

Based on the generally accepted classification for DHNs (Rorat 2006; Close 1996), it was found that, the eleven poplar DHNs were assigned to four out of the five subgroups, the K$_n$ subgroups of DHNs were the most numerous, being represented by five members. Y$_n$SK$_n$ and K$_n$S were each represented by 2 members, and SK$_n$ DHN subgroups were represented by just one member (Table 2, Fig. 4, ESM Table S4 and Fig. 5b). Interestingly, the remaining one protein (PtrDHN-10) with the SKS composition, cannot be assigned to a certain class of DHNs (Fig. 5b and ESM Table S4). It probably represents an intermediate form of SK$_n$ and K$_n$S, which had been documented to occur in one DHN proteins of *Stellaria longipes* (Z21500; SK$_3$S; Close 1996). It was note worthy that one K$_n$ subgroup of DHN member (PtrDHN-1.2) was characterized by 13 repeated K-segments (K$_{13}$; Fig. 5b), of which a maximum repeated number has previously been documented of being from spinach CAP85 (K$_{11}$; Kaye et al. 1998). K$_n$ subgroup of poplar DHNs become the most predominant DHNs with highest proportional occurrence (5/11) than those of *Arabidopsis* (1/10) and barley (1/13; Fig. 4, ESM Fig. S5, ESM Table S4, S5 and S6). However, it had been confirmed that the K$_n$ subgroup DHNs from various plants were induced by cold temperature, dehydration, and ABA, and seem to be directly involved in cold acclimation processes (Kosova et al. 2007; Rorat 2006). Thus, the significantly enriched K$_n$ subgroup DHNs are present in *Populus* genome, suggesting that the more DHNs of the direct responsibility for cold acclimation is required for woody plants. This should be a possible reason why woody plants contain more cold-inducible K$_n$ type DHN genes than herbaceous plants. Furthermore, our evidence also indicated that the Y$_n$SK$_n$ subgroup of DHNs were the highest proportional occurrence on *Arabidopsis* (5/10) and barley (9/13; ESM Table S5 and S6), in contrast, fewer members (2/11) are assigned to the Y$_n$SK$_n$ subgroup in *Populus* DHNs (Table 2 and ESM Table S4). It could be explained that the presence of more K$_n$ subgroup than *Arabidopsis* and barley were caused by the loss of gene sequence encoding Y-segments in *Populus* DHNs after the occurrence of evolutional divergence between *Populus*, *Arabidopsis*, and barley.

Divergence within *Populus* DHN genes

An unrooted tree was, respectively, generated by both Neighbor-Joining (Saitou and Nei 1987) and Minimum-Evolution methods using MEGA 5.0 (Tamura et al. 2011) based on complete protein sequences of all the DHN genes in *Populus*. The tree topologies generated by the two methods were comparable without modifications at branches, and supported by their high bootstrap values of >55, suggesting



```
PtrDHN-7     -------------------MAGIMHKIEDTLN-IGGKKDERKGETQGGYNQQDHRDGAQGQRKEGFVDQMKDKIPGGGGGVHKGETQGGYNHQEQGGYNQ 80
PtrDHN-6     -------------------MAGIMHKIEETLN-IGGKKDERKGETQGGYNQQDHRDGAQGQRKEGFVDQMKDKIPGGGGGVHKGETQGGYNHQEQGGYNQ 80
PtrDHN-4     -------------------MAGIMHKIEETLN-IGGKKDERKGETQGGYNQQDHRGAAQGQRKEGFADQMKDKIPGGGGGVHKGETQGGYNHQDQGGYNQ 80
PtrDHN-5     MTQGGYNQQEHRGGAQGQRKEGFVGQMKDKIPGAGGGGGGGGGMTQGGYNQQEHRGGAQGQRKEGFVDKIKGKIPGGGG-------ASGVGGMTQGGYNQ 93
PtrDHN-9     -------------------MAGIMHKIEETLH-IGGKKEEQKSGKSG----EEHKSGKSGEHKG------------------------------------ 40
PtrDHN-2     ------------------MIAATIVDELGNPIQLTDEHGNPVKLTDEHGNPVHIAGVATTKQPP-------------------TLGDIISTDTVPGTGHL 63
PtrDHN-8     -------------------MAATIRDEQGNPIQLTDEYGNPVQLTDEHGNPVQITGIATTKQPP-------------------TLGN-VSSDQVPGTGLL 61
PtrDHN-3     ------------------MAEENKSHEYETK--VGEESG--AVETKDRG-LFDFLGKKEEEKPQ---EEVIG-----------TEFEEKLQVSEPETKVE 63
PtrDHN-1.1   ------------------MAGVNKSHEYETKTSAGEESG--ATETKDRG-LFGFMGKKKEEVPATEYEEKIHRSDNSYPGDGEKKHEDTTVPSDTETPLT 79
PtrDHN-1.2   ------------------MAGVNKSHEYETKTSAGEESG--ATETKDRG-LFGFMGKKKEEVPATEYEEKIHRSDNNYPGDGEKKHEDTTVPSNTETPLT 79
PtrDHN-10    ------------------MASCQCSKPVEHPCNQDQKSHSSGQKVEKQAEGGVVKTGTR------------------------SSSQSHSPGSTNGMTPA 58

PtrDHN-7     QQHRGG------------------------------------------------------------------------------------------AQGE 90
PtrDHN-6     QQHRGG------------------------------------------------------------------------------------------AQGE 90
PtrDHN-4     QEHRGG------------------------------------------------------------------------------------------AQGE 90
PtrDHN-5     QDHRGG------------------------------------------------------------------------------------------AQGV 103
PtrDHN-9     -EHKGE------------------------------------------------------------------------------------------YRGE 49
PtrDHN-2     SSTARS-----------------------------------------------------------------------------------------EDAMK 74
PtrDHN-8     SSTAMS-----------------------------------------------------------------------------------------EDATK 72
PtrDHN-3     EEHKK-------------KEEE---------------------------------------------------EKKPTLFE---------KLHRSGSS 88
PtrDHN-1.1   PEKKKSYFEQAKGIIPAYKKTEEVPPLELKLFIRQGHPLPQRRR------------------------RRHPLTPEKKKSYFEQAKDIIPAYKKTEDGPP 155
PtrDHN-1.2   QEKKKSYFEQAKDMIPAYKKIEEVPPPTEAVYPSETPLTPEKKKSYFEQAKDMIPAYKKTEEVPPPTESETPLTPEKKKSYFEQAKGIIPAYKKTEDGPP 179
PtrDHN-10    PACNAN------------------------------------------------------------------------------------------KRGE 68

PtrDHN-7     RKEGFVDQMKDKLPGGGGG--------------------------------------------------------------------------------- 109
PtrDHN-6     RKEGFVAQMKDKLPGGGGG--------------------------------------------------------------------------------- 109
PtrDHN-4     RQEGFVGQMKDKLPGGGGGG-------------------------------------------------------------------------------- 110
PtrDHN-5     RKEGFVDKIKGKIPGGGGASGVG----------------------------------------------------------------------------- 126
PtrDHN-9     HKEGFVDKIHG---DGHG---------------------------------------------------------------------------------- 64
PtrDHN-2     G-GLRETGHHGEVAGDQGG--------------------------------------------------------------------------------- 92
PtrDHN-8     GTDIHETGQHGGFAADQGG--------------------------------------------------------------------------------- 91
PtrDHN-3     SSSSSSDEEEGDDEEKKKK---------------------------------------KK---------------------------------------- 109
PtrDHN-1.1   SPAEAAVHPTGTPLEPEKKSYFEQAKDRYQHIRKLKS---------------PLPVQLKPQFIPQ----------------------------------- 205
PtrDHN-1.2   SPAEAAVHPTGTPLEPEKKSYFEQAKDIIPAYKKTEDSPPSPAEAAVHPTGSRPLEPEKKKSYLEQAKDIIPAYKKTEDSPPSPAEAAAHPTGSRPLEPE 279
PtrDHN-10    RKKGLFQRIKDGISGHSDG--------------------------------------------------------------------------------- 87

PtrDHN-7     ----------------------------------------MTQGGYNQQEHRGG---------------------------------------------- 123
PtrDHN-6     ----------------------------------------MTQGGYNQQEHRGG---------------------------------------------- 123
PtrDHN-4     ---------------------------------------GMTQGGYNQQEHRGG---------------------------------------------- 125
PtrDHN-5     ---------------------------------------GMTQGGYNQQDHRGG---------------------------------------------- 141
PtrDHN-9     ------------------------------------------EGEKKKKEKKEK---------------------------------------------- 76
PtrDHN-2     ----------------------------------------HKKEEHDETSSTSSSGSSEDD--------------------------------------- 113
PtrDHN-8     ----------------------------------------HKKEEQEEISSTSSSGTSEDD--------------------------------------- 112
PtrDHN-3     ----------------------------------------EKRSLKEKMKISGE--------------------------------------------KG 125
PtrDHN-1.1   ----------------------------------------EKKSFFEQAKERTPGF-------------------------------------------KKT 224
PtrDHN-1.2   KKKSYFEQAKDIIPAYKKTEDSPPSPAEAAAHPTGSRPLEPEKKSYFEQAKERTPGYKKTEEVSPRPAEAVVHHTETPLEPEKKGFLDQAKERTPGYKKT 379
PtrDHN-10    ------------------------------------------GGSSSSSESESD---------------------------------------------- 99

PtrDHN-7     -------------AQGGYNQQEHR--------------------------------GDAQGERKEGFVDKIKGKIPGGGGSGVRG------------ 164
PtrDHN-6     -------------AQGGYNQQEHR--------------------------------GDAQGERKEGFVDKIKGKIPGGGGSGVRGEG---------- 166
PtrDHN-4     -------------AQG------------------------------------------------ERQEGFVGQMKDKLPGG---VAVA----------- 149
PtrDHN-5     -------------AQG------------------------------------------------VRKEGFVDKIKGKIPGGGASGVGGMRRWEINK----- 177
PtrDHN-9     -------------KKK------------------------------------------------EKKHGHDGHSSSDSDSD--------------------- 96
PtrDHN-2     ---------------------------------------------------------GQGGRRKKKGLKQKIKEKLTGGKHKEEHGHTVDVHTTT----- 152
PtrDHN-8     ---------------------------------------------------------GRG---GRKGLKEKIKEKLTCGKH------------------- 133
PtrDHN-3     EEK---------EHEDTSVPVEVVH---------------------------------TETPHEPEDKKGFLDKIKEKLPGHKKADEVPPPAPEHVSP----- 181
PtrDHN-1.1   EEVSPRPAKAAAHHTETPFEPEKKKSFFEQAKERIPSYKKTEEVIS---------TETPLEPEEKKGFFDQAKERIPAHK-TEEVPPHPAQSASN--SFS 312
PtrDHN-1.2   EEVSPRPAKAAVHHTETPLEPEEKKGFFDQAKERTSGFKKTEEVSPRPAKAAAHLTETPLEPEDKKSFFDLAKERIPSHKKTEEVPPHPAKSASNEGAFS 479
PtrDHN-10    ------------------------------------------------------------DEKCGKRKN--------------------------- 108

PtrDHN-7     --------------------------------------------------------------------GVEEVEVLSTSVADYH------ 180
PtrDHN-6     --------------------------------------------------------------------GEKKKKKDKKKNEDGHSSSSDS 188
PtrDHN-4     ------------------------------------------------------------------------------------------ 149
PtrDHN-5     ---------------------------------------------------------ITGAARKVCARKGVEEVEVLSTSVADYH------ 205
PtrDHN-9     ------------------------------------------------------------------------------------------ 96
PtrDHN-2     ---------------------------------------------------------TGPAGEQYQEQEKKSVIEKIKGKLPGHHSHH--- 183
PtrDHN-8     ------------------------------------------------------------------------------------------ 133
PtrDHN-3     ------------------------------------------------------EAAVSHEG---DAKEKKGLLEKIKEKLPGYHPKTEE- 214
PtrDHN-1.1   HTETPFEPEEKKGFLDKVKEKDLARK------KTEEKRGFLKEKVPTHKKTEEFPFPAKPAYTEAAVSNTNTPLEPEEKRGLLDKIKEKMPG-HKKTEEV 405
PtrDHN-1.2   QTETPFEPEEK-GFLDKVKEKVPAHKTEEVPPPAEEKRGFLKEKVPTHKKTEEFPFPAKPAYTEAAVSNTNTPLEPEEKRGLLDKIKEKMPG-HKKTEEI 577
PtrDHN-10    ------------------------------------------------------------------------------------------ 108

PtrDHN-7     ---------------------------------------- 180
PtrDHN-6     D--------------------------------------- 189
PtrDHN-4     ---------------------------------------- 149
PtrDHN-5     ---------------------------------------- 205
PtrDHN-9     ---------------------------------------- 96
PtrDHN-2     ---------------------------------------- 183
PtrDHN-8     ---------------------------------------- 133
PtrDHN-3     --------------------------EKEKEKESASQR-- 226
PtrDHN-1.1   PPSEFDSTENVVSHKGEPMVKKGMMEKIKEKLPGHRPQI 444
PtrDHN-1.2   PPSEFDSIENVVSHKGEPMVKKGMMEKIKEKLPGHRPQI 616
PtrDHN-10    ---------------------------------------- 108
```

Fig. 4 Amino acid sequence alignment of all identified poplar DHNs. K-, Y-, and S-segments were respectively represented with *open blue*, *yellow*, and *red boxes*

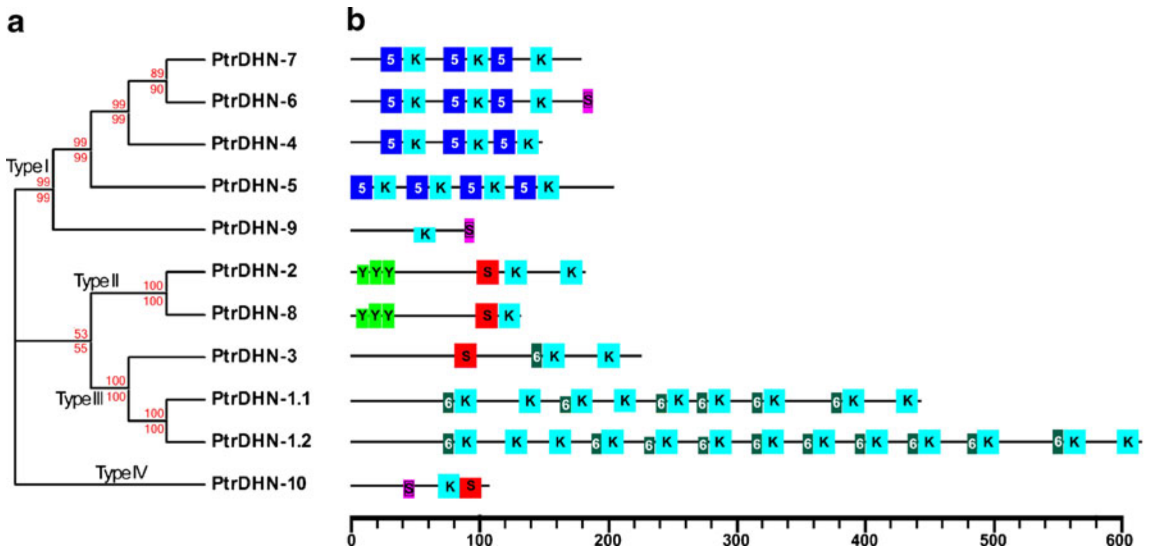


Fig. 5 Phylogenetic relationships and motif compositions of poplar DHN proteins. **a** Phylogenetic analysis of poplar DHN proteins. Neighbor-joining bootstrap and Minimum-Evolution values for clans supported above the 55% level were respectively indicated above and below the branches in red font. All poplar DHN protein names and their individual corresponding ID number for phylogenetic analysis are listed as Table 1. **b** Schematic view of the conserved motifs in the DHN proteins from *Populus* elucidated MEME (Bailey and Elkan 1994). Each motif is represented by a *capital or number in the colored box*, in which *K* represents K-segment, *S* and *S* represents S-segment, *Y* represents Y-segment, *5* represents motif-5, and *6* represents motif-6. The height of the motif "block" is proportional to −log (*p* value), truncated at the height for a motif with a *p* value of 1e−10. The *black lines* represent the non-conserved sequences. Refer to ESM Fig. S4 for the details of individual motif

that we constructed a reliable unrooted tree topology, in which the 11 poplar DHNs were grouped into four distinct clans, including Type I, Type II, Type III, and Type IV (Fig. 5a). The four distinct types generated by their evolutional divergence mostly corresponded to the subgroups identified by motif analysis above. The PtrDHN-2 and PtrDHN-8 belonging to Y_nSK_n subgroup were assigned to type II, and the SK_nS subgroup of PtrDHN-10 representing intermediate form of SK_n and K_nS to type IV. Type I contains two K_nS subgroup of DHNs (PtrDHN-6 and -9) and three K_n subgroup of DHNs (PtrDHN-7, -4, and -5; Fig. 5a). The latter differs from other two K_n subgroup DHNs of PtrDHN-1.1 (K_9) and -1.2 (K_{13}) by the presence of a novel repeating motif (motif-5; Fig. 5 and ESM Fig. S4e). The two K_n subgroup of DHNs (PtrDHN-1.1 and -1.2), together with one SK_n subgroup of DHN (PtrDHN-3), were assigned to type III because of their presence of the other one novel motif (motif-6; Fig. 5 and ESM Fig. S4f).

Table 2 Biochemical properties of all identified poplar DHN proteins

Novel simplified nomenclature	JGI ACS. number	Subclass	Number of AA	MW (kDa)	p*I*	GRAVY
PtrDHN-7	663123	K_3	180	19.12	6.45	−1.345
PtrDHN-6	571250	K_3S	189	20.10	9.08	−1.617
PtrDHN-4	571249	K_3	149	15.68	7.03	−1.313
PtrDHN-5	582807	K_4	205	20.75	9.96	−0.880
PtrDHN-9	665494[a]	KS	96	10.68	8.66	−1.995
PtrDHN-2	649369	Y_3SK_2	183	19.47	6.45	−1.099
PtrDHN-8	195568[a]	Y_3SK	133(175)	13.94	5.01	−0.968
PtrDHN-3	818850	SK_2	226	25.78	5.18	−1.632
PtrDHN-1.1	550802	K_9	444	50.81	8.98	−1.277
PtrDHN-1.2	NA	K_{13}	616	68.87	6.12	−1.225
PtrDHN-10	817405	SKS	108	11.28	9.08	−1.304

[a] Denotes that the genes and their encoding proteins had been revised in the present study

NA denotes not available

The similar conserved motifs of DHN proteins within the same types might provide additional supports for the unrooted tree topology. Also, proteins encoded by one paralogous pairs in *DHN* gene family well correspond to the same types, for instance, the paralogous pairs of PtrDHN-1/3 were assigned to type II, PtrDHN-2/8 to type III, Cluster I/PtrDHN-9 to type I. This evidence further supports the expansion of DHN gene family in the *Populus* genome caused by segmental duplication as well as tandem duplication events.

Biochemical properties of poplar DHN proteins

Generally, DHNs are characterized by the presence of abundant Gly and polar amino acid, but lack Cys and Trp (Close 1997, 1996). Analysis of the amino acid compositions of all poplar DHN proteins indicated that they share the common feature, only one exceptional example is PtrDHN-10 (817405) of the SKS subgroup with relatively high content of Cys (4.6%; ESM Table S7). Together with their relatively low GRAVY values in the range of −1.995 to approximately −0.880 (Table 2), confirm the presence of the very hydrophilic nature in *Populus* DHN proteins, which is in agreement with other plant DHNs (Kosova et al. 2007). For example, the K_n subgroup of PtrDHN-7 (K_3) with molecular mass of 19.1 kDa, Gly, Gln, Lys, Glu, and Asp represent 60.0% of the total amino acids, whereas no Cys and Trp were found (Table 2 and ESM Table S7). Calculation on MW of all poplar DHN proteins shows that they are characterized by a range of molecular masses from 10.7 to 68.9 kDa, most (9/11) of which are relatively small falling in a range of 10~26 kDa, only two are larger, respectively, being 50.8 and 68.9 kDa (Table 2). However, their unique amino acids composition led to the presence of discrepancy that the apparent MW on electrophoretic gels significantly higher than the actual MW of these proteins calculated from their amino acid sequence (Close 1997; Kosova et al. 2007). Like barley DHN5, its MW on SDS gels is evaluated into about 84 kDa according to standards of protein marker though its actual MW is only 58.5 kDa (Kosova et al. 2007). Accordingly, further experiment is required for confirming actual MW corresponding to apparent MW of each poplar DHNs.

In addition, isoelectric point (p*I*) value is also considered to be important biochemical properties for subdivision of DHNs of plants because DHNs of different acidic or basic features within the same subgroup might respond to various environmental factors (Allagulova et al. 2003). Theoretical p*I* values of *Populus* DHNs fluctuate in a wide range from 5.01~9.96, with five acidic DHNs, five basic and one neutral DHNs (Table 2), which is consistent with p*I* range (5.21~9.52) of barley DHNs (Kosova et al. 2007).

Tissue location of *DHN* gene expression in *Populus*

Numerous studies of *DHN*s have confirmed that they not only accumulated during seed desiccation and in response to water deficit induced by drought, low temperature, or salinity, but were also present in nearly all vegetative tissues during optimal growth conditions (Kosova et al. 2007; Rorat 2006; Tunnacliffe et al. 2010). To investigate all poplar *DHN* gene expression pattern of normal developmental tissues, we reanalyzed the poplar Affymetrix microarray data (Wilkins et al. 2009), using these matched probe sets to each poplar *DHN* (ESM Table S8). Similarly, poplar *DHN* genes expressed across nearly all vegetative tissues except for mature leaves (ML; Fig. 6). It is notable that the largest fraction of *DHN* genes preferentially expressed in continuous light-grown seedling (CL; 7/11), MC (6/11), FC (6/11), and young leaf (YL; 6/11). Relatively large fraction of *DHN* genes expressed in dark-grown seedlings (DS; 4/11), xylem (X; 4/11), etiolated dark-grown seedling transferred to light for 3 h (DL; 3/11), and root (R; 2/11).

Furthermore, several previous studies obtained from different species, indicated that different types of DHN proteins can localize to common tissues during development under normal growth conditions (Battaglia et al. 2008; Rorat 2006). Our in silico expression study of all poplar *DHN* genes confirms this conclusion, for example, the same tissue expression pattern are found between the K_nS type of

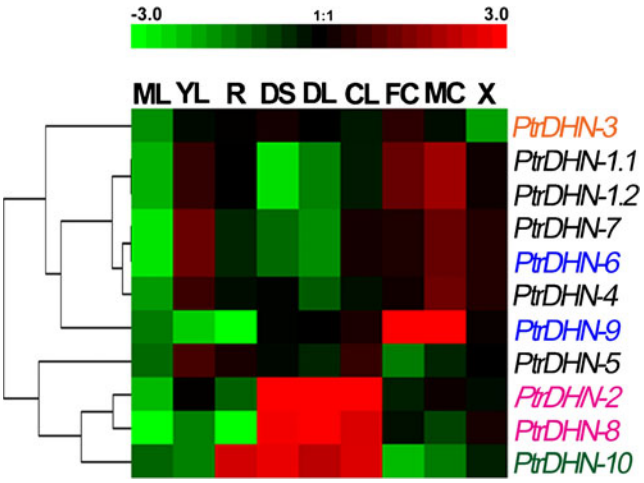


Fig. 6 Relative transcript abundance profiles of *Populus DHN* genes across different tissues. A heat map displaying the transcript abundance is produced using the genome-wide microarray data generated by Wilkins et al. (2009). The transcript abundance levels for the *Populus DHN* genes were clustered using hierarchical clustering based on Pearson correlation. *Color scale* at the bottom of each dendrogram represents log2 expression values, *green color* represents low level and *red color* represents high level of transcript abundances. *Populus DHN* within the same subgroup are marked in the common color. *ML* mature leaf; *YL* young leaf; *R* root; *DS* dark-grown seedlings; *DL* etiolated dark-grown seedling transferred to light for 3 h; *CL* continuous light-grown seedling; *FC* female catkins; *MC* male catkins; *X* xylem

PtrDHN-6 and the K_n type of *PtrDHN-4* and *-7*, as well as between the SK_nS type of *PtrDHN-10* and the Y_nSK_n type of *PtrDHN-2* and *-8* (Fig. 6). However, we also found that *Populus DHN* genes belonging to the same types preferentially expressed in the common tissues under normal growth conditions. For example, *PtrDHN-4*, *-7*, *-1.1*, and *-1.2* belonging to the same K_n type share the similar tissue expression patterns that preferentially expressed in MC, FC and YL, few accumulated in ML, R, CL, DL, and DS. *PtrDHN-2* (Y_3SK_2) and *-8* (Y_3SK) belonging to the common Y_nSK_n type share the same expression patterns with the highest transcript abundances especially present in seedlings under specific conditions (CL, DL, and DS), which is consistent pattern with this type of *DHN*s in other plants, such as Indian mustard *BjDHN1* (Y_3SK_2) and oilseed rape *BnDHN1* (Y_3SK_2; Yao et al. 2005). The evidence that poplar *DHN* genes within the same type preferentially share similar expression patterns across the nine tissues during normal growth conditions, would provide one useful data resource for exploring correlation between *DHN* type and their tissue localization.

Conclusion

Considerable research effort has been performed in characterization of the DHNs in herbaceous plants, such as barley, rice, and *Arabidopsis*, but such effort has not yet been directed towards woody trees. In this work, the above issues are addressed using the method of genome-wide identification and in silico analysis. This comprehensive analysis will be an important starting point for future efforts to elucidate the function role of all DHN proteins in poplar.

Acknowledgments We are grateful for the financial support from the National Basic Research Priorities Programme (Grant No. 2009CB119102), the Excellent Doctor Degree Dissertation in the Northeast Forestry University (Grant No. 140-602055), and Special Fund of Forestry Industrial Research for Public Welfare of China (Grant No. 201004040). The authors declare no conflicts of interest.

References

Abdel-Haleem H (2007) The origins of genome architecture. J Hered 98(6):633–634

Allagulova C, Gimalov F, Shakirova F, Vakhitov V (2003) The plant dehydrins: structure and putative functions. Biochemistry (Mosc) 68(9):945–951

Alsheikh MK, Svensson JT, Randall SK (2005) Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. Plant Cell Environ 28(9):1114–1122

Bae EK, Lee H, Lee JS, Noh EW (2009) Differential expression of a poplar SK2-type dehydrin gene in response to various stresses. BMB Rep 42:439–443

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. Menlo Park, California, AAAI Press, pp 28–36

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34(suppl 2):W369–W373

Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, DiLoreto DS, Yellanki P, Carlson JE (2009) The cinnamyl alcohol dehydrogenase gene family in Populus: phylogeny, organization, and expression. BMC Plant Biol 9(1):26

Battaglia M, Olvera-Carrillo Y, Garciarrubio A, Campos F, Covarrubias A (2008) The enigmatic LEA proteins and other hydrophilins. Plant Physiol 148(1):6–24

Caruso A, Morabito D, Delmotte F, Kahlem G, Carpin S (2002) Dehydrin induction during drought and osmotic stress in *Populus*. Plant Physiol Biochem 40(12):1033–1042

Choi DW, Zhu B, Close T (1999) 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(8):1234–1247

Close T (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol Plant 97(4):795–803

Close T (1997) Dehydrins: a commonalty in the response of plants to dehydration and low temperature. Physiol Plant 100(2):291–296

de Hoon M, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics (Oxford, England) 20 (9):1453–1454

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform 23(1):205–211

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res 20 (1):45–58

Finn R, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R (2006) Pfam: clans, web tools and services. Nucleic Acids Res 34 (Database Issue):D247–D251

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K (2010) The Pfam protein families database. Nucleic Acids Res 38(Database issue):D211–D222

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: vol 41. Nucleic Acids Symposium Series, pp 95–98

HongXia X, JunWei C, Ming X (2009) The role of dehydrin in plant response to cold stress. Acta Botanica Boreali-Occidentalia Sinica 29(1):199–206

Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G (2010) Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biol 10(1):145

Hundertmark M, Hincha DK (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics 9(1):118

Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. Nucleic Acids Res 32 (17):5096–5103

Kalluri UC, DiFazio SP, Brunner AM, Tuskan GA (2007) Genome-wide analysis of Aux/IAA and ARF gene families in *Populus trichocarpa*. BMC Plant Biol 7(1):59

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(4):1367–1377

Kosova K, Vitamvas P, Prásil I (2007) The role of dehydrins in plant response to cold. Biol Plant 51(4):601–617

Pruitt KD, Tatusova T, Maglott DR (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 33(suppl 1):D501–D504

Pruitt KD, Tatusova T, Maglott DR (2006) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35(suppl 1):D61–D65

Rinne PLH, Welling A, Schoot C (2010) Perennial life style of *Populus*: dormancy cycling and overwintering. Gen Genomics Populus 8(3):171–200

Rodriguez E, Svensson J, Malatrasi M, Choi DW, Close T (2005) Barley Dhn13 encodes a KS-type dehydrin with constitutive and stress responsive expression. Theor Appl Genet 110 (5):852–858

Rohde A, Ruttink T, Hostyn V, Sterck L, Van Driessche K, Boerjan W (2007) Gene expression during the induction, maintenance, and release of dormancy in apical buds of poplar. J Exp Bot 58(15–16):4047–4060

Rorat T (2006) Plant dehydrins-tissue location, structure and function. Cell Mol Biol Lett 11(4):536–556

Saitou N, Nei M (1987) The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Mol Biol Evol 4(4):406–425

Saldanha AJ (2004) Java Treeview-extensible visualization of microarray data. Bioinformatics 20(17):3246–3248

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28(10):2731–2739

Tunnacliffe A, Wise M (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften 94(10):791–812

Tunnacliffe A, Hincha D, Leprince O, Macherel D (2010) LEA proteins: versatility of form and function. Dormancy and Resistance in Harsh Environments: 91–108

Tuskan G, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313(5793):1596–1604

Wang BB, Brendel V (2006) Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci: 7175–7180

Wang XS, Zhu HB, Jin GL, Liu HL, Wu WR, Zhu J (2007) Genome-scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). Plant Sci 172(2):414–420

Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. Plant Physiol 149(2):981–993

Yao K, Lockhart KM, Kalanack JJ (2005) Cloning of dehydrin coding sequences from *Brassica juncea* and *Brassica napus* and their low temperature-inducible expression in germinating seeds. Plant Physiol Biochem 43(1):83–89