Plant Omics Journal POJ

POJ 6(6):415-424 (2013) ISSN:1836-3644

Identification and expression analysis of the heat shock transcription factor (*HSF*) gene family in *Populus trichocarpa*

Haizhen Zhang, Jingli Yang, Yunlin Chen, Xuliang Mao, Zhanchao Wang, Chenghao Li*

State Key Laboratory of Forest Genetics and Tree Breeding, Northeast Forestry University, 26 Hexing Road, Harbin 150040, China

***Corresponding author: chli0@163.com**

Abstract

In plants, heat shock transcription factors (HSFs) play key roles in regulating the heat shock (HS) response. Therefore, the genes encoding HSFs are important for adaptation to high temperatures and for tolerance to other abiotic stresses. There have been many studies on these genes in herbaceous plants, but few on the *HSF* family in woody plants. In this study, we identified 31 *HSF* genes in *Populus trichocarpa* and investigated their phylogenetic relationships with *HSF* genes in *Arabidopsis* and rice. Analyses of chromosomal duplications revealed that tandem/segmental duplications contributed to the expansion of the *HSF* gene family in *P. trichocarpa*. Gene structure was analyzed by investigating exon/intron organization and by using the MEME motif finder. Changes in gene expression were investigated using exPlot and digital northern analyses. Interestingly, nine of the *HSF* genes showed significant variations in expression patterns, suggesting that they have roles in stress responses. We evaluated changes in the transcript levels of the HSFs in response to abiotic stresses (heat, cadmium, salt, abscisic acid, and drought stresses) by analyzing $\log_2$ fold change data. The transcript levels of the *PoptrHSFA4a*, *PoptrHSFA4b* and *PoptrHSFA5a* genes markedly increased in response to a wide range of stresses. The results of this study provide further information for cloning and expression of *HSF* genes, and for functional studies on the roles of these genes during development and in the responses to various environmental stimuli. Our results may help researchers design more efficient strategies to study the *P. trichocarpa HSF* family.

Keywords: heat stress transcription factor; woody plant; bioinformatics analysis; gene expression; abiotic stress.
Abbreviations: BLAST_Basic Local Alignment Search Tool; DBD_DNA-binding domain; GSDS_Gene Structure Display Server; HS_heat shock; HSPs_heat shock proteins; HSFs_heat shock factors; KEGG_Kyoto Encyclopedia of Genes and Genomes; MEME_ Motif Elucidation; NCBI_National Center for Biotechnology Information.

Introduction

The protective mechanisms of plants allow them to survive under complex stress conditions. Much research has been conducted on these protective mechanisms, particularly the rapid cellular defense mechanism commonly known as the heat shock (HS) response, which is characterized by large increases in the abundance of heat shock proteins (HSPs) (Queitsch et al., 2000; Lee et al., 1994). These proteins function as molecular chaperones, and prevent protein unfolding and aggregation to help maintain cellular protein homeostasis. The HSPs are activated by heat shock stress transcription factors (HSFs), which bind to heat stress elements (5'-AGAAnnTTCT-3') in the promoters of the *HSP* genes (Pelham et al., 1982). HSFs are the terminal components of the signal transduction chain that activates genes in response to various inducers, including high temperatures, oxidants, heavy metals, and pathogens (Morimoto, 1998). HSFs have a modular structure with a highly conserved helix-turn-helix motif in the N-terminal region, an adjacent domain with heptad hydrophobic A/B repeats, which is involved in oligomerization, a nuclear localization signal, and a C-terminal activation domain (Döring et al., 2000). Based on the presence of the conserved DNA-binding domain (DBD) and the adjacent HR-A/B region, plant *HSF* genes can be classified into three classes: A, B, and C (Nover et al., 2001). Class A *HSF* genes have an AHA transactivator domain in their C-terminal region, while class B and C *HSF* genes lack this transactivator domain (Nover et al., 2001; Baniwal et al., 2004). The *HSF* gene family has been well characterized in *Arabidopsis*, rice, and tomato (Nover et al., 2001; Guo et al., 2008a). In *Arabidopsis*, there are 21 *HSF* genes assigned to three classes and 14 groups. The *HSF* genes A1a, A1b, A2, A4a, A4c, A5, A9, and B1 have been functionally characterized in detail (Guo et al., 2008b; Mishra et al., 2002; Li et al., 2005; Charng et al., 2007; Schramm et al., 2008; Shim et al., 2009; Liu et al., 2009). It was reported that expression of *OsHSFA2a* in particular increased in root and shoot tissues in response to heat stress in rice. The authors found that *OsHSFA3* expression was more sensitive to cold and drought stress, while *OsHSFA7* and *OsHSFA9* were specifically expressed in developing seeds (Chauhan et al., 2011). There is a large body of research on the roles of *HSF* genes in the responses to various abiotic stresses. For example, *HSFA1* showed a unique role as a master regulator of thermotolerance in tomato (Mishra et al., 2002). In *Arabidopsis*, *AtHSFA2* was shown to modulate the expression of stress responsive genes and enhance tolerance to heat and oxidative stress (Li et al., 2005), and *AtHSF3* was shown to depress the HS response (Prändl et al., 1998). *HSFA4a* conferred cadmium tolerance in wheat and rice (Shim et al., 2009), and the rice *OsHSF7* gene was shown to function as a high-temperature receptor and response factor

(Liu et al., 2009). Black cottonwood (*Populus trichocarpa*) is a perennial woody deciduous plant that is cultivated mainly in western North America. This species makes a substantial contribution to the ecology and economy of the regions in which it is grown. It was the first tree for which a complete genome sequence was available (Tuskan et al., 2006). The completion of the whole genome sequence for *P. trichocarpa* provided an opportunity to analyze the structure and function of *HSF* genes. Currently, 28 different *HSF* genes have been identified in *P. trichocarpa* by screening cDNA libraries in the EST database. These libraries represent data from bioinformatics studies and analyses of expression patterns in 11 different tissues and organs (Wang et al., 2012). To further confirm the responsiveness of *HSF* genes to abiotic stress, quantitative real-time PCR was used to analyze changes in the transcript levels of 28 *P. trichocarpa HSF* genes in roots, stems, and leaves of poplar under heat stress (Wang et al., 2012). However, to date, there have been no reports on the *HSF* gene family and expression patterns at the transcriptional level under other abiotic stresses in *P. trichocarpa*. In this study, we revealed three additional *HSF* genes in *Populus* to extend the total number of HSF genes to 31. As well, we used the *Arabidopsis* and rice *HSF* gene datasets reported by Guo et al. (Guo et al., 2008a) to classify the *P. trichocarpa* genes. We examined the phylogenetic relationships among these genes, determined their chromosomal locations, and analyzed their exon/intron structures and motifs. We used exPlot, digital northern, and quantitative real-time polymerase chain reaction (qRT-PCR) analyses to explore their transcription patterns. The aim of this work was to provide a foundation to explore the functions and structure of the *HSF* gene family in *P. trichocarpa* and to reveal their expression patterns in response to various abiotic stresses. These data will undoubtedly be useful in future gene cloning and functional studies.

Results

Identification and phylogenetic analysis of HSF genes

The characteristic features, protein lengths, molecular weights, isoelectric points and Ensembl IDs of the *P. trichocarpa HSF* genes are shown in Table 1. A phylogenetic tree was constructed from the full-length aligned protein sequences of three model plants (*P. trichocarpa*, rice, and *Arabidopsis*) (Fig. 1). In total, 30 full-length genes encoding putative *P. trichocarpa HSFs* were identified in the Heatster database. Using Pfam nomenclature, the *P. trichocarpa HSF* gene family was divided into three groups: *HSF* A, B, and C genes. Class A could be further divided into nine subtypes and class B into five different subtypes, while class C had only one type; thus, the *HSF* genes were designated as A1–A9, B1–B5, and C, respectively. Class A comprised 17 genes, class B comprised 12, and class C had only a single gene. However, 31 full-length *HSF* genes were identified from KEGG. According to the multiple sequence alignment, KEGG contained all 30 of the *HSF* genes identified in Heatster, but had no clear delineation of the remaining gene. Based on the multiple sequence alignment and phylogenetic analysis, this gene was defined as *PoptrHSFB5c*. In the phylogeny, all class A *HSF* genes were grouped in a single major clade with two distinct sub groups; (A1, A3, A4) and (A2, A6, A7, A8, A9). The evolutionary tree indicated that class B and class C genes were derived from class A ones.

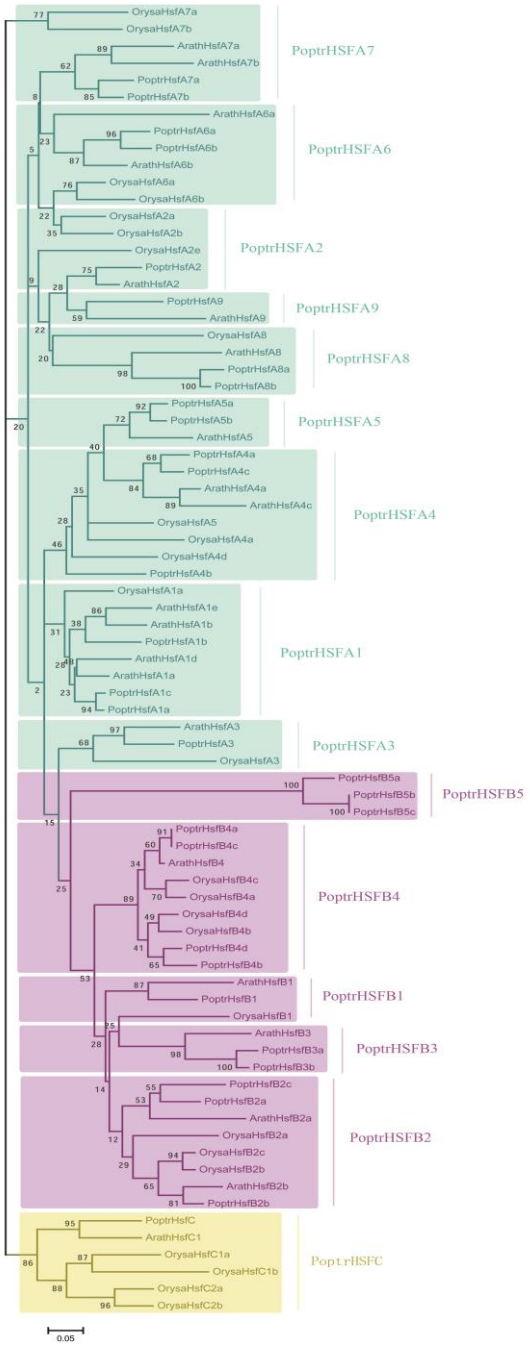


Fig 1. Phylogenetic analysis of HSF genes from P. trichocarpa, rice and Arabidopsis. The phylogenetic relationships are based on amino acid sequence comparisons of the conserved N-terminal halves, the DBD and HR-A/B regions and the linker between both regions. Three classes of HSF gene are distinguished by different color. Dark green, purple and yellow represent class A, B and C respectively, at the same time, each rectangle represents a subfamily.

Chromosomal location, gene structure, and conserved motifs of P. trichocarpa HSF genes

The 31 *P. trichocarpa HSF* genes were distributed across all linkage groups (LGs) except for LGs XIII and XIX (Fig. 2). The distribution of *HSF genes* in the *P. trichocarpa* genome was heterogeneous and gene clusters were observed.

Table 1. Classification and physicochemical properties of the *P. trichocarpa* HSF proteins.

Gene name	Class	Ensembl-ID	Length (aa)	MW(Da)	pI
PoptrHSFA1a	A1a	POPTRDRAFT_757199	507	55694.9	4.84
PoptrHSFA1b	A1b	POPTRDRAFT_1097130	499	55091.5	5.54
PoptrHSFA1c	A1c	POPTRDRAFT_813327	510	55849.1	4.82
PoptrHSFA2	A2	POPTRDRAFT_763234	388	43856.9	4.96
PoptrHSFA3	A3	POPTRDRAFT_282464	476	53261.4	4.86
PoptrHSFA4a	A4a	POPTRDRAFT_568665	406	46266.6	5.23
PoptrHSFA4b	A4b	POPTRDRAFT_1099462	443	50781.7	5.82
PoptrHSFA4c	A4c	POPTRDRAFT_758729	413	47388.9	5.19
PoptrHSFA5a	A5a	POPTRDRAFT_597074	485	54364.2	5.54
PoptrHSFA5b	A5b	POPTRDRAFT_641884	490	54702.6	5.77
PoptrHSFA6a	A6a	POPTRDRAFT_231057	358	41335.3	5.16
PoptrHSFA6b	A6b	POPTRDRAFT_766391	348	40084.0	5.15
PoptrHSFA7a	A7a	POPTRDRAFT_1079166	359	40692.4	5.50
PoptrHSFA7b	A7b	POPTRDRAFT_410842	330	38122.2	5.98
PoptrHSFA8a	A8a	POPTRDRAFT_820699	393	44808.0	4.80
PoptrHSFA8b	A8b	POPTRDRAFT_822047	392	44694.8	4.7
PoptrHSFA9	A9	POPTRDRAFT_416510	484	54007.8	5.45
PoptrHSFB1	B1	POPTRDRAFT_1083203	258	28183.4	8.71
PoptrHSFB2a	B2a	POPTRDRAFT_422013	301	33366.2	5.04
PoptrHSFB2b	B2b	POPTRDRAFT_548499	343	36826.6	4.97
PoptrHSFB2c	B2c	POPTRDRAFT_253824	286	31751.7	5.14
PoptrHSFB3a	B3a	POPTRDRAFT_717386	226	26377.0	8.75
PoptrHSFB3b	B3b	POPTRDRAFT_576364	228	26485.9	6.78
PoptrHSFB4a	B4a	POPTRDRAFT_412394	364	40440.6	8.15
PoptrHSFB4b	B4b	POPTRDRAFT_557710	272	31530.7	6.67
PoptrHSFB4c	B4c	POPTRDRAFT_1110420	368	41038.1	8.16
PoptrHSFB4d	B4d	POPTRDRAFT_641429	270	31292.3	6.63
PoptrHSFB5a	B5a	POPTRDRAFT_758907	209	24028.4	9.16
PoptrHSFB5b	B5b	POPTRDRAFT_784255	211	24383.6	9.35
PoptrHSFB5c	B5c	POPTRDRAFT_940628	132	15152.1	9.82
PoptrHSFC	C	POPTRDRAFT_261844	339	37981.9	5.43

Two adjacent *HSF* gene pairs were found within a distance of less than 200 kb on the duplication blocks, which may have resulted from tandem duplication, according to Leister's description (Leister, 2004). Six gene pairs were located in conserved positions on homologous gene blocks of different chromosomes, suggesting that these genes may have resulted from segmental duplication (Fig. 2). Two pairs of tandemly duplicated genes were arranged on LGVIII and LGXI. Four genes were found on LG VI, three on LGs I, VIII, and XI, two on LGs II, III, IV, X, XIV, and XVIII, and the remaining genes were distributed one per chromosome. To gain further insight into the structural diversity of the *P. trichocarpa HSF* genes, we constructed a separate phylogenetic tree using only the full-length HSF protein sequences of *P. trichocarpa* (Fig. 3A). The *P. trichocarpa HSF* genes were classified into three classes as described above. We then compared the exon/intron organization in the coding sequences of the genes. All of the *P. trichocarpa HSF* genes had two exons and one intron, except that *HSFA8a* lacked an intron. There were striking differences in the arrangement of introns and intron phases among the subfamilies of the *P. trichocarpa HSF* genes, but the intron phases were remarkably well-conserved within each subfamily (Fig. 3B). We predicted conserved motifs using MEME motif detection software to reveal the diversification of the *P. trichocarpa HSF* genes. The details of the 15 putative motifs are shown in Table 2. As shown in Fig. 3C, most of the closely related members in the phylogenetic tree shared common motifs and all *HSF* genes contained motifs 1, 2, and 4. This suggested that *HSF* genes of similar structure were assigned to the same subfamily.

Expression analyses of HSF genes

Expression profiling and functional analysis of 27 *HSF* genes from 17 different *Populus* samples was performed with exPlot. During the seed germination process in *Populus balsamifera* (GSM327380), *PoptrHSFB2a* showed the highest expression value, and members of classes A (*PoptrHSFA5a*, *PoptrHSFA4a* and *PoptrHSFA4c*) and C (*PoptrHSFC*) also showed high expression values. Class A (*PoptrHSFA4c*, *PoptrHSFA4a*) and class B (*PoptrHSFB4d*, *PoptrHSFB4a* and *PoptrHSFB4c*) genes were highly expressed throughout the stem tissue. Several genes were highly expressed in leaves infected with rust (*Melampsora medusae* f. sp. *tremuloidae*; mmt) (GSM244438, GSM244439), with *PoptrHSFB5b* showing the highest expression value. In another dataset for mmt-infected leaf discs (GSM412656-GSM412664), class A (*PoptrHSFA4c*, *PoptrHSFA4a*, *PoptrHSFA5a*, *PoptrHSFA3*, *PoptrHSFA1c*) and class C (*PoptrHSFB2a*) genes were highly expressed. In suspension cells treated with the phytotoxin thaxtomin a, class A (*PoptrHSFA1b*, *PoptrHSFA8b*, *PoptrHSFA4c*, *PoptrHSFA6b*) and class B (*PoptrHSFB5a*) genes showed higher expression levels than those of the other *HSF* genes. In mature leaves after 9 weeks of drought treatment (GSM529949), members of class A and class B were expressed at very high levels. In summary, the exPlot expression data showed that the class A genes *PoptrHSFA5a*, *PoptrHSFA4a* and *PoptrHSFA4c*, and the class B gene *PoptrHSFB2a* were expressed at high levels in all samples, suggesting that these *Populus HSF* genes participate in both abiotic and biotic stress responses (Fig. 4).

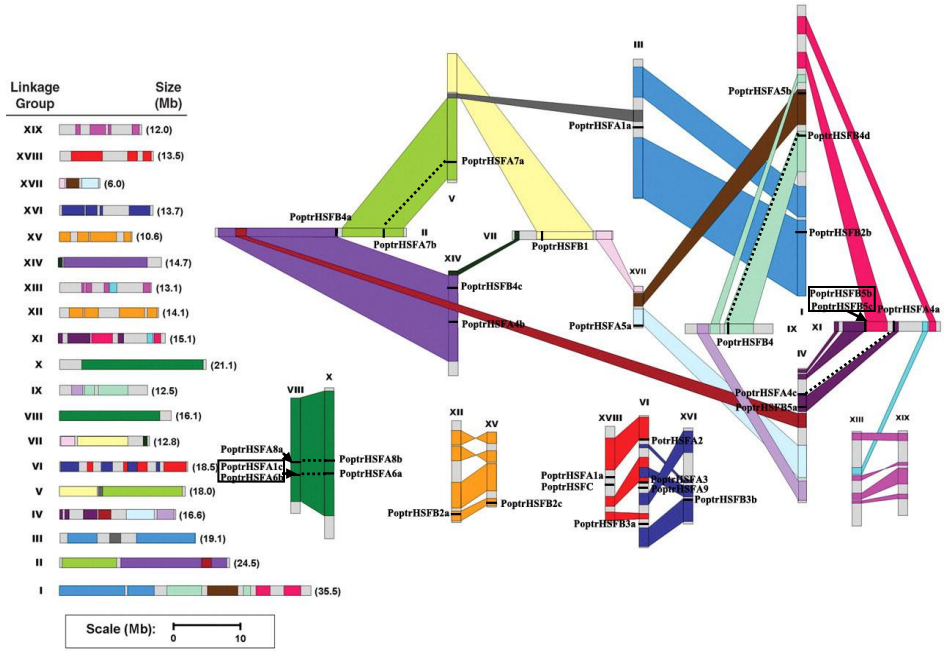


Fig 2. Locations of *P. trichocarpa HSF* genes on the chromosomes LGI–XIX. Common colors refer to homologous genome blocks and presumed to have arisen from the salicoid-specific genome duplication (Tuskan et al., 2006). Homologous gene blocks are indicated with the same color and segmental duplicated genes are connected with lines. Tandemly duplicated genes are encompassed in the black boxes. The scale represents mega bases (Mb). The LG numbers are indicated at the top of each bar.

Table 2. Motif sequences of *HSF* genes identified in *P. trichocarpa* by MEME tools.

Motif	Width	Best possible match
1	36	LNTYGFRKIDPDRWEFANECFRRGQKHLLCNIHRRK
2	50	YDMVDDPHTDHIVSWNRDGNSFVVWDPPEFARDLLPKYFKHNNFSSFVRQ
3	50	RDKNVLMMELVKLRQQQQTTDHCIQAMEQRLQGMECRQQQMMSFLAKAMQ
4	11	KNGPPPFLTKT
5	33	CTALMEENERLRKENCMLMSELTHMKKLCNDII
6	21	VNDGFWEQFLTENPGYYDIEE
7	21	AVGACVEVGRFGYWGEIERLK
8	29	FLAQLVQQKEMRWRLIEAMSKKRRRPIDQ
9	29	NKPTDHGNHWWKMQHMDNLTEQMGHLTPA
10	46	APDGQIVKYQPPMNEAADAMHAPIMKMEAPRRLEPYMTNWKDFFIG
11	28	HYLHLMKEEEGCKTKLFGVPLHVKKRRH
12	50	HLACKIEAMDFSAYSKKRRLPQVDHPMPIAENSFVENHCSSRPESNVIHQ
13	12	YFIQNHVKPVPP
14	46	CGPAPYATANHVTSNGSLVQKPLNQLLGYYPTTAPNNPKQIPQVHV
15	50	VGGQGAPAIPCPQADVDMPPKSPGIDMNPEPIADIPCEPYMPPETCAGTF

Note: Numbers correspond to the motifs described in Figure 3. Sequences obtained from the analysis of three groups of *Populus trichocarpa* HSF complete proteins with the MEME tools.

The expression profiles of 18 *PoptrHSF* and 15 *PoptrHSP* genes were further analyzed by comparing their digital expression profiles among 17 different EST libraries. The results are shown as a dendrogram, which clusters genes with similar functions together. Comparison of the different libraries revealed that *PoptrHSFA5a* (POPTR_0017s08630) was highly expressed in flower buds but expressed at lower levels in the other tissues. *PoptrHSFA8b* (POPTR_0010s11490) was expressed at high levels in female catkins, with weak expression in the bark and low expression in the other tissues. On the whole, the remaining genes were inactive in the different libraries examined (Fig. 5).

Expression of HSF genes under abiotic stress conditions

To further investigate the expression patterns of this gene family, qRT-PCR analysis was performed for nine selected *HSF* genes that showed drought-responsive expression in leaf tissues in the exPlot expression analysis. Two paralogous sets (*PoptrHSFA1b*/*PoptrHSFA1c* and *PoptrHSFA4a*/*PoptrHSFA4b*/*PoptrHSFA4c*) were included, to determine whether they showed similar trends in expression. The transcript levels of these genes were determined under different abiotic stresses by qRT-PCR. When tissue-cultured cells were subjected heat stress (42°C), the transcript level of *PoptrHSFA4a* was clearly increased at 0.5 h, to a level 8-fold higher than that in the untreated control, while *PoptrHSFB2a* and *PoptrHSFB4d* showed slight increases in their transcript levels. In contrast, none of the other genes showed increased transcript levels in response to heat stress; in fact, the transcript level of *PoptrHSFA5a* decreased (Fig. 6A). The transcript level of *PoptrHSFA4b* increased more than 10-fold after cadmium treatment (10 mM) for 10 h compared with that in the control, but the other genes did not show obvious changes in

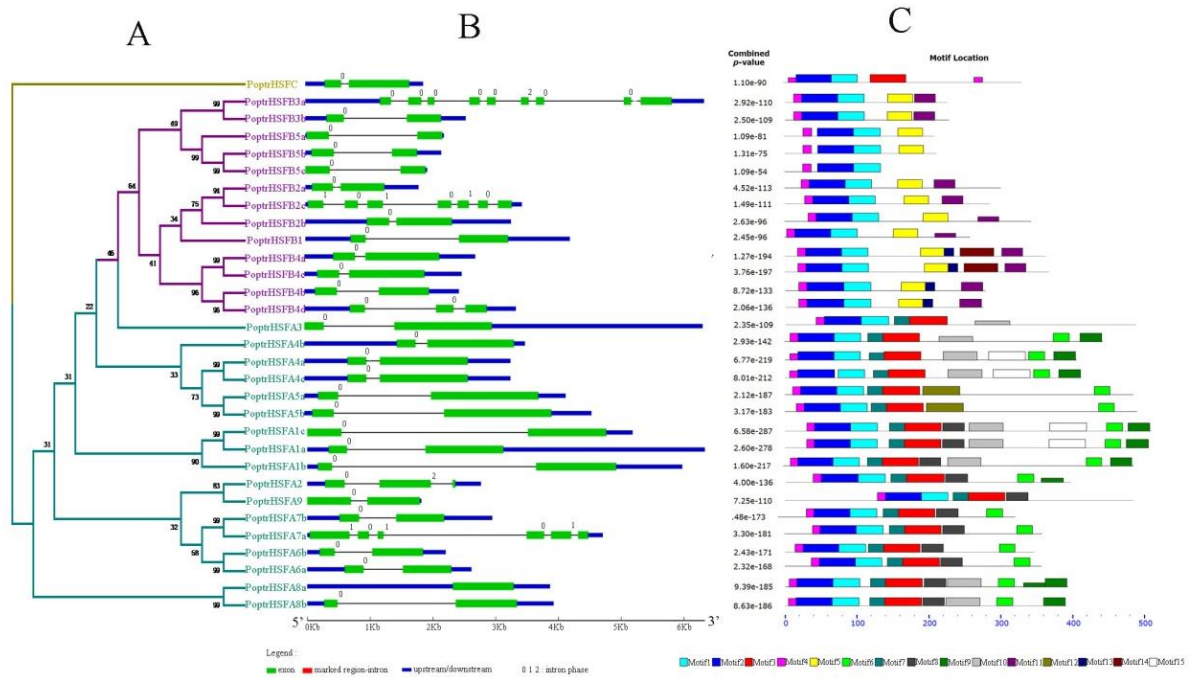
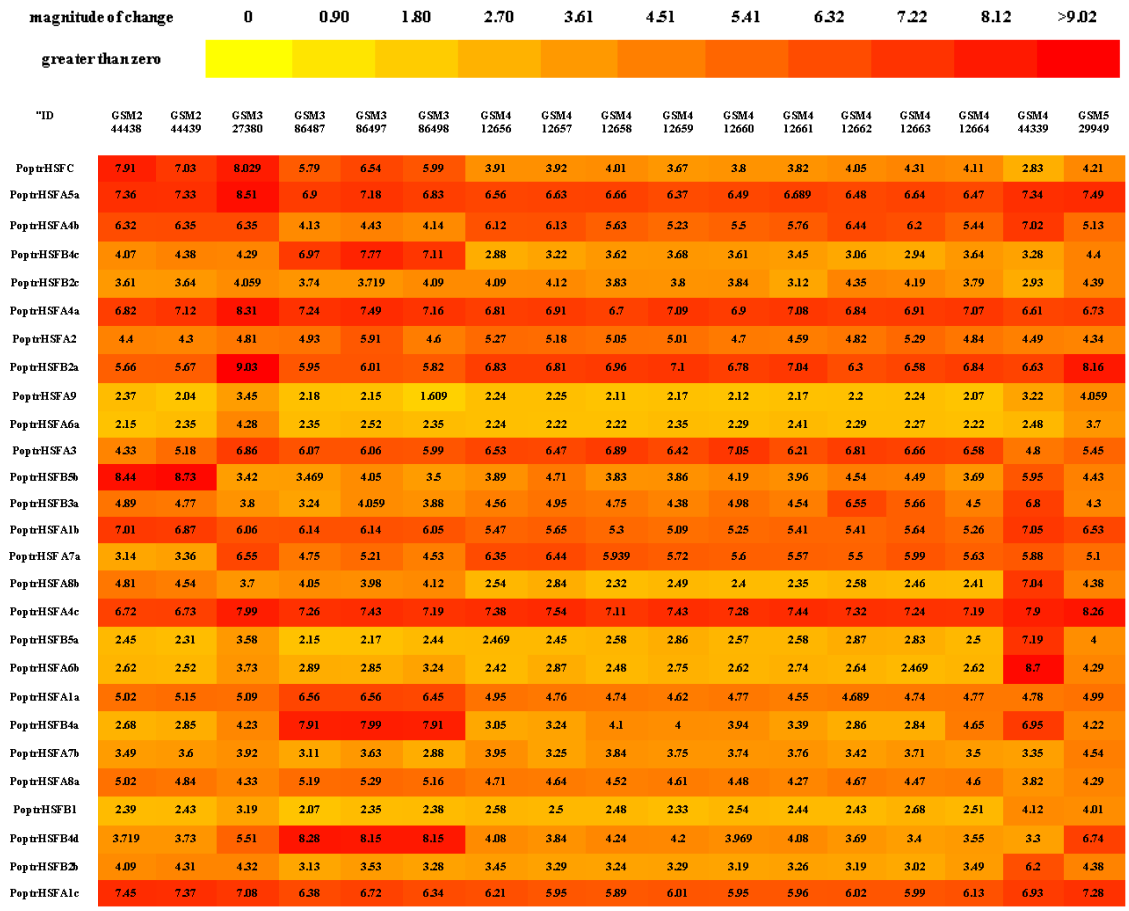


Fig 3. Phylogenomic analysis of 31 *HSF* genes in *P. trichocarpa* (A) with the integration of exon/intron structures (B) and MEME motifs (C). Exon-intron structure was obtained from the Gene Structure Display Server. Motifs were identified with the MEME software using the complete amino acid sequences of the *HSF* genes.

Table 3. The *HSF* genes identified from the *P. trichocarpa* genome.

Genes	NCBI ID	Map position (bp)	Peptide-Name	Gene model (V2.1)
PoptrHSFA1a	XM_002303393.1	LGIII:9463415-9466122	18216587_peptide	POPTR_0003s09370.1
PoptrHSFA1b	XM_002319751.1	LGXIII:7022204-7027082	18221247_peptide	POPTR_0013s07730.1
PoptrHSFA1c	XM_002326502.1	LGVIII:10692315-10693075	18237171_peptide	POPTR_0001s02140.1
PoptrHSFA2	XM_002309456.1	LGVI:15181094-15182694	18211418_peptide	POPTR_0006s24330.1
PoptrHSFA3	XM_002326210.1	LGVI:9089145-9090553	18212887_peptide	POPTR_0006s11680.1
PoptrHSFA4a	XM_002316773.1	LGXI:8481675-8483040	18231972_peptide	POPTR_0011s06820.1
PoptrHSFA4b	XM_002321033.1	LGXIV:6843226-6845304	18223403_peptide	POPTR_0014s13780.1
PoptrHSFA4c	XM_002305016.1	LGIV:3204764-3206146	18224442_peptide	POPTR_0004s06090.1
PoptrHSFA5a	XM_002328876.1	LGXVII:5476275-5478745	18210583_peptide	POPTR_0017s08630.1
PoptrHSFA5b	XM_002298639.1	LGI:24867108-24869860	18237079_peptide	POPTR_0001s32810.1
PoptrHSFA6a	XM_002315715.1	LGX:9112826-9114092	18241482_peptide	POPTR_0010s09210.1
PoptrHSFA6b	XM_002311606.1	LGVIII:10692120-10693570	18247875_peptide	POPTR_0008s15740.1
PoptrHSFA7a	XM_002306760.1	LGV:14856683-14858754	18205996_peptide	POPTR_0005s23640.1
PoptrHSFA7b	XM_002302055.1	LGII:3133642-3135170	18246773_peptide	POPTR_0002s04900.1
PoptrHSFA8a	XM_002311501.1	LGVIII:9024854-9027973	18248062_peptide	POPTR_0008s13620.1
PoptrHSFA8b	XM_002315811.1	LGX:10883415-10887046	18241529_peptide	POPTR_0010s11490.1
PoptrHSFA9	XM_002309178.1	LGVI:8449484-8450337	18211973_peptide	POPTR_0006s15050.1
PoptrHSFB1	XM_002310106.1	LGVII:8975331-8978142	18243268_peptide	POPTR_0007s11030.1
PoptrHSFB2a	XM_002318787.1	LGXII:12971230-12971977	18229765_peptide	POPTR_0012s13430.1
PoptrHSFB2b	XM_002298017.1	LGI:8039253-8041616	18234944_peptide	POPTR_0001s08990.1
PoptrHSFB2c	XM_002321866.1	LGXV:10141139-10141519	18232817_peptide	POPTR_0015s13390.1
PoptrHSFB3a	XM_002307987.1	LGVI:3262338-3264327	18213031_peptide	POPTR_0006s04770.1
PoptrHSFB3b	XM_002323284.1	LGXVI:3691176-3692925	18251345_peptide	POPTR_0016s05680.1
PoptrHSFB4a	XM_002301140.1	LGII:9443178-9444224	18244758_peptide	POPTR_0002s12640.1
PoptrHSFB4b	XM_002313973.1	LGIX:6173087-6174392	18228106_peptide	POPTR_0009s07220.1
PoptrHSFB4c	XM_002327119.1	LGXIV:2299827-2301609	18223097_peptide	POPTR_0014s02700.1
PoptrHSFB4d	XM_002298424.1	LGI:19901615-19902931	18234351_peptide	POPTR_0001s28040.1
PoptrHSFB5a	XM_002305104.1	LGIV:5014509-5016326	18225496_peptide	POPTR_0004s04260.1
PoptrHSFB5b	XM_002329783.1	LGXI:4422646-4424054	18230550_peptide	POPTR_0011s05130.1
PoptrHSFB5c	XM_002333987.1	LGXI:4422883-4424054		
PoptrHSFC	XM_002324382.1	LGXVIII:8165594-8166232	18214571_peptide	POPTR_0018s05770.1



"ID	GSM244438	GSM244439	GSM327380	GSM386487	GSM386497	GSM386498	GSM412656	GSM412657	GSM412658	GSM412659	GSM412660	GSM412661	GSM412662	GSM412663	GSM412664	GSM444339	GSM529949
PoptrHSFC	7.91	7.03	8.029	5.79	6.54	5.99	3.91	3.92	4.01	3.67	3.8	3.82	4.05	4.31	4.11	2.83	4.21
PoptrHSFA5a	7.36	7.33	8.51	6.9	7.18	6.83	6.56	6.63	6.66	6.37	6.49	6.689	6.48	6.64	6.47	7.34	7.49
PoptrHSFA4b	6.32	6.35	6.35	4.13	4.43	4.14	6.12	6.13	5.63	5.23	5.5	5.76	6.44	6.2	5.44	7.02	5.13
PoptrHSFB4c	4.07	4.38	4.29	6.97	7.77	7.11	2.88	3.22	3.62	3.68	3.61	3.45	3.06	2.94	3.64	3.28	4.4
PoptrHSFB2c	3.61	3.64	4.059	3.74	3.719	4.09	4.09	4.12	3.83	3.8	3.84	3.12	4.35	4.19	3.79	2.93	4.39
PoptrHSFA4a	6.82	7.12	8.31	7.24	7.49	7.16	6.81	6.91	6.7	7.09	6.9	7.08	6.84	6.91	7.07	6.61	6.73
PoptrHSFA2	4.4	4.3	4.81	4.93	5.91	4.6	5.27	5.18	5.05	5.01	4.7	4.59	4.82	5.29	4.84	4.49	4.34
PoptrHSFB2a	5.66	5.67	9.03	5.95	6.01	5.82	6.83	6.81	6.96	7.1	6.78	7.04	6.3	6.58	6.84	6.63	8.16
PoptrHSFA9	2.37	2.04	3.45	2.18	2.15	1.609	2.24	2.25	2.11	2.17	2.12	2.17	2.2	2.24	2.07	3.22	4.059
PoptrHSFA6a	2.15	2.35	4.28	2.35	2.52	2.35	2.24	2.22	2.22	2.35	2.29	2.41	2.29	2.27	2.22	2.48	3.7
PoptrHSFA3	4.33	5.18	6.86	6.07	6.06	5.99	6.53	6.47	6.89	6.42	7.05	6.21	6.81	6.66	6.58	4.8	5.45
PoptrHSFB5b	8.44	8.73	3.42	3.469	4.05	3.5	3.89	4.71	3.83	3.86	4.19	3.96	4.54	4.49	3.69	5.95	4.43
PoptrHSFB3a	4.89	4.77	3.8	3.24	4.059	3.88	4.56	4.95	4.75	4.38	4.98	4.54	6.55	5.66	4.5	6.8	4.3
PoptrHSFA1b	7.01	6.87	6.06	6.14	6.14	6.05	5.47	5.65	5.3	5.09	5.25	5.41	5.41	5.64	5.26	7.05	6.53
PoptrHSFA7a	3.14	3.36	6.55	4.75	5.21	4.53	6.35	6.44	5.939	5.72	5.6	5.57	5.5	5.99	5.63	5.88	5.1
PoptrHSFA8b	4.81	4.54	3.7	4.05	3.98	4.12	2.54	2.84	2.32	2.49	2.4	2.35	2.58	2.46	2.41	7.04	4.38
PoptrHSFA4c	6.72	6.73	7.99	7.26	7.43	7.19	7.38	7.54	7.11	7.43	7.28	7.44	7.32	7.24	7.19	7.9	8.26
PoptrHSFB5a	2.45	2.31	3.58	2.15	2.17	2.44	2.469	2.45	2.58	2.86	2.57	2.58	2.87	2.83	2.5	7.19	4
PoptrHSFA6b	2.62	2.52	3.73	2.89	2.85	3.24	2.42	2.87	2.48	2.75	2.62	2.74	2.64	2.469	2.62	8.7	4.29
PoptrHSFA1a	5.02	5.15	5.09	6.56	6.56	6.45	4.95	4.76	4.74	4.62	4.77	4.55	4.689	4.74	4.77	4.78	4.99
PoptrHSFB4a	2.68	2.85	4.23	7.91	7.99	7.91	3.05	3.24	4.1	4	3.94	3.39	2.86	2.84	4.65	6.95	4.22
PoptrHSFA7b	3.49	3.6	3.92	3.11	3.63	2.88	3.95	3.25	3.84	3.75	3.74	3.76	3.42	3.71	3.5	3.35	4.54
PoptrHSFA8a	5.02	4.84	4.33	5.19	5.29	5.16	4.71	4.64	4.52	4.61	4.48	4.27	4.67	4.47	4.6	3.82	4.29
PoptrHSFB1	2.39	2.43	3.19	2.07	2.35	2.38	2.58	2.5	2.48	2.33	2.54	2.44	2.43	2.68	2.51	4.12	4.01
PoptrHSFB4d	3.719	3.73	5.51	8.28	8.15	8.15	4.08	3.84	4.24	4.2	3.969	4.08	3.69	3.4	3.55	3.3	6.74
PoptrHSFB2b	4.09	4.31	4.32	3.13	3.53	3.28	3.45	3.29	3.24	3.29	3.19	3.26	3.19	3.02	3.49	6.2	4.38
PoptrHSFA1c	7.45	7.37	7.08	6.38	6.72	6.34	6.21	5.95	5.89	6.01	5.95	5.96	6.02	5.99	6.13	6.93	7.28

Fig 4. *Populus HSF* gene expression in response to different biotic and abiotic stress conditions. Data were obtained from the exPlot tool in PopulusDB. The subfamily names for each of the *Populus HSF* genes are presented in the left panel and the stress conditions are represented by GSM sample numbers. Yellow indicates low expression and red indicates high expression

their transcript levels (Fig. 6B). The transcript levels of the genes showed the same trend after a 1-h NaCl (100 mM) treatment. The transcript level of *PoptrHSFA4b* increased to approximately 6-fold higher than that in the control, but decreased as the duration of the stress treatment extended. There were no obvious changes in the transcript levels of the other genes under salt stress (Fig. 6C). Under abscisic acid stress (ABA, 500 μM) and osmotic stress (PEG 6000, 3%) treatments, the transcript levels of all of the genes increased. The transcript levels of *PoptrHSFA4b* and *PoptrHSFA5a* increased significantly after ABA treatment, especially at 6 h, when the value was 10–12-fold higher than that of the control (Fig 6D). All genes showed almost the same trend in transcription in the PEG 6000 treatment; their transcript levels did not significantly increase (Fig 6E). In summary, the transcript levels of *PoptrHSFA4a* and *PoptrHSFA4b* markedly changed in response to various stresses, especially those of *PoptrHSFA4b*, which markedly increased under Cd, NaCl, and ABA stress conditions.

Discussion

HSF genes of P. trichocarpa

Because *HSF* genes encode key transcription factors in the HS response, they have been extensively analyzed in the model plants *Arabidopsis* and rice. In total, 21 *Arabidopsis* and 25 rice *HSF* genes have been identified and analyzed (Wu, 1995). A previous study identified 28 *HSF* genes in *P. trichocarpa* (Wang et al., 2012). The results of our study expand the number of known *HSF* genes in *P. trichocarpa* to 31. We have also provided the accession number of each site for further analyses. The expression patterns of some of the 31 *HSF* genes under different stress conditions were already known, and were further confirmed by the Plant Transcription Factor Database (PlnTFDB 2.0). Phylogenetically, these genes can be divided into three classes: A, B and C, which can be further divided into 15 subfamilies. We defined and analyzed their genetic structure and conserved domains (motifs), chromosome locations and expression patterns, and performed qRT-PCR analyses to clarify their transcription patterns in response to stresses.

Chromosomal location and gene duplication of P. trichocarpa HSF genes

Although the *Populus* genome is as large as that of *Arabidopsis*, it contains more *HSF* genes than do other genomes. Gene duplication events play an important role in genomic rearrangement and expansion (Vision et al., 2000). Gene duplication can be divided into tandem and segmental duplication events. Our analyses indicated that in *P. trichocarpa*, two tandem and six segmental duplication events have occurred in the chromosomes (Fig. 2). These

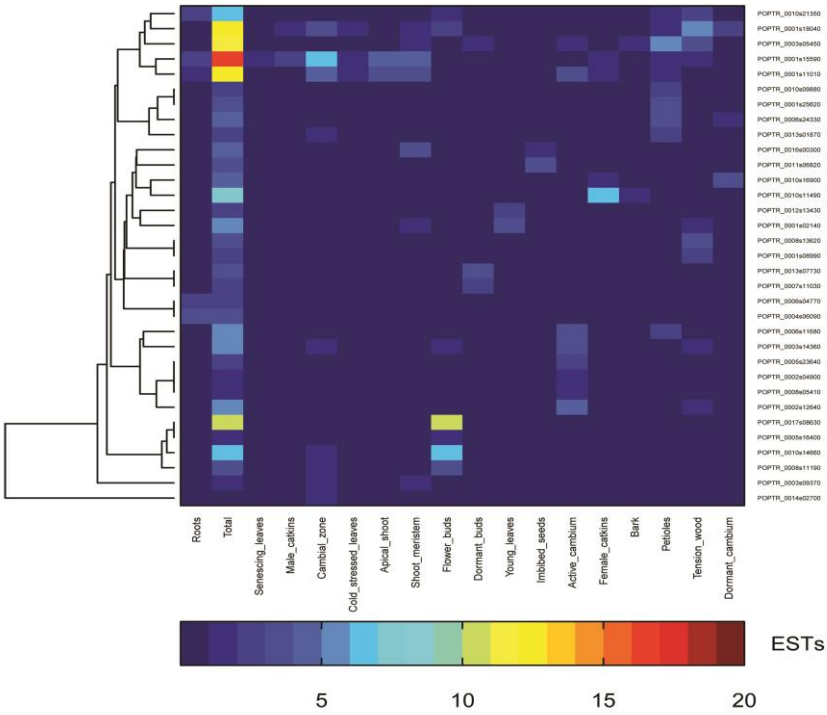


Fig 5. Expression patterns of a cluster of *Populus HSF* and *HSP* genes in a wide variety of tissues and organizations. The color bar represents the expression values: indigo represents low level expression, yellow shows medium level expression and crimson signifies high level expression. Gene names are shown on the right.

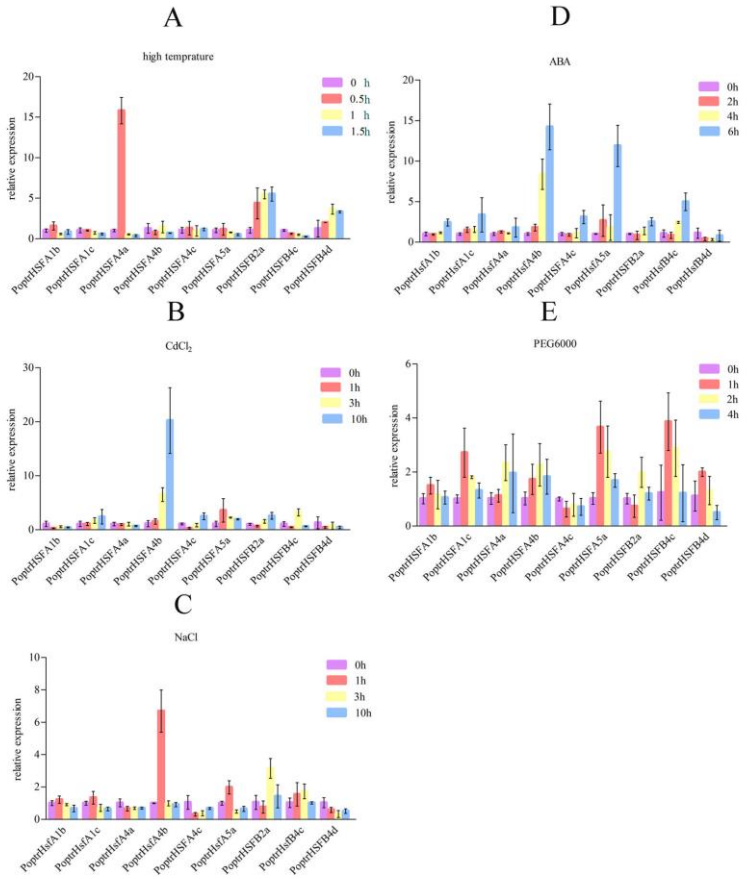


Fig 6. Expression analysis of nine selected *P. trichocarpa HSF* genes in mature leaves under high temperature (A), $CdCl_2$ (B), salinity (C), ABA (D) and drought stresses (E) by qRT-PCR. The data were normalized using the *P. trichocarpa* actin gene. Standard deviations were derived from three replicates of each experiment.

tandem and segmental duplications have probably played a major role in expanding the *HSF* gene family in *P. trichocarpa*. This has also been observed in other *Populus* gene families (Lan et al., 2009; Wilkins et al., 2009).

Gene structure and conserved motifs of P. trichocarpa HSF genes

Exon/intron structure and motif divergence play pivotal roles in the evolution of multiple gene families. Our results showed that genes in the same family generally showed similar conserved motifs and exon/intron structures (Fig. 3), e.g., all *HSF* genes had motifs 1, 2, and 4 when searched against the Pfam (http://pfam.sanger.ac.uk/search) and SMART (http://smart.embl-heidelberg.de) databases. These motifs comprise the DBD. As expected, we can determine putative functional domains from similar predicted motifs. At the same time, the structures of different genes from the same family have diverged during evolution.

Expression profile of Populus HSF genes and qRT-PCR analyses of P. trichocarpa HSF gene transcription

In *Arabidopsis* and rice, *HSF* genes are differentially expressed in a wide variety of tissues and organs (Swindell et al., 2007; Mittal et al., 2009). In our study, digital northern analysis showed that *PoptrHSF* and *PoptrHSP* genes are transcribed at various levels in a tissue-specific manner under unstressed conditions. This is similar to the *HSF* expression patterns of rice, in which some *HSF* genes are expressed only in the panicle and developing seed (Chauhan et al., 2011). Our analyses revealed that in *P. trichocarpa*, *PoptrHSFA5a* (POPTR_0017s08630) and *PoptrHSFA8b* (POPTR_0010s11490) are transcribed at higher levels in flower buds and female catkins than in other tissues. These genes may have a particular role in flower differentiation and development; for example, they may be associated with floral organ formation and/or the regulation of flowering time. Our heat map data complemented these results and suggested important independent clues as to the links between *HSF* and *HSP* genes. Examination of gene expression patterns under biotic and abiotic stresses in exPlot suggested that class A *HSF* genes (*PoptrHSFA1c*, *PoptrHSFA3*, *PoptrHSFA4a*, *PoptrHSFA4b*, *PoptrHSFA4c*, *PoptrHSFA5a*, *PoptrHSFB2a*) are expressed at high levels under all stresses. *PoptrHSFB2a*, which belongs to the class B *HSF* genes, is also actively expressed in response to all stresses. Our results indicated that these genes play a vital role during external stress to enhance or maintain normal growth in *Populus*. To verify these conclusions, we used qRT-PCR analyses to examine transcript levels of nine genes that showed drought-responsive expression in the exPlot analysis. In simulated drought conditions (osmotic stress), there were increased transcript levels of eight out of the nine genes in mature leaves; only *PoptrHSFA4c* showed decreased transcript levels under these conditions. The reason for this inconsistency may be that our stress treatment was not long enough or that other genes with high homology to *PoptrHSFA4c* replaced its function. Among the other four stress conditions, the increased transcript level of *PoptrHSFA4a* observed under high temperatures was similar to that of *HSFA1a*, a master regulator of induced thermotolerance in tomato, and that of *HSFA2a*, which shows exceptionally high expression in rice under heat stress. The highest transcript levels of *PoptrHSFA4b* were observed in response to Cd, NaCl, and ABA stress treatments; this transcription pattern is similar to that of *HSFA4a*, which confers cadmium tolerance in wheat and rice (Shim et al., 2009). The results of this study also showed that genes showing the highest degree of homology in the phylogenetic tree showed similar patterns of expression.

Materials and Methods

Database searches for HSF genes and analyses of physicochemical characteristics

The latest nucleotide and protein sequence data for *HSF* genes in *P. trichocarpa* were downloaded from KEGG (http://www.genome.jp/kegg/catalog/org_list.html) and Heatster (http://www.cibiv.at/services/hsf/) (Scharf et al., 2012). All protein sequences were selected according to the existence of an *HSF* domain by the SMART program. Theoretical isoelectric points (pI) and molecular weights were determined using the ProtParam Tool (http://web.expasy.org/protparam/).

Phylogenetic analysis

We used the *Arabidopsis* and rice *HSF* gene dataset reported by Guo et al. (2008a) to classify the *P. trichocarpa* genes. Classifications were determined by multiple sequence alignment of the HR–A/B region using protein sequences. The 77 amino acid sequences encoded by *HSF* genes were aligned using Clustal W (Larkin et al., 2007) and manually edited using Jalview to reduce gaps (Clamp et al., 2004). Neighbor-joining (NJ) trees (Saitou et al., 1987) were generated using the MEGA5 program with p-distances and the complete deletion option. Nodal support was estimated by bootstrap analysis and an interior branch test on the basis of 1000 re-samplings.

Chromosome localization

The *HSF* genes were plotted onto the *Populus* chromosomes by identifying their chromosomal position from the Phytozome (http://www.phytozome.net/) and Joint Genome Institute (http://genome.jgi-psf.org/pages/blast.jsf?db=Poptr1_1) websites (Table 3). A schematic view of the chromosomes was re-organized based on the most recent whole-genome duplication in *Populus* (Tuskan et al., 2006).

Exon/intron structure and motif analysis

Diagrams of exon/intron structure were obtained from the Gene Structure Display Server (GSDS) (http://gsds.cbi.pku.edu.cn/chinese.php). The CDS and genome sequences of the *P. trichocarpa* genes were obtained from NCBI (http://www.ncbi.nlm.nih.gov/). We found similar motifs among the full-length *HSF* amino acid sequences using the Motif Elucidation (MEME) system (http://meme.sdsc.edu/meme/cgi-bin/meme.cgi), with the following parameters: maximum number of motifs: 15, distribution of motif occurrences: zero or one per sequence, optimum motif width set to ≥ 6 and ≤ 50, and optimum number of sites for each motif, ≥ 2 and ≤ 600. All *P. trichocarpa* amino acid sequences used were downloaded from Heatster (http://www.cibiv.at/services/hsf/) (Scharf et al., 2012).

exPlot and digital northern analysis

We searched the *Populus HSF* gene expression data available at PopGenIE (http://popgenie.org/gp). A detailed description of samples was downloaded from Sample List

(http://popgenie.org/content/experiment-search) (Table S1). For digital northern analysis, digital northern heat maps were produced from EST libraries representing gene models within the PopulusDB (Segerman et al., 2007, Sterky et al., 2004). These included 17 libraries that were derived from several taxa of *Populus*; aspen (*Populus tremula*), a hybrid aspen (*P. tremula* × *tremuloides* T89), and black cottonwood (http://popgenie.org/book/digital-northern).

Stress treatments and qRT-PCR analysis

Plants were exposed to 42°C for 0, 0.5, 1, and 1.5 h; 100 mM NaCl for 0, 1, 3, and 10 h; 10 mM $CdCl_2$ for 0, 1, 3, and 10 h; 500 μM ABA for 0, 2, 4, and 6 h and 3% PEG 6000 for 0, 1, 2, and 4 h. Leaves were collected and immediately frozen in liquid nitrogen and then stored at −80°C for RNA extractions. Total RNA was extracted and its quality determined by electrophoresis on 1.2% agarose gels. qRT-PCR was performed using an ExTaq RT PCR kit and SYBR green dye (Takara, Dalian, China) in 96-well optical reaction plates (Applied Biosystems, Foster City, CA, USA). The transcript levels obtained for the different stages were standardized to that of actin using the $2^{-\Delta\Delta CT}$ method. We selected nine *HSF* genes (*PoptrHSFA1b*, *PoptrHSFA1c*, *PoptrHSFA4a*, *PoptrHSFA4b*, *PoptrHSFA4c*, *PoptrHSFA5a*, *PoptrHSFB2a*, *PoptrHSFB4c* and *PoptrHSFB4d*) showing drought-responsive expression in leaf tissue under drought stress from the exPlot expression analysis. We designed the primers using Primer Premier 5 to produce amplicons with lengths of 180–200 bp (Table S2). Amplifications were performed in 20-μl reaction mixtures containing 10 μl 2×SYBR Premix, 1.6 μl cDNA template, and 0.8 μl each specific primer to a final concentration of 0.4 μM. The reactions were performed under the following conditions: an initial denaturation step of 95°C for 30 s, followed by a two-step thermal cycling profile of denaturation at 95°C for 5 s and combined primer annealing/extension at 60°C for 30 s for 40 cycles. No-template controls were included for each primer pair and each PCR reaction was performed in triplicate.

Conclusions

The plant *HSF* family is an important target for study to elucidate the mechanisms of a variety of stress responses. Our study represents a comprehensive and specific analysis of gene structure, chromosome localization and expression of the *P. trichocarpa HSF* gene family. We demonstrated that *PoptrHSFA4a*, *PoptrHSFA4b* and *PoptrHSFA5a* play important roles in resistance to external biotic and abiotic stresses among all 31 *HSF* genes. We also predicted *P. trichocarpa HSF* gene expression and function through similar genes that have been well-studied in model or other plants. This information not only provides evidence for the role of *HSF* genes in the abiotic stress-response pathway, but may also be used to produce stress-tolerant *P. trichocarpa* cultivars suitable for different stress environments.

Acknowledgments

This research was supported by the National High Technology Research and Development Program of China (2011AA100203), Key Project of Chinese National Programs for Fundamental Research and Development (973 Program: 2009CB119100) and Fundamental Research Funds for the Central Universities (DL11EA02).

References

Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, Port M, Scharf KD (2004) Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. J Biosciences. 29: 471–487.

Charng YY, Liu HC, Liu NY, Chi WT, Wang CN, Chang SH, Wang TT (2007) A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in *Arabidopsis*. Plant Physiol. 143: 251–262.

Chauhan H, Khurana N, Agarwal P, Khurana P (2011) Heat shock factors in rice (*Oryza sativa* L.): genome-wide expression analysis during reproductive development and abiotic stress. Mol Genet Genomics. 286: 171–187.

Clamp M, Cuff J, Searle SM, Barton GJ (2004) The jalview java alignment editor. Bioinformatics. 20: 426–427.

Döring P, Treuter E, Kistner C, Lyck R, Chen A, Nover L (2000) The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. Plant Cell. 12: 265–278.

Guo J, Wu J, Ji Q, Wang C, Luo L, Yuan Y, Wang Y, Wang J (2008a) Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. J Genet Genomics. 35: 105–118.

Guo L, Chen S, Liu K, Liu Y, Ni L, Zhang K, Zhang L (2008b) Isolation of heat shock factor HsfA1a-binding sites in vivo revealed variations of heat shock elements in *Arabidopsis thaliana*. Plant Cell Physiol. 49: 1306–1315.

Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet. 20: 116–122.

Lan T, Yang ZL, Yang X, Liu YJ, Wang XR, Zeng QY (2009) Extensive functional diversification of the *Populus* glutathione s-transferase supergene family. Plant Cell. 21: 3749–3766.

Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R (2007) Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947–2948.

Lee Y, Nagao RT, Key JL (1994) A soybean 101-kD heat shock protein complements a yeast HSP104 deletion mutant in acquiring thermotolerance. Plant Cell. 6: 1889–1897.

Li C, Chen Q, Gao X, Qi B, Chen N, Xu S, Chen J, Wang X (2005) AtHsfA2 modulates expression of stress responsive genes and enhances tolerance to heat and oxidative stress in *Arabidopsis*. Sci China Ser C. 48: 540–550.

Liu JG, Qin Q, Zhang Z, Peng RH, Xiong A, Chen JM, Yao QH (2009) *OsHSF7* gene in rice, *Oryza sativa L.*, encodes a transcription factor that functions as a high temperature receptive and responsive factor. BMB Rep. 42: 16–21.

Mishra SK, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, Scharf KD (2002) In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. Genes Dev. 16: 1555–1567.

Mittal D, Chakrabarti S, Sarkar A, Singh A, Grover A (2009) Heat shock factor gene family in rice: genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. Plant Physiol Bioch. 47: 785–795.

Morimoto RI (1998) Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev. 12: 3788–3796.

Nover L, Bharti K, Döring P, Mishra SK, Ganguli A, Scharf KD (2001) *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperone. 6: 177–189.

Pelham H, Bienz M (1982) A synthetic heat-shock promoter element confers heat-inducibility on the herpes simplex virus thymidine kinase gene. EMBO J. 1: 1473–1477.

Prändl R, Hinderhofer K, Eggers-Schumacher G, Schöffl F (1998) HSF3, a new heat shock factor from *Arabidopsis thaliana*, derepresses the heat shock response and confers thermotolerance when overexpressed in transgenic plants. Mol Gen Genet. 258: 269–278.

Queitsch C, Hong SW, Vierling E, Lindquist S (2000) Heat shock protein 101 plays a crucial role in thermotolerance in *Arabidopsis*. Plant Cell. 12: 479–492.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4: 406–425.

Segerman B, Jansson S, Karlsson J (2007) Characterization of genes with tissue-specific differential expression patterns in *Populus*. Tree Genet Genomes. 3: 351–362.

Scharf KD, Berberich T, Ebersberger I, Nover L (2012) The plant heat stress transcription factor (Hsf) family: structure, function and evolution. BBA-Gene Regul Mech. 1819: 104–119.

Schramm F, Larkindale J, Kiehlmann E, Ganguli A, Englich G, Vierling E, Koskull Döring V (2008) A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of *Arabidopsis*. Plant J. 53: 264–274.

Shim D, Hwang JU, Lee J, Lee S, Choi Y, An G, Martinoia E, Lee Y (2009) Orthologs of the class A4 heat shock transcription factor HsfA4a confer cadmium tolerance in wheat and rice. Plant Cell. 21: 4031–4043.

Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandre K, Strauss SH (2004) A *Populus* EST resource for plant functional genomics. Proc Natl Acad Sci USA. 101: 13951–13956.

Swindell WR, Huebner M, Weber AP (2007) Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC genomics. 8: 125.

Tuskan GA, Di Fazio 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 G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, de Pamphilis 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, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé 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, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui H, 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.

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science. 290: 2114–2117.

Wang F, Dong Q, Jiang H, Zhu S, Chen B, Xiang Y (2012) Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. Mol Biol Rep. 39: 1877–1886.

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: 981–993.

Wu C (1995) Heat shock transcription factors: structure and regulation. Annu Rev Cell Dev Bl. 11: 441–469.