BIOLOGIA PLANTARUM 56 (4): 648-656, 2012 DOI: 10.1007/s10535-012-0239-3

Characterization of eight cytosolic *sHSP* genes and their expression in *Capsella bursa-pastoris*

P. TAO, L. LIU and J.B. WANG*

Key Laboratory of the MOE for Plant Developmental Biology, College of Life Sciences, Wuhan University, Wuhan 430072, P.R. China

Abstract

Small heat shock proteins (sHSPs) are crucial components of the plant response to heat shock. We identified and analyzed eight *sHSP* genes of *Capsella bursa-pastoris* to better understand the ability of this species to adapt. Eight genes were initially cloned and sequenced from the mature embryo cDNA pool. They belong to the cytosolic I (CI), cytosolic II (CII), and cytosolic III (CIII) subfamilies. One CI *sHSP* gene was homologous to that of *C. rubella*. Sequence analysis using 3'RACE revealed that there are two or more variable 3'-untranslated regions (UTRs) in these sHSP transcripts. The transcriptional levels of the eight *sHSP* genes were analyzed in different organs and developmental stages *via* qRT-PCR. Eight genes were significantly up-regulated in young leaves exposed to heat stress at 42 °C, and also showed differential responses to ABA treatment. We also compared expression of these genes with corresponding *Arabidopsis sHSP* genes and found some differences between the two species.

Additional key words: *Arabidopsis thaliana*, *Capsella rubella*, heat shock, qRT-PCR, RACE, 3'-UTR.

Introduction

Small heat shock proteins (sHSPs), which are ubiquitous in plants, play diverse roles in various developmental processes, including embryogenesis, seed germination, pollen development, and fruit maturation, and in the cellular response to environmental stresses and diseases (Wang *et al*. 2004, Sun and MacRae 2005, Kalemba and Pukacka 2008, Xue *et al*. 2010). They allow plants to quickly adapt to fluctuations in temperature, irradiance and to oxidative stress (Sun *et al*. 2002, Sundby *et al*. 2005, Dafny-Yelin *et al*. 2008). At least seven different plant sHSP subfamilies have been identified based on genome analysis of *Arabidopsis thaliana* (Scharf *et al*. 2001). The largest *sHSP* subfamily in *A. thaliana* consists of six cytosolic I genes (Siddique *et al*. 2008, Waters *et al*. 2008). Recent genome analysis of *A. thaliana*, *Oryza sativa*, and *Populus trichocarpa* revealed one additional mitochondrial and three additional cytosolic classes of *sHSPs* (Waters *et al*. 2008). Some novel properties have been characterized in these sHSPs (Siddique *et al*. 2008, Nautiyal and Shono 2010).

Much of our knowledge of the structure of sHSPs is based on crystal structures analysis of archaeabacterium (*Methanococcus jannaschii*) HSP16.5 (Kim *et al.* 1998) and wheat (*Triticum aestivum*) HSP16.9 (Van Montfort *et al.* 2001) and their comparisons revealed that most structural features are conserved (Van Montfort *et al*. 2002).

As a small genus within the mustard family, *Capsella* contains only three species: two diploids, *Capsella rubella* (2n=2x=16) and *Capsella grandiflora* (2n=2x=16), and one tetraploid *Capsella bursa-pastoris* (2n=4x=32) (Hurka and Neuffer 1997). *C. bursa-pastoris* is a good model plant for analysis of adaptation and evolution (Slotte *et al.* 2006, 2007). It is one of the widespread flowering plants, and it is efficiently adapted to different environments (Hurka *et al*. 2003). Most research on sHSPs has focused on *A. thaliana*. Little is known about cytosolic *sHSP* genes in *C. bursa-pastoris*, and several questions remain to be solved. For example, do the sequences of the *C. bursa-pastoris sHSP* genes resemble those of *Arabidopsis*? As a tetraploid plant, do *sHSP* families expand in *C. bursa-pastoris*? Are there new characteristics

Received 26 September 2011, *accepted* 30 October 2011.

Abbreviations: sHSP - small heat shock protein; qRT-PCR - quantitative reverse transcriptase - polymerase chain reaction; RACE - rapid amplification of cDNA ends; 3'-UTR - 3'-untranslated region.

Acknowledgments: This work was supported by the State Key Basic Research and Development Plan of China (2007CB108700), the National Natural Science Foundation of China (31070204, 30570112) and the Chinese 111 Project.

* Corresponding author; fax: (+86) 27 68752213, e-mail: jbwang@whu.edu.cn

in the 3'-UTRs of sHSP mRNAs? Are the expression patterns of the newly identified *sHSP* genes similar to those of the corresponding *sHSP* genes of *A. thaliana*? The 3'-UTR is rich in single-feature polymorphisms that can be used to distinguish between closely related transcripts (Bi *et al*. 2006, Eveland *et al*. 2008). In this study, we used 3'RACE to analyze the coding regions and information-rich 3'-UTRs of these *sHSP* genes.

Materials and methods

Plant materials, growth conditions and stress treatments: *Capsella bursa-pastoris* (L.) Medik. embryos were manually separated from ovules and devided into six groups according to developmental stages: globular-heart, torpedo, cotyledon, premature, mature and embryos 4 months after ripening. In addition, the different organs were prepared for expression analysis using qRT-PCR. Further, 4-week-old *C. bursa-pastoris* seedlings were grown in pots filled with a mixture of soil and sand (2:1) at temperature of 20 °C, air humidity 70 % and 14-h photoperiod with irradiance of 150 μmol m^{-2} s^{-1}. For ABA treatment, seedlings were sprayed with 200 μM ABA. Young leaves were sampled 0, 3, 6 and 12 h after spraying, frozen in liquid nitrogen, and prepared for total RNA isolation. For heat stress treatment, seedlings were transferred to another chamber maintained at 42 °C. Young leaves were harvested 0, 3, 6 and 12 h after heat treatment, frozen in liquid nitrogen, and used for total RNA extraction. Two replicates from each preparation and each treatment were collected to minimize variation during processing.

RNA extraction, cDNA synthesis, 3'RACE analysis and cloning of *sHSP* genes: Total RNA was extracted using *TRIzol* reagent (*Invitrogen*, USA) and was treated with RNase-free DNase I (*Fermentas*, Canada) to digest remaining genomic DNA. First-strand cDNA was synthesized from total RNA using oligo-$(dT)_{18\sim25}$ and moloney murine leukemia virus reverse transcriptase (*Promega*, USA). Hereafter, the resulted single-stranded cDNA was stored for future qRT-PCR. Additionally, DNaseI-treated RNA was also primed with oligo(dT)-containing adapter primer (AP) and transcribed into cDNA using reverse transcriptase (RT) *SuperScriptII* (*Invitrogen*). Abridged universal amplification primer (AUAP) was used as antisense primer for PCR. According to homology of different species, degenerate primers were designed for downstream 3'RACE cDNA cloning (Table 1). The 3'RACE PCR was run using degenerate primer and AUAP. The 3'RACE PCR conditions were as follows: 5 min at 94 °C, 30 s at 94 °C, 30 s at 55 °C, 80 s at 72 °C, followed by 36 cycles, and a final extension step of 10 min at 72 °C. Successful 3'RACE products were acquired and checked on the agarose gels. The amplified PCR products were purified and cloned with the dual promoter TA cloning ® kit (*Invitrogen*). Clones, showing different sizes on the agarose gel, were selected for sequencing in colony PCR. Sequencing results were analyzed using *BLASTX* algorithms (http://www.ncbi. nlm.nih.gov/ BLAST/).

Sequence analysis, cellular location and protein secondary structure prediction: Coding region, stop codon and 3'-UTR were analyzed using open reading frame finder in *NCBI* (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Molecular mass and isoelectric point (pI) of these sHSPs were calculated using *ProtParam* (http://web.expasy.org/protparam/). To gain a better understanding of the possible cellular locations of these sHSPs identified here, we performed cellular location predictions using the following programs: *WoLF PSORT* (http://wolfpsort.org/), *Predotar* (http://urgi.versailles.inra.fr/predotar/predotar.html), and *TargetP* (http://www.cbs.dtu.dk/services/TargetP/). The secondary structure of HSP16 from wheat (*T. aestivum*) was obtained from the protein sequence database (1GME) and was compared to secondary structure predictions of these sHSPs using *SPDBV_4.01_PC* (Guex and Peitsch 1997). The results were obtained from The *SWISS-MODEL Workspace* (http://swissmodel.expasy.org/workspace/) (Arnold *et al*. 2006). Additionally, *3D-PSSM* (http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html) was also used for secondary structure predictions.

Phylogenetic analysis: To understand how the sHSPs are related to each other, the complete amino acid sequences of CI, CII, and CIII sHSP of *Arabidopsis thaliana*, *Oryza sativa*, and *P. trichocarpa* were obtained from *GenBank*, and a phylogenetic tree was constructed using Neighbor-joining with *MEGA v. 3.1* (Kumar *et al*. 2004). Distance matrices were based on the Jones-Taylor-Thornton substitution matrix for the amino acid data. Bootstrapping (1000 replicates) was performed to quantify the relative support for branches of the inferred phylogenetic tree.

qRT-PCR: For qRT-PCR amplification experiments, it is necessary to design specific primer sets for closely related sequences. Primer sets were designed avoiding conserved regions using *Primer5* software. In our primer pairs, at least one of the primer sets was highly specific. PCR was carried out in a reaction volume of 20 mm^3 containing 1 mm^3 of diluted cDNA, 0.6 mm^3 of each reverse and forward primers and 17.8 mm^3 of the PCR master-mixture (*SYBR* qPCR mix, *Thunderbird,* Toyobo, Japan). The following PCR program was used: 95 ºC for 10 min, 95 ºC for 15 s, 60 ºC for 1 min; steps 2 and 3 were repeated 40 times. The specificity of the PCR amplification was checked with a heat dissociation curve (60 - 95 °C) following the final cycle of the PCR. The relative quantification analysis was performed according to threshold values (CT) generated from the *ABI StepOne*™ RT-PCR system (*Applied Biosystems*, USA). For calculating the gene expression level, the results were

standardized using *ACT7* as internal control. The relative quantification analysis was performed using the comparative ΔΔCt method (Schefe *et al*. 2006) and relative gene expression was expressed as fold change. Two biological replicates were programmed, and each cDNA sample was diluted and run in duplicate, and relative gene expression was expressed as mean ± SD.

Table 1. PCR primers used in this study.

Name	Sequence (5' - 3')	Comments
AP	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT	used for RT-PCR
AUAP	GGCCACGCGTCGACTAGTAC	3'RACE for cloning
DP1	ATGKCDMTNATYCCRAG	3'RACE for cloning
DP2	ATGGANTTNGKNAGGTTTCC	3'RACE for cloning
DP3	ATGGABTTRGABTTDGGRAG	3'RACE for cloning
DP4	ATGAGTGCTGTNGCGATHA	3'RACE for cloning
CbHSPCIaF	AATTCCCCTCGTCATCGCT	3'RACE for cloning/qRT-PCR for gene expression
CbHSPCIaR	CCTCTATCTCCACTTTGACCTCC	real-time qRT-PCR for gene expression
CbHSPCIbF	TCCGAGCATCTTTGGAGGT	3'RACE for cloning/ qRT-PCR for gene expression
CbHSPCIbR	CTTGGCTGGTGCGTTTG	real-time qRT-PCR for gene expression
CbHSPCIcF	CCCGAGCATTTTTGGAGGC	3'RACE for cloning/qRT-PCR for gene expression
CbHSPCIcR	CAACCCCGATGGCGTCAT	real-time qRT-PCR for gene expression
CbHSPCIdF	AACACCATCTTCTTCGCTACT	3'RACE for cloning/qRT-PCR for gene expression
CbHSPCIdR	CGCTGATCTTCAAAACACTCT	real-time qRT-PCR for gene expression
CbHSPCIIaF	TATCTCTGCTGTTTGTAACGAT	3'RACE for cloning/qRT-PCR for gene expression
CbHSPCIIaR	GGCGACTTGAACTTGAATT	qRT-PCR for gene expression
CbHSPCIIbF	TGTAACGACGGAGTCTTGAAG	3'RACE for cloning/qRT-PCR for gene expression
CbHSPCIIbR	AGGCGACTTGAACTTGAATG	qRT-PCR for gene expression
CbHSPCIIIaF	TAACCACTTGTTCGGCTTG	3'RACE for cloning/qRT-PCR for gene expression
CbHSPCIIIaR	GGAGACTCCAGGGATGTCA	qRT-PCR for gene expression
CbHSPCIIIbF	TTTTTATCTCGACATCCCC	3'RACE for cloning/qRT-PCR for gene expression
CbHSPCIIIbR	CTTAGTAGACCCTTCTTCACTCT	qRT-PCR for gene expression
ACT7F	GAGCGAGAAATTGTCCGTGA	qRT-PCR for gene expression
ACT7R	GAGCGATGGCTGGAATAGAA	qRT-PCR for gene expression

Results

Total mRNA of mature embryos was used as the template for 3'RACE. Successful 3'RACE products were obtained using DP1, DP2, and DP4 degenerate primers (Fig. 1), suggesting that *sHSP* genes related to DP3 are not expressed in the mature embryos. To obtain accurate sequences, each *sHSP* gene was cloned and sequenced three times. The results were identical across replicates for all genes. Four *sHSP* genes were obtained by amplification with DP1 degenerate primers, two *sHSP* genes were obtained using DP2 degenerate primers, and two *sHSP* genes were obtained using DP4 degenerate primers. To obtain homologous information of these sHSPs, we analyzed them using *BLASTX* in *NCBI* (Table 2). The eight different genes were named according to their homology, seven of which were similar to *A. thaliana* sequences; the other was similar to *C. rubella* sequences.

Table 2. Analysis of transcripts using *Open Reading Frame Finder* in *NCBI* and analysis of sequence homology using *BLASTX* in *NCBI* databases.

Name	Coding regions (nt)	Stop codon	Variable 3'-UTR (nt)	Homology	ID	e-value
CbHSPCIa	468	TGA	175, 187	HSP17.8-CI [*A. thaliana*]	NP_172220	1e-67
CbHSPCIb	471	TAA	89, 127	ATHSP17.4 [*A. thaliana*]	NP_190209	3e-62
CbHSPCIc	471	TAA	119, 156, 194	ATHSP17.4 [*A. thaliana*]	NP_190209	4e-69
CbHSPCId	483	TGA	120, 147	putHs42 [*C. rubella*]	ABW81130	4e-61
CbHSPCIIa	468	TAA	95, 113, 116, 169, 194	HSP17.6II [*A. thaliana*]	NP_196763	2e-66
CbHSPCIIb	468	TAA	153, 196, 209	HSP17.6II [*A. thaliana*]	NP_196763	2e-66
CbHSPCIIIa	462	TGA	66, 155, 165, 186.	HSP17.4III [*A. thaliana*]	NP_175807	7e-69
CbHSPCIIIb	462	TGA	73, 96, 141, 232.	HSP17.4III [*A. thaliana*]	NP_175807	6e-69

All eight genes had an integrated coding region and a corresponding 3'-UTR, and we calculated the length of each gene-coding region (Table 2). Table 3 summarizes the characteristics of these sHSP proteins, including their putative molecular masses, isoelectric points, and cellular localization.

Fig. 1. The results of 3'RACE with AUAP and the following degenerate primer: A - DP1, B - DP2 and C - DP4.

Deduced amino acid sequences of all eight sHSPs identified in *C. bursa-pastoris* were used for the phylogenetic analysis. To determine the phylogenetic relationships among these sHSPs and the CI, CII, and CIII sHSP of *A. thaliana*, *O. sativa*, and *P. trichocarpa*, the complete amino acid sequences of these latter proteins were obtained from *GenBank* and a Neighbor-joining phylogenetic tree was constructed (Fig. 2). The eight sHSPs were easily identified in the phylogenetic tree. Four of the sHSPs (CbHSPCIa, CbHSPCIb, CbHSPCIc, and CbHSPCId) belonged to the CI subfamily. CbHSPCIIa and CbHSPCIIb were classified into the CII subfamily, and CbHSPCIIIa and CbHSPCIIIb were classified into the CIII subfamily. Based on the phylogenetic tree, 36 sHSPs from 4 species were grouped into the largest sHSP subfamily, CI. The CIII subfamilies of *A. thaliana*, *O. sativa*, and *P. trichocarpa* each contained only one member, whereas the CIII subfamily of *C. bursa-pastoris* included two members. In addition, the CII and CIII subfamilies were more closely related to each other than to the CI subfamily. The *C. bursa-pastoris* sHSP subfamilies were closely related to those of *A. thaliana* with high bootstrap values.

To evaluate the 3'-UTRs of these *sHSP* genes, 3'RACE was performed with specific primer (CbHSPCIaF, CbHSPCIbF, CbHSPCIcF, CbHSPCIdF, CbHSPCIIaF, CbHSPCIIbF, CbHSPCIIIaF, and CbHSPCIIIbF) and AUAP primer. The eight 3'RACE products were extracted and cloned into specific vectors. The clones, showing length polymorphisms in the agarose gels, were selected for sequencing and the lengths of the variable 3'-UTRs were calculated (Table 2). Variable 3'-UTR lengths were observed in each of the eight *sHSP* genes. Moreover, there were variable 3'-UTRs in all CI, CII, and CIII subfamilies, although the UTRs of CII and CIII genes were more variable than those of CI genes.

Nucleotide sequences were aligned using *ClustalX*. *CbHSPCIIa* and *CbHSPCIIb* showed similar nucleotide sequences in the 3'-UTR, and less variation between their coding regions. A 40-nucleotide fragment present in the 3'-UTR of *CbHSPCIIb* did not exist in that of *CbHSPCIIa*. In addition, *CbHSPCIIIa* and *CbHSPCIIIb* transcripts showed 17 nucleotide changes in the coding region and 40 nucleotide changes in the 3'-UTR.

The eight sHSPs were predicted to localize in the cytosol, consistent with previous results on the sHSPs of *A. thaliana, O. sativa*, and *P. trichocarpa*. In addition, all sHSPs contained critical domains and motifs composed of eight β-strands (β2 - β9). However, there are number of secondary structural features that are conserved across CI, CII, and CIII sHSPs, with β2, β3, β8, and β9 being the

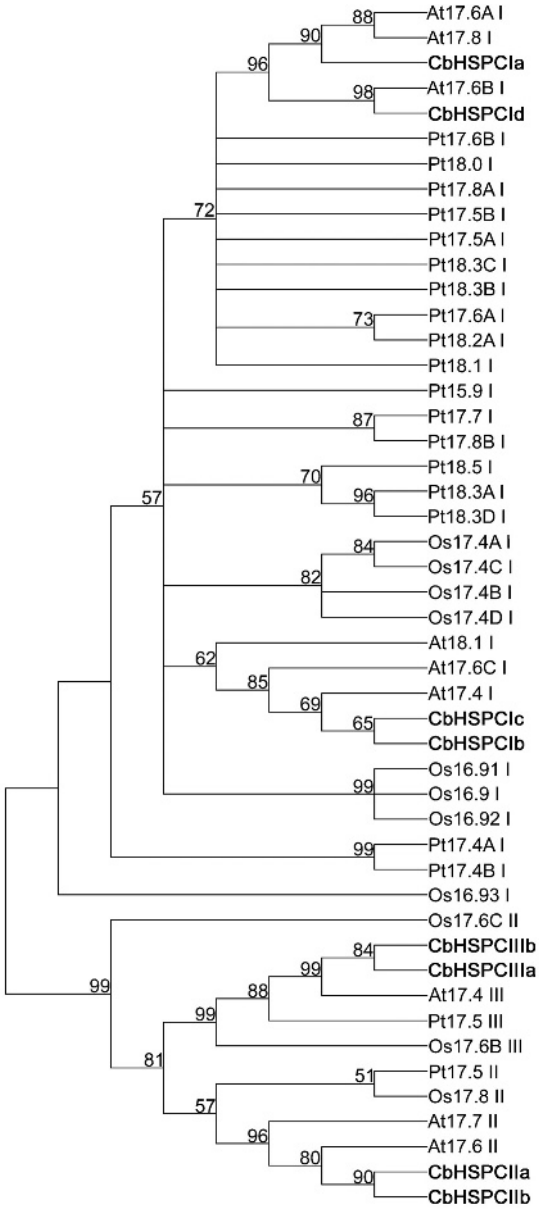

Fig. 2. The tree was derived by Neighbor-joining method with bootstrap analysis (1000 replicates) from alignment of amino acid sequences of conserved ACD of CI, CII and CIII sHSPs from *Capsella bursa-pastoris*, *Arabidopsis thaliana*, *Oryza sativa* and *Populus trichocarpa*. Numbers above the major branches indicate bootstrap values ≥ 50 %. The abbreviations of species are as follows: Cb - *Capsella bursa-pastoris*, At - *Arabidopsis thaliana*, Os - *Oryza sativa*, Pt - *Populus trichocarpa*. sHSPs of *C. bursa-pastoris* are shown in bold.

Table 3. Prediction of characteristics of the eight sHSP proteins localized in cytosol.

Name	Molecular mass [kD]	Isoelectric points
CbHSPCIa	17.7180	5.98
CbHSPCIb	17.5780	5.98
CbHSPCIc	17.5669	5.56
CbHSPCId	18.3925	5.97
CbHSPCIIa	17.6743	6.18
CbHSPCIIb	17.7024	6.76
CbHSPCIIIa	17.0744	6.16
CbHSPCIIIb	17.1024	5.96

most highly conserved across subfamilies. In particular, the "GVL" sequence motif of β9 is conserved across all subfamilies of all four species.

To further understand the expression of the eight *sHSP* genes, their transcriptional levels were analyzed in different organs and embryo stages by qRT-PCR. Expression of the eight genes was found to be highest in the stamen and lowest in the petals of reproductive organs. There were similar expression levels in sepals and pistils. In vegetative organs, little or no transcript was detectable in roots, stems, and leaves. During embryonic development, there were identified some differences in the expression patterns of four CI *sHSP* genes. *CbHSPCIc* showed distinct up-regulation during embryogenesis, and had the highest expression level in embryos after ripening. *CbHSPCIb* and *CbHSPCId* had similar expression patterns, and had the highest expression in mature embryos during embryogenesis (Table. 4). We compared the expressions of *CbHSPCIIa* and *CbHSPCIIb* and found that they both had similar expression profiles, but the *CbHSPCIIa* transcripts were expressed approximately five times higher than those of *CbHSPCIIb*. Both *CbHSPCIIa* and *CbHSPCIIb* were up-regulated during embryogenesis, similar to *CbHSPCIc*, and had the highest expression in embryos after ripening (Table 4). *CbHSPCIIIb* transcript levels were slightly higher than that of *CbHSPCIIIa* in reproductive organs such as petals, sepals, stamens, and pistils. The expression of *CbHSPCIIIa* was clearly higher than that of *CbHSPCIIIb* during embryonic development, but both had a maximum expression in mature and 4 months after ripening embryos (Table. 4).

To determine the effect of ABA and heat-stress on the *sHSP* genes, *C. bursa-pastoris* plants were treated with 200 μM ABA or 42 °C. RNA was isolated from young leaves 0, 3, 6 and 12 h after ABA treatment, qRT-PCR was performed. Exogenous ABA significantly enhanced the expression of *CbHSPCIa*, *CbHSPCIb*, *CbHSPCId*, *CbHSPCIIa*, *CbHSPCIIb*, *CbHSPCIIIa*, and *CbHSPCIIIb*. In particular, *CbHSPCIa*, *CbHSPCId, CbHSPCIIa*, and *CbHSPCIIb* were clearly up-regulated. *CbHSPCIa*,

Table 4. qRT-PCR showing differential gene expression of 4 CI *sHSP* genes, 2 CII *sHSP* genes and 2 CIII *sHSP* genes at different organs and embryo developmental stages. GH - mixes of globular and heart embryos, T - torpedo embryos, C - cotyledon embryos, P - premature embryos, M - mature embryos, AR - embryos 4 months after ripening. *ACT7* was used as an internal control. Data represent means ± SD of two biological replicates.

Gene	Root	Stem	Leaf	Petal	Sepal	Stamen	Pistil	GH	T	C	P	M	AR
CbHSPCIa	0	0	0	1.4±0.2	4.7±1.5	11.0±1.4	5.1±3.4	0	0.3±0.1	0.4±0.2	1.3±0.5	3.6±1.1	4.3± 0.6
CbHSPCIb	0	0	0	0.5±0	2.5±0.9	6.4±2.4	1.6±0.8	0	0.1±0	0.2±0.1	0.5±0.3	4.4±0.6	1.5± 0.8
CbHSPCIc	0	0	0	0.3±0.1	1.3±0.5	3.6±0.8	1.0±0.7	0	0	0	0.7±0.2	10.9±2.0	52.5±22.1
CbHSPCId	0	0	0	0.6±0.1	2.9±0.7	7.4±1.2	2.5±1.7	0	0.2±0.1	0.5±0.3	1.2±0.8	5.6±2.0	3.3± 0.5
CbHSPCIIa	0	0	0	2.1±0.2	6.5±2.4	11.6±2.3	3.8±2.3	0	0.1±0	0.2±0.1	0.8±0.1	8.2±3.3	26.9± 1.0
CbHSPCIIb	0	0	0	0.6±0	1.5±0.7	2.9±0.6	0.7±0.4	0	0	0	0.1±0	1.5±0.6	4.2± 0.2
CbHSPCIIIa	0	0	0	1.4±0.1	3.0±1.0	7.8±1.0	3.0±1.9	0	0.2±0.1	0.2±0.1	1.1±0.3	2.1±0.2	2.1± 0.1
CbHSPCIIIb	0	0	0	1.6±0	3.8±1.3	8.2±1.1	3.7±1.9	0	0.1±0	0.1±0	0.4±0.1	0.9±0.1	1.0± 0

Table 5. qRT-PCR showing expression of 8 *sHSP* genes when young leaves were treated with 200 μM ABA. *ACT7* was used as an internal control. Data represent means ± SD of two biological replicates.

Gene	0 h	3 h	6 h	12 h
CbHSPCIa	0.00129±0.00024	0.13713±0.01332	0.08314±0.03808	0.02541±0.00637
CbHSPCIb	0.00019±0.00002	0.00102±0.00010	0.00070±0.00018	0.00017±0.00008
CbHSPCIc	0.00039±0.00005	0.00058±0.00002	0.00026±0.00007	0.00021±0.00006
CbHSPCId	0.00321±0.00074	0.02382±0.00254	0.09290±0.04357	0.05533±0.00788
CbHSPCIIa	0.00031±0.00013	0.02655±0.00252	0.01485±0.00619	0.00413±0.00086
CbHSPCIIb	0.00020±0.00016	0.01402±0.00029	0.01694±0.00785	0.00599±0.00073
CbHSPCIIIa	0.00494±0.00152	0.02859±0.00507	0.02782±0.01262	0.02784±0.00082
CbHSPCIIIb	0.00379±0.00069	0.01332±0.00030	0.02085±0.01168	0.01863±0.00141

CbHSPCIb, *CbHSPCIc*, and *CbHSPCIIa* showed the highest expression at 3 h, and were down-regulated from 3 to 12 h after ABA treatment. However, the highest expression levels of *CbHSPCId*, *CbHSPCIIb*, *CbHSPCIIIa*, and *CbHSPCIIIb* were detected at 6 h. No distinct changes were detected in the transcriptional levels of *CbHSPCIc* (Table. 5). In contrast, all eight *sHSP* genes clearly responded when *C. bursa-pastoris* plants were exposed to 42 °C for 3, 6 and 12 h. All *sHSP* genes had similar expression patterns, with maximal expression at 42 °C for 3 h, and slight changes between 6 and 12 h (Table 6).

Discussion

Subfamilies of *sHSP* genes from *C. bursa-pastoris*: The eight *sHSP* genes were identified and analyzed using *BLASTX* in *NCBI*, which indicated that seven of the genes were highly homologous to the corresponding *sHSP* genes of *A. thaliana*. One gene (*CbHSPCId*) was obviously homologous to *putHs42* of *C. rubella* (EU162611).

Table 6. qRT-PCR showing expression of 8 *sHSP* genes when young leaves was treated with 42 °C. *ACT7* was used as an internal control. Data represent means ± SD of two biological replicates.

Gene	0 h	3 h	6 h	12 h
CbHSPCIa	0	53.3±31.0	8.2±2.0	12.8±0.2
CbHSPCIb	0	38.3± 5.7	5.1±1.2	6.3±1.8
CbHSPCIc	0	26.6± 5.5	4.4±1.0	6.9±0.4
CbHSPCId	0	32.3± 6.8	4.4±1.0	5.2±0.4
CbHSPCIIa	0	41.3±18.7	8.4±0.1	8.6±1.7
CbHSPCIIb	0	23.0± 3.7	3.9±0.1	4.4±1.0
CbHSPCIIIa	0	7.0± 1.6	0.8±0.2	1.0±0.1
CbHSPCIIIb	0	8.9± 1.8	1.2±0.1	2.4±0.5

CbHSPCId possesses all of the conserved amino acid sequence domains, and has the highest molecular mass of the eight sHSPs. Identification of *CbHSPCId* provided new evidence suggesting that *C. bursa-pastoris* harbors some *C. rubella* alleles, and indicated that *C. rubella* contributed to the gene pool of *C. bursa-pastoris* (Slotte *et al*. 2006). Based on a phylogenetic analysis combining the sHSPs of the other species, eight sHSPs were characterized as CI, CII, and CIII subfamilies. Each subfamily contained *C. bursa-pastoris*, *A. thaliana, O. sativa*, and *P. trichocarpa* sHSPs, indicating that their origins predate the divergence of the common ancestor of these four species (Waters *et al*. 2008). Indeed, it has been shown that at least the CI and CII subfamilies diverged more than 400 million years ago (Waters and Vierling 1999a,b, Waters and Rioflorido 2007). This phylogenetic tree demonstrated that CbHSPCIb and CbHSPCIc, which formed a clade, are closely related to *Arabidopsis* CI 17.4 (NP_190209). CbHSPCIIa and CbHSPCIIb formed the most similar group, corresponding to the CII 17.6 subfamily of *A. thaliana* (NP_196763), whereas CbHSPCIIIa and CbHSPCIIIb formed the next closest group, related to the CIII 17.4 subfamily of *A. thaliana* (NP_175807). Although we could not identify all of the *sHSP* genes in *C. bursa-pastoris*, our results expand the known *sHSP* genes in this species. It is likely that the expansion of some of these genes is due to genome-wide polyploidization events. As a tetraploid plant, genome-wide events would have increased the copy number of all three subfamilies equally. Genome duplication (poly-ploidy) is a driving force of plant gene family evolution (Blanc and Wolfe 2004, Casneuf *et al.* 2006). As a widespread weed, the success of *C. bursa-pastoris* is probably related to such gene expansion, in contrast to the other two species in *Capsella*.

Variable 3' end of *sHSP* genes: In eukaryotes, mRNA molecules include the coding and noncoding regions. Noncoding regions contain the cap, 5'-UTR, 3'-UTR, and poly(A) tail. The 3'-UTR, following the coding region, is a specific section of mRNA. Several regulatory sequences are found in the 3'-UTR, such as the poly(A) signal site that is usually AATAAA (or a slight variant). The 3'-UTR can harbor a poly(A) signal (an AU-rich element that regulates mRNA stability) that regulates the subcellular localization of transcripts (Jansen 2001, Mitchell and Tollervey 2001).

Multiple detailed nucleotide sequence alignments revealed variable 3'-UTRs in the CI, CII, and CIII *sHSP* genes. Five variable 3'-UTRs existed in *CbHSPCIIa*, and three in *CbHSPCIIb*. In addition, there were four variable 3'-UTRs in *CbHSPCIIIa* and *CbHSPCIIIb*, respectively. These transcripts, which have identical coding regions but different 3'-UTR lengths, generated different mRNA products from the same pre-mRNA. These results demonstrated that variable 3'-UTR universally exists in CII and CIII *sHSP* genes. Accordingly, variable 3' ends were also found in all four CI *sHSP* genes in our study. In contrast to the variable 3'-UTRs in the CII and CIII *sHSP* genes, each CI *sHSP* gene had two or three variable 3'-UTRs as well as classic poly(A) signals. Due to these differences in the 3'-UTRs of CI, CII, and CIII *sHSP* genes, these subfamilies would be a good model for future research on 3'-UTR function. The 3'-UTR, located between the stop codon and the poly(A) tail, may possess a variety of regulatory mechanisms. The 3'-UTRs of the two cytosolic glutamine synthetase genes were characterized as regulators of transcript stability in response to gluta-mine in *Medicago sativa* (Simon and Sengupta-Gopalan 2010). Splicing and alternative 3' end processing of mRNAs mediate riboswitch control of gene expression in plants (Wachter *et al*. 2007). Variable 3'-UTRs have been observed in cold-inducible *KIN* genes in *Arabidopsis* (Kurkela and Franck 1990). The existence of variable

3'-UTRs suggests that there are many complicated regulatory mechanisms in *sHSP* genes. This could be important for widespread weeds to adapt to diverse environments.

Comparative analysis of sHSP gene expression in *C. bursa-pastoris* and *Arabidopsis*: *Arabidopsis sHSP* gene expression profiles in different organs and during seed development are available (Kotak *et al*. 2007), making it possible to compare the gene expression of *sHSP* genes between *C. bursa-pastoris* and *A. thaliana*. All eight *C. bursa-pastoris sHSP* genes were assessed and we found that the expression profiles of some *sHSP* genes were similar to the corresponding *Arabidopsis sHSP* genes.

For all eight genes, there was low or no expression in roots, stems, and leaves of *C. bursa-pastoris* under non-stressed conditions. The results were similar to those of the corresponding *Arabidopsis sHSP* genes. In contrast, eight *sHSP* genes were expressed in petals, sepals, pistils, and stamens (the reproductive organs) of *C. bursa-pastoris*, whereas most of the relevant *Arabidopsis sHSP* genes were expressed only in pollen grains (Kotak *et al*. 2007). During embryo development, the expression profiles of five *sHSP* genes (*CbHSPCIa*, *CbHSPCIc*, *CbHSPCId*, *CbHSPCIIIa* and *CbHSPCIIIb*) were similar to those of the corresponding *Arabidopsis* genes. For example, *CbHSPCIc* was homologous to *At17.4 I* (AT3G46230), and was up-regulated and peaked in embryos after ripening, suggesting that it might be required for desiccation tolerance in *C. bursa-pastoris*. Kotak *et al*. (2007) indicated that *At17.4 I* was significantly up-regulated during embryo development of *Arabidopsis via* the regulation of HSFA9. Loss of function of the HSFA9 results in a drastic reduction in the accumulation of seed heat shock proteins including seed-specific sHSPs belonging to cytosolic (CI, CII) classes (Tejedor-Cano *et al*. 2010). The complexity of the *sHSP* genes in tetraploid *C. bursa-pastoris*, combined with the complicated genome, may explain the expression differences. Although *CbHSPCIb* and *CbHSPCIc* formed a clade with *At17.4 I* in our phylogenetic tree, *CbHSPCIb* did not peak in embryos after ripening, which is different from *At17.4 I* (AT3G46230). The *Arabidopsis CII 17.7 sHSP* (AT5G12030) is expressed in pollen grains and is up-regulated during seed development. No or little *CII 17.6* gene (AT5G12020) transcript was detected in *A. thaliana* seeds (Kotak *et al.* 2007). On the contrary, we did not clone the corresponding *CII 17.7 sHSP*, but 3'RACE analysis of two *CII 17.6 sHSP* genes in *C. bursa-pastoris* suggested that the corresponding *CII 17.7 sHSP* is not expressed in embryos, or does not exist in *C. bursa-pastoris*. *CbHSPCIIa* and *CbHSPCIIb* are homologous to *CII 17.6 sHSP* of *A. thaliana*, and were up-regulated, peaking in embryos after ripening of *C. bursa-pastoris*, similar to *Arabidopsis CII 17.7 sHSP* (Kotak *et al*. 2007). This suggests that there is a positive correlation between up-regulation of the two CII *sHSP* genes and desiccation tolerance in such after-ripening embryos of *C. bursa-pastoris*. Thus, overproduction of the *ATCII 17.7 sHSP* gene may increase salt and drought tolerance in *Arabidopsis* (Sun *et al*. 2001).

We also compared the gene expression of two CII *sHSP* genes and two CIII *sHSP* genes in *C. bursa-pastoris* by qRT-PCR. The two former genes had similar gene expression patterns but different expression levels, with the expression of *CbHSPCIIa* being approximately five times higher than that of *CbHSPCIIb* in different organs and during embryo development (Table 4). The transcriptional level of *CbHSPCIIIa* was similar to that of *CbHSPCIIIb* in reproductive organs, but about two times higher in the former during embryo development (Table 4). Recently, Rampino *et al*. (2010) indicated that the *HaHSP17.6a* and *HaHSP17.6b* genes have different expression under heat-stress in sunflower. In *Arabidopsis*, both *CII 17.7 sHSP* and *CIII 17.4 sHSP* response to high irradiance, heat-shock and H_2O_2 treatment, and showed up-regulation in transgenic plants overexpressing *HsfA2* compared with wild-type plants under normal conditions (Nishizawa *et al*. 2006). In addition, we compared the effects of ABA and heat stress on the expression of *sHSP* genes. Eight *sHSP* genes showed more obvious transcriptional changes when exposed to temperatures of 42 °C than those exposed to 200 μM ABA. The sHSPs, as well as other HSPs, are proved to play a significant role in plant stress tolerance, such as extreme temperatures (Campbell *et al*. 2001, Jiang *et al*. 2009, Grigorova *et al*. 2011), dehydration (Campbell *et al.* 2001, Sato and Yokoya 2008, Hussain *et al*. 2011), ABA treatment (Campbell et al. 2001). In conclusion, all *sHSP* genes showed multiple cellular responses to environmental stress, during various developmental processes, and in different reproductive organs.

References

Arnold, K., Bordoli, L., Kopp, J., Schwede, T.: The *SWISS-MODEL* workspace: a web-based environment for protein structure homology modelling. - Bioinformatics **22**: 195-201, 2006.

Blanc, G., Wolfe, K.H.: Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. - Plant Cell **16**: 1679-1691, 2004.

Bi, V.I., McMullen, M.D., Sanchez-Villeda, H., Schroeder, S., Gardiner, J., Polacco, M., Soderlund, C., Wing, R., Fang, Z., Coe E.H., Jr.: Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maize fingerprint contig physical map. - Crop Sci. **46**: 12-21, 2006.

Campbell, J.L., Klueva, N.Y., Zheng, H.G., Nieto-Sotelo, J., Ho, T.D., Nguyen, H.T.: Cloning of new members of heat shock protein *HSP101* gene family in wheat (*Triticum aestivum* (L.) Moench) inducible by heat, dehydration, and ABA. - Biochim. biophys. Acta **1517**: 270-277, 2001.

Casneuf, T., De Bodt, S., Raes, J., Maere, S., Van de Peer, Y.: Nonrandom divergence of gene expression following gene

and genome duplications in the flowering plant *Arabidopsis thaliana*. - Genome Biol. **7**: R13, 2006.

Dafny-Yelin, M., Tzfira, T., Vainstein, A., Adam Z.: Non-redundant functions of sHSP-CIs in acquired thermotolerance and their role in early seed development in *Arabidopsis*. - Plant mol. Biol. **67**: 363-373, 2008.

Eveland, A.L., McCarty, D.R., Koch, K.E.: Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. - Plant Physiol. **146**: 32-44, 2008.

Grigorova, B., Vaseva, I., Demirevska, K., Feller, U.: Combined drought and heat stress in wheat: changes in some heat shock proteins. - Biol. Plant. **55**: 105-111, 2011.

Guex, N., Peitsch, M.C.: *SWISS-MODEL* and the *Swiss-Pdb Viewer*: an environment for comparative protein modelling. - Electrophoresis **18**: 2714-2723, 1997.

Hurka, H., Bleeker, W., Neuffer, B.: Evolutionary process associated with biological invasions in the *Brassicaceae*. - Biol. Invasions **5**: 281-292, 2003.

Hurka, H., Neuffer, B.: Evolutionary processes in the genus *Capsella* (*Brassicaceae*). - Plant Syst. Evol. **206**: 295-316, 1997.

Hussain, S.S., Iqbal, M.T., Arif M.A., Amjad, M.: Beyond osmolytes and transcription factors: drought tolerance in plants *via* protective proteins and aquaporins. - Biol. Plant. **55**: 401-413, 2011.

Jansen, R.P.: mRNA localization: message on the move. - Nat. Rev. mol. cell. Biol. **2**: 247-256, 2001.

Jiang, C., Xu, J., Zhang, H., Zhang, X., Shi, J., Li, M., Ming, F.: A cytosolic class I small heat shock protein, RcHSP17.8, of Rosa chinensis confers resistance to a variety of stresses to *Escherichia coli*, *yeast* and *Arabidopsis thaliana*. - Plant Cell Environ. **32**: 1046-1059, 2009.

Kalemba, E.M., Pukacka, S.: Changes in late embryogenesis abundant proteins and a small heat shock protein during storage of beech (*Fagus sylvatica* L.) seeds. - Environ. exp. Bot. **63**: 274-280, 2008.

Kim, R., Kim, K.K., Yokota, H., Kim, S.H.: Small heat shock protein of *Methanococcus jannaschii*, a hyperthermophile. - Proc. nat. Acad. Sci. USA **95**: 9129-9133, 1998.

Kotak, S., Vierling, E., Bäumlein, H., Von Koskull-Döring, P.: A novel transcriptional cascade regulating expression of heat stress proteins during seed development of *Arabidopsis*. - Plant Cell **19**: 182-195, 2007.

Kumar, S., Tamura, K., Nei, M.: *MEGA3*: integrated software for molecular evolutionary genetics analysis and sequence alignment. - Brief Bioinform. **5**: 150-163, 2004.

Kurkela, S., Franck, M.: Cloning and characterization of a cold- and ABA-inducible *Arabidopsis* gene. - Plant mol. Biol. **15**: 137-144, 1990.

Mitchell, P., Tollervey, D.: mRNA turnover. - Curr. Opin. Cell Biol. **13**: 320-325, 2001.

Nautiyal, P.C., Shono, M.: Analysis of the role of mitochondrial and endoplasmic reticulum localized small heat shock proteins in tomato. - Biol. Plant. **54**: 715-719, 2010.

Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., Shigeoka S.: *Arabidopsis* heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. - Plant J. **48**:535-547, 2006.

Rampino, P., Mita, G., Assab, E., De Pascali, M., Giangrande, E., Treglia1, A.S., Perrotta1, C.: Two sunflower *17.6HSP* genes, arranged in tandem and highly homologous, are induced differently by various elicitors. - Plant Biol. **12**: 13-22, 2010.

Sato, Y., Yokoya, S.: Enhanced tolerance to drought in transgenic rice plants overexpressing a small heat-shock protein, sHSP17.7. - Plant Cell Rep. **27**: 329-334, 2008.

Scharf, K.D., Siddique, M., Vierling, E.: The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing alpha-crystallin domains (Acd proteins). - Cell Stress Chaperon. **6**: 225-237, 2001.

Schefe, J.H., Lehmann, K.E., Buschmann, I.R., Unger, T., Funke-Kaiser, H.: Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. - J. Mol. Med. **84**: 901-910, 2006.

Siddique, M., Gernhard, S., Von Koskull-Döring, P., Vierling, E., Scharf, K.D.: The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. - Cell Stress Chaperon. **13**: 183-197, 2008.

Simon, B., Sengupta-Gopalan, C.: The 3' untranslated region of the two cytosolic glutamine synthetase (GS1) genes in alfalfa (*Medicago sativa*) regulates transcript stability in response to glutamine. - Planta **232**: 1151-1162, 2010.

Slotte, T., Ceplitis, A., Neuffer, B., Hurka, H., Lascoux, M.: Intrageneric phylogeny of *Capsella* (*Brassicaceae*) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. - Amer. J. Bot. **93**: 1714-1724, 2006.

Slotte, T., Holm, K., McIntyre, L.M., Lagercrantz, U., Lascoux, M.: Differential expression of genes important for adaptation in *Capsella bursa-pastoris* (*Brassicaceae*). - Plant Physiol. **145**: 160-173, 2007.

Sun, W., Bernard, C., Van de Cotte, B., Montagu, M.V., Verbruggen, N.: *At-HSP17.6A*, encoding a small heat-shock protein in *Arabidopsis*, can enhance osmotolerance upon overexpression. - Plant J. **27**: 407-415, 2001.

Sun, W., Montagu, M.V., Verbruggen, N.: Small heat shock proteins and stress tolerance in plants. - Biochim. biophys. Acta **1577**: 1-9, 2002.

Sun, Y., MacRae, T.H.: Small heat shock proteins: molecular structure and chaperone function. - Cell. Mol. Life Sci. **62**: 2460-2476, 2005.

Sundby, C., Härndahl, U., Gustavsson, N., Åhrman, E., Murphy, D.J.: Conserved methionines in chloroplasts. - Biochim. biophys. Acta **1703**: 191-202, 2005.

Tejedor-Cano, J., Prieto-Dapena, P., Almoguera, C., Carranco, R., Hiratsu, K., Ohme-Takagi, M., Jordano, J.: Loss of function of the HSFA9 seed longevity program. - Plant Cell Environ. **33**: 1408-1417, 2010.

Van Montfort, R.L., Basha, E., Friedrich, K.L., Slingsby, C., Vierling, E.: Crystal structure and assembly of a eukaryotic small heat shock protein. - Nat. Struct. Biol. **8**: 1025-1030, 2001.

Van Montfort, R.L., Slingsby, C., Vierling, E.: Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. - Adv. Protein Chem. **59**: 105-156, 2002.

Wachter, A., Tunc-Ozdemir, M., Grove, B.C., Green, P.J., Shintani, D.K., Breaker, R.R.: Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. - Plant Cell **19**: 3437-3450, 2007.

Waters, E.R., Aevermann, B.D., Sanders-Reed, Z.: Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. - Cell Stress Chaperon. **13**: 127-142, 2008.

Waters, E.R., Rioflorido, I.: Evolutionary analysis of the small heat shock proteins in five complete algal genomes. - J. mol. Evol. **65**: 162-174, 2007.

Waters, E.R., Vierling, E.: Chloroplast small heat shock proteins: Evidence for atypical evolution of an organelle-localized protein. - Proc. nat. Acad. Sci. USA **96**: 14394-14399, 1999a.

Waters, E.R., Vierling, E.: The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. - Mol. Biol. Evol. **16**: 127-139, 1999b.

Wang, W., Vinocur, B., Shoseyov, O., Altman, A.: Roles of plant heat-shock proteins and molecular chaperones in the abiotic stress response. - Trends Plant Sci. **9**: 244-252, 2004.

Xue, Y., Peng, R., Xiong, A., Li, X., Zha, D., Yao, Q.: Over-expression of heat shock protein gene *hsp26* in *Arabidopsis thaliana* enhances heat tolerance. - Biol. Plant. **54**: 105-111, 2010.