Genetics and Molecular Research

Genome-wide identification, classification, and expression analysis of sHSP genes in Chinese cabbage (*Brassica rapa* ssp *pekinensis*)

P. Tao[1], W.L. Guo[2], B.Y. Li[1], W.H. Wang[1], Z.C. Yue[1], J.L. Lei[1] and X.M. Zhong[1]

[1]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, China
[2]Zhejiang Institute of Communication and Media, Hangzhou, China

Corresponding author: X.M. Zhong
E-mail: zhongxm@mail.zaas.ac.cn

Genet. Mol. Res. 14 (4): 11975-11993 (2015)
Received January 23, 2015
Accepted May 15, 2015
Published October 5, 2015
DOI http://dx.doi.org/10.4238/2015.October.5.11

ABSTRACT. Small heat shock proteins (sHSPs) are essential for the plant's normal development and stress responses, especially the heat stress response. The information regarding sHSP genes in Chinese cabbage (*Brassica rapa* ssp *pekinensis*) is sparse, hence we performed a genome-wide analysis to identify sHSP genes in this species. We identified 26 non-redundant sHSP genes distributed on all chromosomes, except chromosome A7, with one additional sHSP gene identified from an expressed sequence tag library. Chinese cabbage was found to contain more sHSP genes than *Arabidopsis*. The 27 sHSP genes were classified into 11 subfamilies. We identified 22 groups of sHSP syntenic orthologous genes between Chinese cabbage and *Arabidopsis*. In addition, eight groups of paralogous genes were uncovered in Chinese cabbage. Protein structures of the 27 Chinese cabbage sHSPs were modeled using Phyre2, which revealed that all of them contain several conserved β strands across different subfamilies. In general, gene structure was conserved within each subfamily between

Chinese cabbage and *Arabidopsis,* except for peroxisome sHSP. Analysis of promoter motifs showed that most sHSP genes contain heat shock elements or variants. We also found that biased gene loss has occurred during the evolution of the sHSP subfamily in Chinese cabbage. Expression analysis indicated that the greatest transcript abundance of most Chinese cabbage sHSP genes was found in siliques and early cotyledon embryos. Thus, genome-wide identification and characterization of sHSP genes is a first and important step in the investigation of sHSPs in Chinese cabbage.

Key words: sHSP; Gene structure; *Brassica rapa* ssp *pekinensis;* Gene expression

INTRODUCTION

Heat stress adversely affects crop productivity by causing protein dysfunction. In order to prevent irreversible aggregation of denatured proteins under heat stress, many plants produce heat shock proteins (HSPs) that act as chaperones for maintaining correct protein folding. Based on their molecular weights, HSPs are divided into five families: HSP100s, HSP90s, HSP70s, HSP60s, and small HSPs (sHSPs) (Waters, 2013). sHSPs that have low molecular weights (usually less than 30 kDa) and contain a conserved α-crystallin domain (ACD; 80-100 AA), are ancient, ubiquitous, and diverse in living organisms. sHSPs usually form oligomers ranging from 9 to 50 subunits (200-800 kDa) (Basha et al., 2012; Lopes-Caitar et al., 2013). The ACD is located in the C-terminal region. The variable N-terminal region, preceding the ACD and a short C-terminal extension, are the source of structural diversity among sHSPs of different subfamilies. The crystal structures of wheat (*Triticum aestivum*) HSP16.9 (van Montfort et al., 2001) and archaebacterium (*Methanococcus jannaschii*) HSP16.5 (Kim et al., 1998) have been identified and applied for structure analysis of other sHSPs. Although sHSP proteins of different subfamilies differ in primary sequence and subcellular localization, they share a compact β-sheet sandwich structure. Compared with animals, plants contain a greater diversity of sHSP genes (Sarkar et al., 2009). In *Arabidopsis thaliana*, more than seven different sHSP subfamilies have been identified by genome analysis (Scharf et al., 2001). Six cytosolic-I genes constitute the largest sHSP subfamily in *A. thaliana* (Siddique et al., 2008; Waters et al., 2008). Comparative analysis of *A. thaliana*, *Populus trichocarpa*, and *Oryza sativa* (rice), has revealed one additional mitochondrial and three additional cytosolic sHSP subfamilies (Waters et al., 2008). The sHSP proteins of subfamilies CI to VI are localized in the cytosol or nucleus, with the remaining sHSPs targeted to peroxisomes (PX), the endoplasmic reticulum (ER), chloroplasts (CP), and mitochondria (MT-I and MT-II) (Scharf et al., 2001; Waters et al., 2008). These proteins are involved not only in cellular response to environmental stress, but also in various developmental processes such as pollen development, embryogenesis, fruit maturation, and seed germination (Wang et al., 2004; Sun and MacRae, 2005). sHSPs are crucial for enhancing the adaptation of plants subjected to heat, cold, drought, high light, or oxidative stress (Sun et al., 2002; Sundby et al., 2005; Dafny-Yelin et al., 2008).

Different *cis*-acting elements are involved in regulating the expression of sHSP genes in plants that have been subjected to stresses such as heat, low temperature, and exogenous abscisic acid (ABA) (Yi et al., 2006; Yi and Liu, 2009). Heat-induced expression of sHSP

genes is regulated mainly by heat shock transcription factors (HSFs) interacting with a conserved palindromic heat shock element (HSE) consisting of the adjacent and inverse repeats of the motif 5'-nGAAn-3' (Schöffl et al., 1998). In addition to HSE, some *cis*-acting elements, including ABA-responsive elements (ABREs), dehydration-responsive elements (DREs), and low temperature-responsive elements (LTREs), have been shown to be involved in quantitative regulation of the expression of different heat shock genes (Yi et al., 2006). As a *cis*-acting element, the ABRE motif plays an essential role in ABA-dependent pathways associated with the plant response to ABA (Nakashima et al., 2006; Huang et al., 2012), while the DRE/LTRE motif, is involved in gene expression regulation in response to dehydration, low temperature, and high-salt stress via ABA-independent pathways (Heidarvand and Amiri, 2010; Huang et al., 2012). The existence of numerous *cis*-acting elements, indicates the complexity of the regulatory network that controls the expression of sHSP genes.

Brassica, a genus of particular economic importance belonging to the mustard family, contains three diploid species, *Brassica rapa* (AA), *Brassica oleracea* (CC), and *Brassica nigra* (BB), and three allotetraploid species-*Brassica juncea* (AABB), *Brassica napus* (AACC), and *Brassica carinata* (BBCC). Chinese cabbage (*Brassica rapa* ssp *pekinensis*), an important vegetable crop originating from China, is considered to be a classic representative of the A genome of *Brassica.* The *B. rapa* genome is derived from the genome of a hexaploid ancestor, a triplicated diploid ancestral genome closely related to the *A. thaliana* genome (Wang et al., 2011; Cheng et al., 2013). The hexaploid ancestor of *B. rapa* is believed to have had three subgenomes, that have been designated on the basis of gene density and rate of gene loss (fractionation) as LF (the least fractionated), MF1 (the medium fractionated), and MF2 (the most fractionated) (Wang et al., 2011). *Brassica rapa* may have had a two-step origin; the first step being tetraploidization, followed by substantial genome fractionation, thus producing MF1 and MF2 genomes. The tetraploidized genome then subsequently hybridized with a third, less-fractionated genome (LF) (Cheng et al., 2012; Tang et al., 2012; Cheng et al., 2013).

In recent years, the Chinese cabbage genome (Chiifu-401-42) has been sequenced and assembled, and the results have been published (Wang et al., 2011). This sequenced genome provides an opportunity to carry out genome-wide investigations of important genes in Chinese cabbage. Thus far, genome-wide analysis of sHSP genes has been performed in at least four plant species: *A. thaliana*, *P. trichocarpa*, *Glycine max* (soybean), and *O. sativa* (Scharf et al., 2001; Waters et al., 2008; Sarkar et al., 2009; Lopes-Caitar et al., 2013). However, comprehensive analysis of sHSP genes in Chinese cabbage has not been performed, and several questions remain unsolved. For example, the number, location, and genomic organization of sHSP genes in the Chinese cabbage genome still need to be determined, and their *cis*-acting elements require characterization. Moreover, a comparison of the number of sHSP genes and their expression patterns in Chinese cabbage relative to *A. thaliana* would be of interest. To address these questions, in this study we identified 30 new sHSP genes through homology and synteny analyses, 29 from the Chinese cabbage genome and one from an expressed sequence tag (EST) library. Three sHSP genes were eliminated because of the incomplete ACD region of their corresponding proteins. The remaining 26 sHSP genes were found to be distributed over nine chromosomes, and these were classified into different groups according to a phylogenetic analysis. 3D structure analysis showed that all 27 sHSPs possess a conserved ACD region. Finally, we used published data to analyze sHSP gene structure, promoter motifs, and transcriptional levels in different organs.

MATERIAL AND METHODS

Identification and physical locations of sHSP genes in Chinese cabbage

Upon completion of the genome sequence project of Chinese cabbage (Wang et al., 2011), published data related to genomic sequence, gene sequence and protein sequence are available (http://brassicadb.org/brad/index.php). The nucleotide sequence and protein sequence of 19 *Arabidopsis* sHSP genes, 36 *P. trichocarpa* sHSP genes and 23 *O. sativa* sHSP genes were downloaded from the references (Waters et al., 2008). These sequences were used to search for sHSP genes in Chinese cabbage genome. We collected the resulting sequences showing similarity in coding region for further analysis. Additionally, we also acquired the sHSP genes by searching syntenic genes between *B. rapa* and *A. thaliana* with syntenic gene analysis (http://brassicadb.org/brad/searchSyntenytPCK.php) (Cheng et al., 2012). Chromosome location image of these sHSP genes was produced using the MapInspect software by mapping back to genome (the *B. rapa* v1.5)

Analysis of promoter motifs of sHSP genes in Chinese cabbage

The transcription start site (TSS) and TATA box of each sHSP gene was deduced by SoftBerry (http://linux1.softberry.com/berry.phtml). The 1000-bp region upstream relative to TSS of each sHSP gene was downloaded from Brassica Database. HSEs were identified by MEME (Version 4.9.1) (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) (Bailey et al., 2006) and the other cis-acting regulatory elements by PLACE (http://www.dna.affrc.go.jp/PLACE/signalscan.html).

Analysis of protein characteristics and phylogenetic relationship of sHSP proteins in Chinese cabbage

Molecular weight (MW) and isoelectric point (pI) of the Chinese cabbage sHSPs were predicted by an online tool ProtParam (http://web.expasy.org/protparam/). To analyze the evolutionary relationship of sHSPs, amino acid sequences of sHSP proteins of Chinese cabbage, and those of *A. thaliana*, *P. trichocarpa* and *O. sativa* (Waters et al., 2008), were aligned by ClustalX (version 1.83). A phylogenetic tree was produced by MEGA v. 5.2 with neighbor-joining method (Tamura et al., 2011). Distance matrices were based on the Jones-Taylor-Thornton substitution matrix with the pairwise deletion option. The reliability of the phylogenetic tree was assessed by performing one thousand bootstrap replicates.

Identification of syntenic *A. thaliana-B. rapa* (At-Br) orthologs and paralogs in Chinese cabbage

To show genome fractionation of sHSP genes and obtain each syntenic At-Br orthologs, syntenic genes were searched between *A. thaliana* and *B. rapa* by an online tool "syntenic gene analysis" (http://brassicadb.org/brad/searchSyntenytPCK.php) (Cheng et al., 2012).

Analysis of conserved motifs and 3D structure of sHSP proteins in Chinese cabbage

Motifs of Chinese cabbage sHSPs were identified by MEME (Version 4.9.1) (http://meme.nbcr.net/meme/cgi-bin/meme.cgi), the parameter settings were as follows: minimum number of sites, maximum number of sites, minimum motif width, maximum motif width, and maximum number of motifs to find were respectively 6, 38, 6, 30, and 10; the other parameters were default values (Bailey et al., 2006). 3D structure of the Chinese cabbage sHSPs were modeled using Phyre2 as quality control in order to exclude any deducing sHSPs not containing ACD domain.

Gene structure analysis

Gene structures of sHSP genes were analyzed by alignment of genomic DNA sequence and coding sequence of each sHSP gene. The gaps were identified as introns of sHSP genes. To indicate the corresponding site of each intron in second structure of sHSP proteins, the intron was mapped back to the second structure map.

Gene expression and EST expression profile

Raw RNA-seq data of *B. rapa* in different organs and embryo development stages have been published (Tong et al., 2013; Zhang et al., 2014), and these were downloaded respectively from NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession No. GSE43245 and the supplementary material of a recent publication (Zhang et al., 2014). The expression data of sHSP genes of *B. rapa* in different organs and embryo development stages were collected and used to produce a heat map by MeV (Saeed et al., 2003).

RESULTS

Identification and physical locations of sHSP genes in Chinese cabbage

To obtain putative sHSP genes, 19 *Arabidopsis* sHSP genes were compared against the Chinese cabbage genome (http://brassicadb.org/brad/index.php) using BLASTn. In total, 29 putative sHSP genes were acquired from the *B. rapa* genome. Three deduced sHSP genes (Bra037959, Bra012949, and Bra038854), showed different secondary structures in their corresponding proteins and were excluded. We did not obtain any Cytosolic-VI (C-VI) sHSP genes from the *B. rapa* genome, but one was acquired (accession No. GR722435) by searching the EST library at NCBI. Table 1 lists the remaining 27 sHSPs, all of which contained a conserved ACD domain. The number of sHSP genes in Chinese cabbage (27) was greater than the number in *Arabidopsis* (19) and rice (23) but less than in *P. trichocarpa* (36). The 27 sHSP genes were named according to their homologs in *Arabidopsis* (Table 1). To map their physical locations, the 26 sHSP genes were marked on the corresponding chromosomes of the *B. rapa* v1.5 genome sequence (Figure 1). No sHSP gene was located on chromosome 7. The genes encoding Br17.7A I and Br18.0A I were tandemly arranged. We analyzed pIs and MWs of the Chinese cabbage sHSPs according to their amino acid sequences. The sHSPs were found to be of low MW (15.0-25.1 kDa). Most sHSP proteins were acidic; the exceptions were Br15.5 IV, Br16.2, Br16.4, Br16.0 PX, and Br25.1 CP (Figure 2).

Table 1. Prediction of characteristics of 27 sHSP proteins and analysis of sequence homology using BLASTn against the *Brassica* database.

Protein name	Subfamily	Gene ID	pI	MW (kDa)
Br17.2A I	Cytosolic I	Bra006697	6.20	17.2
Br17.2B I	Cytosolic I	Bra020295	6.20	17.2
Br17.7A I	Cytosolic I	Bra018383	6.33	17.7
Br17.7B I	Cytosolic I	Bra018216	5.55	17.7
Br18.0A I	Cytosolic I	Bra018384	6.33	18.0
Br18.0B I	Cytosolic I	Bra031725	6.33	18.0
Br18.0C I	Cytosolic I	Bra002539	6.77	18.0
Br15.9 II	Cytosolic II	Bra006137	6.76	15.9
Br17.3 II	Cytosolic II	Bra008920	5.97	17.3
Br17.6 II	Cytosolic II	Bra023360	6.85	17.6
Br17.2 III	Cytosolic III	Bra030910	5.79	17.2
Br15.5 IV	Cytosolic IV	Bra020865	7.9	15.5
Br15.8 IV	Cytosolic IV	Bra013566	6.30	15.8
Br20.7 V	Cytosolic V	Bra022736	5.16	20.7
Br21.6 V	Cytosolic V	Bra002966	5.01	21.6
Br17.8VI	Cytosolic VI	GR722435	4.31	17.8
Br16.0 PX	Peroxisome	Bra028141	7.89	16.0
Br20.1 ER	ER	Bra027999	5.85	20.1
Br22.3 ER	ER	Bra000703	5.43	22.3
Br25.1 CP	CP	Bra026317	8.80	25.1
Br18.0 MT	MT	Bra028253	5.97	18.0
Br22.0 MT	MT	Bra029174	6.86	22.0
Br23.8 MT	MT	Bra013872	5.51	23.8
Br21.8 MTII	MTII	Bra018972	5.28	21.8
Br16.2		Bra024931	7.75	16.2
Br16.4		Bra024977	9.41	16.4
Br17.1		Bra022078	6.84	17.1

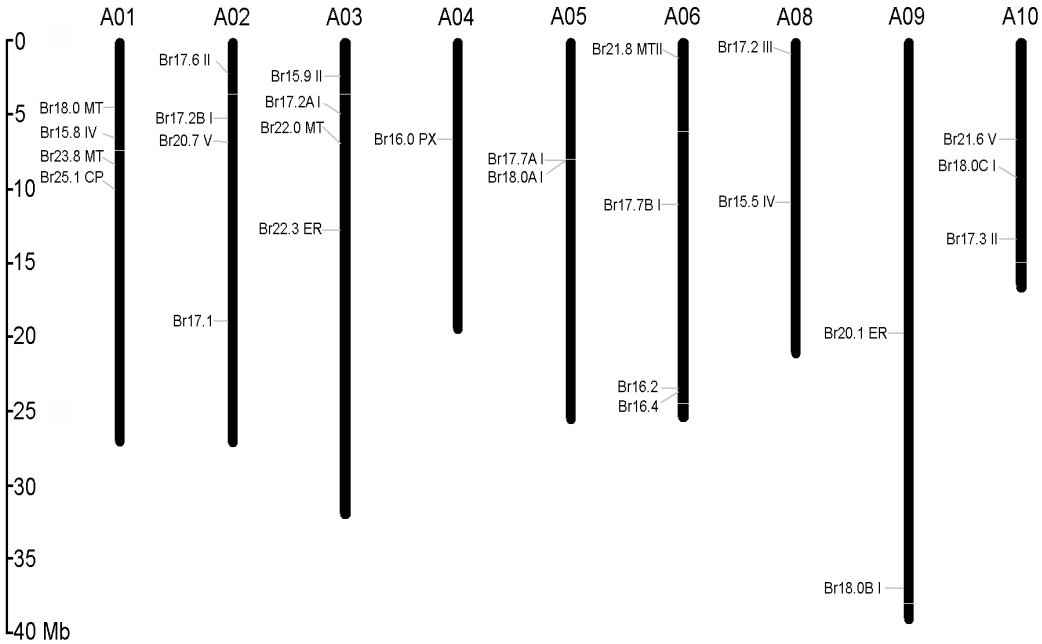


Figure 1. Chromosome localization of 26 sHSP genes from *Brassica rapa* ssp *pekinensis*.

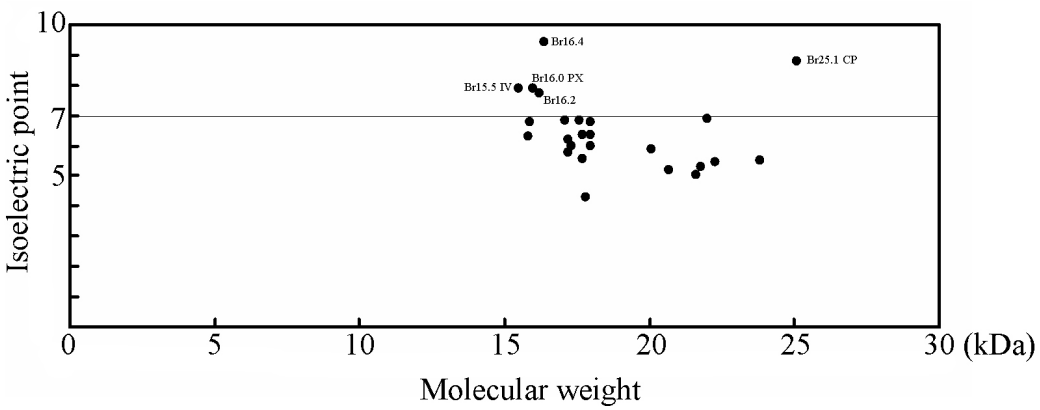


Figure 2. Putative isoelectric point and molecular weight of sHSP proteins from Chinese cabbage.

Paralogous sHSP genes in Chinese cabbage and orthologous sHSP genes between Chinese cabbage and *Arabidopsis*

According to an analysis of the draft genome sequence of Chinese cabbage (N = 10) by Wang et al. (2011), the hexaploid ancestor of Chinese cabbage contained LF, MF1, and MF2 subgenomes. To evaluate sHSP gene triplication, we identified orthologous sHSP genes between Chinese cabbage and *Arabidopsis* using the "syntenic genes" online colinearity comparison tool (Cheng et al., 2012). As shown in Table 2, we identified 22 groups of orthologous sHSP genes between Chinese cabbage and *Arabidopsis*. Most *A. thaliana* sHSP genes had one to three syntenic orthologs in the *B. rapa* genome. *At17.6A I* (AT1G59860), *At17.6C I* (AT1G53540), and *At17.8 I* (At1g07400) had no syntenic orthologs in Chinese cabbage, while *Br18.0B I* (Bra031725), *Br20.1 ER* (Bra027999), and *Br16.2* (Bra024931) conversely had no syntenic orthologs in *A. thaliana* (Table 2). In total, two *A. thaliana* sHSP genes had three syntenic orthologs each in Chinese cabbage, six *A. thaliana* sHSP genes had two orthologs each in Chinese cabbage, and seven *A. thaliana* sHSP genes had a single ortholog in Chinese cabbage. With respect to sHSP gene distributions in the three subgenomes of *B. rapa* (Wang et al., 2011), 14 sHSP genes were located in the LF subgenome, 7 in MF1, and 5 in MF2 (Table 2). Furthermore, a total of eight groups of paralogous sHSP genes were identified in Chinese cabbage based on the syntenic comparison (Table 2). The sHSP gene family of *B. rapa* and *A. thaliana* also experienced a tandem duplication event within the same chromosome (e.g., *At17.6 II* and *At17.7 II*; *Br17.7A I* and *Br18.0A I*). Tandem duplication of *Br17.7A I* and *Br18.0A I* was accompanied by the partial loss of the nucleotide sequence of the *Br17.7A I* gene.

Identification of promoter motifs in sHSP genes of Chinese cabbage

sHSP genes play an essential role in response to various stresses, especially heat stress, hence, the characterization of *cis*-regulatory elements of sHSP genes that regulate heat shock-responsive, ABA-responsive, low temperature-induced, and drought-responsive gene expression is important. To search for probable HSEs, we carried out sequence analysis using MEME (Bailey et al., 2006). Two motifs, similar to HSEs (5'-nGAAnnTTCnnGAAn-3' or 5'-nTTCnnGAAnnTTCn-3') and their variants, were identified (Figure 3). The HSE mod-

ule existed in Motif 1 (5'-nGAAnnTTCnnGAAnnTTCnnGAAn-3') (Figure 3A) and Motif 5 (5'-nTTCnnGAAnnTTCn-3') (Figure 3B). All 26 sHSP genes showed one or two HSEs and variants in the 1000-bp region upstream of TSS (Figure 3), most between 0 and 200 bp. The results of precise sequence searches for HSE (5'-nGAAnnTTCn-3' or 5'-nTTCnnGAAn-3'), DRE (RCCGAC), ABRE (ACGTG), and LTRE (CCGAM) are shown in Table 3. No HSE 5'-nGAAnnTTCn-3' sequence was found in the 1,000-bp TSS upstream region of *Br16.2*, *Br15.8 IV*, *Br20.7 V*, or *Br23.8 MT*, whereas the remaining sHSP genes contained one or more HSEs in this region. Most sHSP genes had greater numbers of HSEs than ABREs, DREs, and LTREs.

Table 2. Orthologous sHSP genes between Chinese cabbage and *Arabidopsis thaliana*, and syntenic paralogs of sHSP genes in three subgenomes of Chinese cabbage.

At ortholog	Paralogs of sHSP genes in Chinese cabbage		
	LF	MF1	MF2
AT3G46230 (At17.4 I)	Bra018216	No	No
AT1G59860 (At17.6A I)	No	No	No
At2G29500 (At17.6B I)	**Bra018383 / Bra018384**	No	No
-	-	-	Bra031725
AT1G53540 (At17.6C I)	No	No	No
At1g07400 (At17.8 I)	No	No	No
AT5G59720 (At18.1 I)	Bra002539	Bra006697	Bra020295
At5G12020 (At17.6 II)/At5G12030 (At17.7 II)	Bra008920	Bra006137	Bra023360
AT1G54050 (At17.4 III)	Bra037959	Bra030910	No
-	Bra012949	-	-
At4G21870 (At15.4 IV)	Bra013566	No	Bra020865
At5G54660 (At21.7 V)	Bra002966	No	Bra022736
At2g19310 (At18.5VI)	Bra038854	No	No
At5G37670 (At15.7 PX)	Bra028141	No	No
At4G10250 (At22.0 ER)	No	Bra000703	No
-	-	Bra027999	-
At4G27670 (At21 CP)	Bra026317	No	No
At4G25200 (At23.6 MT)	Bra013872	No	No
At5G51440 (At23.5 MT)	Bra028253	Bra029174	No
At1G52560 (At26.5MTII)	Bra018972	No	No
At5G47600 (At14.2)	Bra024977	Bra022078	No
-	Bra024931	-	-

LF, less-fractioned subgenome; MF1, medium-fractionated subgenome; MF2, more-fractioned subgenomes. Paralogs of sHSP genes in Chinese cabbage, whose corresponding proteins have incomplete ACD are highlighted by a single underline. Tandem-arranged sHSP genes in Chinese cabbage are marked in bold. "-" indicates no syntenic orthologs in the corresponding subgenome of Chinese cabbage or *Arabidopsis*.

Phylogenetic analysis of sHSP proteins of Chinese cabbage

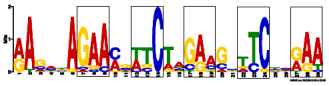
We identified a total of 27 sHSP genes from the Chinese cabbage genome and the NCBI data. To examine evolutionary relationships among sHSPs from *B. rapa*, *A. thaliana*, *P. trichocarpa*, and *O. sativa*, we generated a neighbor-joining phylogenetic tree of their amino acid sequences. In the phylogenetic tree, sHSP proteins from Chinese cabbage were clearly grouped into different subfamilies: 7 into the cytosolic I (C-I) class (the largest subfamily), 3 into cytosolic II (C-II), 1 into cytosolic III (C-III), 2 into cytosolic IV (C-IV), 2 into cytosolic V (C-V), 1 into cytosolic VI (C-VI), 1 into PX, 2 into ER, 1 into CP, 3 into MT, and 1 into MT-II. The remaining three sHSPs of Chinese cabbage (Br16.2, Br16.4, and Br17.1) were closely related to the orphan sHSP (At14.2) of *A. thaliana* with high bootstrap support (100%). All the sHSP proteins of Chinese cabbage and *Arabidopsis* gathered in smaller clusters than *P. trichocarpa* and *O. sativa*. In addition, C-II and C-III subfamilies were more closely related to

each other than to the other subfamilies. All sHSP genes corresponding to those of *Arabidopsis* were found in the Chinese cabbage genome (Figure 4).

A

Name	Strand	Start	*p*-value	Sites
Br15.9 II	+	105	3.45e-12	CAATTTCACA GAGAAAGAACGATCAAGACGCTTCTGGAA CTTCAAACTG
Br17.2B I	-	40	1.88e-11	GGGGATGATG AAGAGAGAACGTTCTTCGAGTTTCTTGAA TTAGAAATTG
Br17.6 II	+	85	2.17e-11	ATTTTTCACA AAGCCAGAACGATCAAGATGCCTCCCGAA CAACTCCAAA
Br17.3 II	+	104	7.38e-11	GTCAAAAAAA AAACTAGAACGATCAAGACACTTCTCGAA CAGTACGGAA
Br17.2A I	-	46	2.24e-10	GCGGAGGATG AAGAGAGAACGTTCTTTGAGTTTCTTGAA TTAGAAAGTG
Br18.0C I	-	47	2.83e-10	GTGGAGGATG AAAAGAGAACGTTCTTCGAGTTTCTTGAA TTAGAAAGTG
Br21.6 V	-	807	1.29e-09	ATGTTCTCTT GACGTAGAAATTGCTCGAAGGGTCACAAA AGAAGTCTGT
Br20.7 V	-	916	1.29e-09	TGTTCTCTTT GACGTAGAAATTGCTCGAAGGGTCACAAA GAAATGTGCA
Br17.7B I	-	63	1.43e-09	AGAGATAAGA GAGACAGAACGTTCTGCAGGTTTCTTTAA AAATCACGAT
Br17.7A I	-	102	1.92e-09	AAGATATGTG AATCCAGAAACTTCGAGGAGGTCCGAGGA GGATATAATC
Br18.0B I	-	91	2.57e-09	ATTCTCTCCC AACTTAGAAACATCTAGAAGATCCCAAGA CGAAAATATC
Br18.0A I	-	91	2.57e-09	ATTCTCTCCC AACTTAGAAACATCTAGAAGATCCCAAGA CGAAAATATC
Br15.5 IV	-	514	3.21e-08	GAGGGTACAG AAGAAAGAAACTGCAGGACCCATCTCAGT GTTAATCTTA
Br23.8 MT	+	105	4.96e-08	CTAAGCTTGG GAGCCACGCAGCTCTAGAACTATCCGGAA ACATGTTCTA
Br25.1 CP	+	620	7.50e-08	AGTGGTGTGA ATAAGAGGACGACCTAGACGCTTGGAGAT CTCTTTCTTA
Br21.8 MTII	-	277	8.02e-08	CTTTACTCTT GAGCTAGCAGTTCCTAGAAAGATCGAAAA TCTATTAATT
Br20.1 ER	+	100	8.57e-08	GCAACCTAAC GAGAGGAAAAAACCTAGAACGTTCTCAAT TTATCTCTCA
Br17.1	-	163	9.78e-08	AAAAAAAAAA AAGTTAGACCAATCGCGGGCCGCCACGAC ACGTGGAGCC

Motif1 E-value=3.1e-019



B

Name	Strand	Start	*p*-value	Sites
Br17.3 II	+	14	5.27e-11	AAACTCTTGT CTACATCTCTCTAGAACAATC CATCTTCACT
Br17.6 II	+	16	6.13e-09	TTCGGTCAGT CTACATCTCTCTAGAAAAATC CATCTCCACT
Br15.9 II	+	35	1.09e-08	TTCACTCTGT CTACATCTCTCTAGAATAATC TCTGCACTCC
Br22.0 MT	+	70	1.64e-08	TACTTGGCAA CCACGTACTTCTAGAACCATC CTTAACACTA
Br18.0 MT	+	25	1.64e-08	TACTTGCGGA CCACGTACTTCTAGAACCTTC TTCTATATAA
Br17.2A I	-	77	1.17e-07	TTAGAAAGTG CACACTTTTTCTAAAACCTTC GCTTATTGTC
Br23.8 MT	+	65	1.45e-07	CATGTTCTAT CCTCTACCCTCTGGAACATTG TCTAGCACCC
Br18.0C I	-	76	1.45e-07	AATTAGAAAG TGCACTTTTTCTAAAACCTTC GTTTATTTGT
Br17.2B I	-	69	1.45e-07	AATTAGAAAT TGCACTTTTTCTAAAACCTTC GTTTATGGTC
Br22.3 ER	+	61	2.24e-07	CTAGAAGAAT CCAAATTTCTCTACAAAAATC TATCTACCAC
Br17.1	-	519	4.58e-07	TGCACGTGAT CCCAAGCTCTCCCAAAGCATC AAATCTGTTA
Br25.1 CP	-	121	6.76e-07	CTCGCGTTTT CTCATCTTCTCCGGAACCATC TTTGTTTTCA
Br20.1 ER	-	41	6.76e-07	GTGTAGAATG TTCATACATTCTCGAACATTC ATGCATGTTT
Br16.4	-	837	8.16e-07	ATAGTGGATC ACGAAACTTTCTCCAACAATC GATTCTTCTC
Br15.5 IV	+	650	2.37e-06	GGTAGAGTGG CTAAAACCCACCAGCACATGC TTACTGATCT
Br17.7B I	+	316	3.30e-06	ACTTTTGAAG CGTAATTTTTCTAGAACACTA ATTTTAGTTT
Br16.0 PX	-	82	3.58e-06	TCTTGTTTTT TATCATCTATCCAGAACATTG GAGAAGACGA
Br18.0B I	-	555	4.55e-06	AAGAATATTC TCGCATTTTTCTTAAACAAAG TTACAGAATT
Br18.0A I	-	555	4.55e-06	AAGAATATTC TCGCATTTTTCTTAAACAAAG TTACAGAATT
Br17.7A I	-	724	4.92e-06	ATTCCTCTGT CCCCGACACTCCCAAGTATTC GTTTTTTTGT
Br17.2 III	+	69	6.68e-06	ATTATTGACA CGCGTAACCACTCGCACATTC TAGATTTATC
Br16.2	+	564	7.20e-06	TTTCTATGTT CTGCATCACTTCCACACACTC AAATGATCTT
Br20.7 V	+	76	1.04e-05	TTTTACTATT CCGTCTCTTACTTGCAAATTC GTTTTAGTAA
Br15.8 IV	-	312	3.25e-05	CACTGATGAC CGACCACTAAATAAAACCATG GTTCCCTACA

Motif5 E-value=1.4e-004

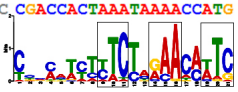


Figure 3. Sequence logos of Motif 1 (**A**) and Motif 5 (**B**) were produced by MEME using nucleotide sequences of the deduced 1000-bp upstream region of the sHSP promoter, relative to TSS. The "Start" indicates the distance from TSS. P values are indicated on the left side and the occurrence of motif is sorted by P value. "+" and "-" indicate sense and anti-sense strand, respectively. The sequences are boxed to show recognized consensus HSEs (5'-nGAAnnTTCnnGAAn-3' or 5'-nTTCnnGAAnnTTCn-3'). The E value is an estimate of the expected number of motifs with the given log likelihood ratio, and with the same width and site count, that one would find in a similarly sized set of random sequences. The heights of the symbols at each motif position indicate sequence conservation.

Table 3. Number of known *cis*-acting regulatory elements in 1000-bp region upstream of TSS in sHSP genes in Chinese cabbage.

Gene name	Number of motif			
	HSE (5'-nGAAn nTTCn-3' or 5'-nTTCnnGAAn-3')	DRE (RCCGAC)	ABRE (ACGTG)	LTRE (CCGAM)
Br17.2A I	4	0	1	1
Br17.2B I	4	1	1	3
Br17.7A I	1	3	0	3
Br17.7B I	5	1	1	2
Br18.0A I	2	0	0	0
Br18.0B I	2	0	0	0
Br18.0C I	4	0	1	2
Br15.9 II	1	1	1	2
Br17.6 II	3	0	0	1
Br17.3 II	1	0	0	1
Br17.2 III	1	0	0	0
Br15.5 IV	1	0	0	0
Br15.8 IV	0	0	0	0
Br21.6 V	1	0	0	0
Br20.7 V	0	0	0	0
Br16.0 PX	2	0	0	2
Br22.3 ER	1	0	1	0
Br20.1 ER	4	0	0	0
Br25.1 CP	2	0	0	2
Br23.8 MT	0	0	0	1
Br18.0 MT	2	0	1	1
Br22.0 MT	1	1	1	2
Br21.8 MTII	1	0	0	1
Br16.2	0	0	1	0
Br16.4	1	0	0	0
Br17.1	1	1	3	2
In total	45	8	12	26

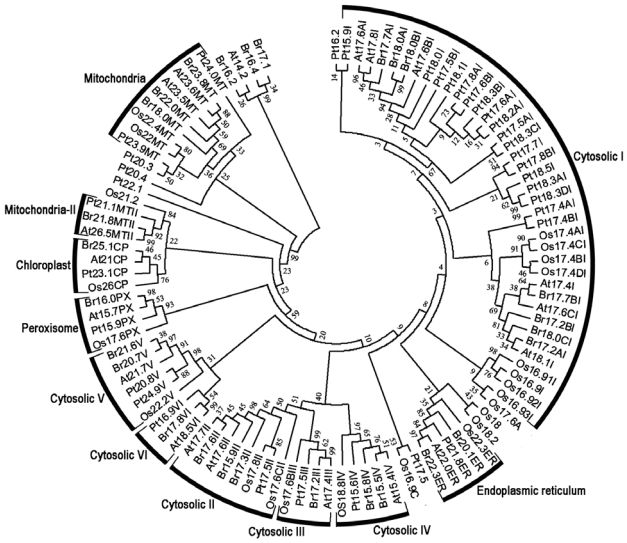


Figure 4. Neighbor-joining tree are derived by MEGA 5.2 with bootstrap analysis (1000 replicates) from alignment of amino acid sequence of sHSP proteins from *Brassica rapa* ssp *pekinensis*, *Arabidopsis thaliana*, *Populus trichocarpa*, and *Oryza sativa* [supplementary material of a recent publication (Zhang et al., 2014)]. The abbreviations of species are as follows: Br-*B. rapa* ssp *pekinensis*, At-*A. thaliana*, Pt-*P. trichocarpa*, Os-*O. sativa*.

There were three orphan sHSP genes in Chinese cabbage (*Br16.2*, *Br16.4*, and *Br17.1*), one in *A. thaliana* (*At14.2*), five in *P. trichocarpa* (*Pt16.2*, *Pt17.5*, *Pt20.3*, *Pt20.4*, and *Pt22.1*), and five in *O. sativa* (*Os16.9C*, *Os17.6A*, *Os18*, *Os18.2*, and *Os21.2*). Based on the phylogenetic tree, the orphan sHSP genes of Chinese cabbage and *A. thaliana* were more closely related to one another than to those of *P. trichocarpa* and *O. sativa*. The orphan sHSPs did not belong to any subfamily. Two orphan sHSPs of *P. trichocarpa* (*Pt20.3* and *Pt20.4*) were closely related to the MT subfamily, and *Pt16.2* was found to be closely related to the C-I subfamily. Two orphan sHSPs of *O. sativa* (*Os18* and *Os18.2*) were related to the ER subfamily, and *Os21.2* was associated with the MT-II subfamily.

Conserved domain analysis and sequence alignment of sHSP proteins in Chinese cabbage

To further check for conserved regions, motif analysis was carried out using the MEME web server. As a result, Motif 1 (β2, β3, and β4) was found to be conserved across all sHSP subfamilies of Chinese cabbage. Motif 5, Motif 6, Motif 10 (the helical regions), and Motif 9 (β10) were representative of the C-I sHSP subfamily, while Motif 8 was characteristic of the C-II sHSP subfamily. Motif 7 appeared exclusively in the orphan sHSPs. On the other hand, Motif 2 (β8 and β9) existed in all sHSP proteins except for the C-V sHSP subfamily (Figure 5A). Detailed sequence logos of the 10 motifs are shown in Figure 5B. Since the results of the MEME motif analysis did not precisely correspond to the β-strand region, we used Phyre2 combined with sequence alignment to predict secondary structures of the 27 sHSP proteins. This analysis indicated that while the sHSPs of Chinese cabbage did not show high conservation in amino acid sequences between subfamilies, they shared a highly conserved secondary structure (data not shown). The C-IV and C-VI families lacked the β6 strand, whereas the C-V family was missing the β10 strand. The "GVL" sequence in the β9 strand was absent in three C-V members, Br16.0PX, Br17.8VI, and Br17.1.

Gene structure

Chinese cabbage sHSP gene structures were analyzed using the genome data. Gene structures of 26 *B. rapa* sHSP genes are shown in Figure 6; since Br17.8VI was obtained from an EST library, its gene structure was not analyzed. The analysis indicated that the sHSP gene structure was conserved within each subfamily (Figure 6). C-I, C-II, and ER subfamily members of Chinese cabbage did not contain any introns, whereas the remaining subfamilies contained one intron. The intron of C-III subfamily members was located on the β4 strand, while the intron of C-IV, C-V, and PX subfamilies was located between β3 and β4 strands. The intron of CP, MT, and MT-II subfamilies and the three orphan genes was situated before the β2 strand. Comparison of sHSP orthologs of Chinese cabbage and *A. thaliana* revealed that *Br16.0 PX* of Chinese cabbage contained one intron whereas *At15.7 PX* did not contain any introns. The remaining sHSP genes of Chinese cabbage and *A. thaliana* shared the same gene organization.

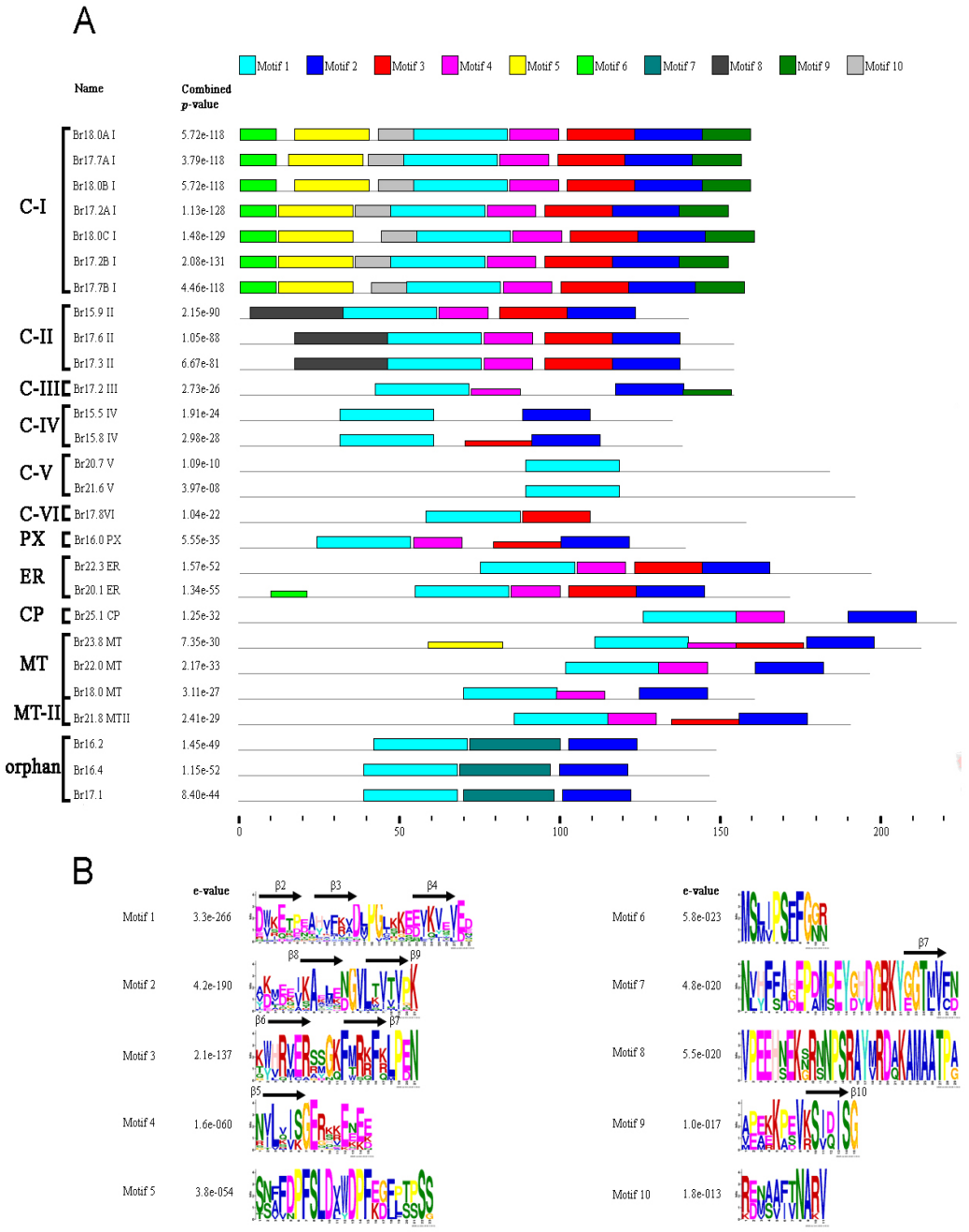


Figure 5. All motifs were identified by MEME using the complete amino acid sequences of 27 sHSPs of Chinese cabbage. **A.** Distribution of conserved motifs in the sHSP family members of Chinese cabbage. Combined P values are indicated on the left side. The height of the motif "block" is proportional to -log (P value), truncated at the height for a motif with a P value of $1e^{-10}$. Different motifs are shown by different colors numbered 1-10. **B.** Detailed sequence logos of the 10 motifs. Different β-strand region are marked by black arrowhead.

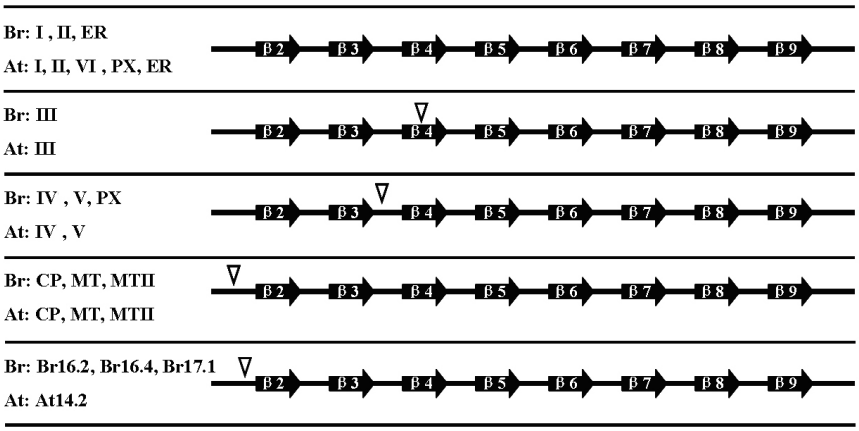


Figure 6. Gene structure of Chinese cabbage and *Arabidopsis* sHSP genes. Exon-intron organization of Chinese cabbage and *Arabidopsis* sHSP genes is depicted for different subfamilies. The eight β strands (β2-β9) are shown in black arrowhead. The intron positions are marked by inverted triangle. Cytosolic IV and VI sHSP proteins lack

Gene expression in different organs and at different embryo stages

High-resolution RNA sequencing data from different organs of Chinese cabbage have recently become available (Tong et al., 2013). We therefore downloaded expression data for all 26 sHSP genes in Chinese cabbage and produced a heat map to visualize gene expression (Figure 7). Gene sets were clustered using hierarchical clustering, and the displayed heat map was based on the transcript abundance pattern. *Br18.0A I*, *Br18.0B I*, and *Br17.1* expression was not detected in any organs, while the remaining 23 sHSP genes (88.5%) were found to be expressed in at least one organ. Nine sHSP genes (34.6%; *Br17.7A I*, *Br17.7B I*, *Br15.9 II*, *Br17.3 II*, *Br17.2 III*, *Br15.5 IV*, *Br20.7 V*, *Br16.0 PX*, and *Br22.0 MT*) were constitutively expressed in all organs. The expression patterns of the 26 sHSP genes were grouped into different classes. *Br18.0A I*, *Br18.0B I*, and *Br17.1*, which were not expressed in any organs, were clustered together. Two C-V sHSP genes (*Br20.7 V* and *Br21.6 V*) shared a similar expression pattern, with the highest expression detected in roots and the lowest in flowers. The highest expression levels for *Br15.5 IV* and *Br15.8 IV* were observed in roots, stems, and leaves. Low levels of *Br16.4* transcripts were detected in siliques. The remaining sHSP genes were clustered into one group; all of them displayed the highest expression in siliques.

In order to examine expression in embryos, expression data from different organs of *B. rapa* were downloaded from a recent publication (Zhang et al., 2014). All 22 sHSP genes for which data were available, were more or less expressed during at least one embryo stage. sHSP gene expression during globular and heart embryo stages was more similar than during the early cotyledon embryo phase (Figure 8). For most sHSP genes, their highest expression levels were observed in early cotyledon embryos, except for *Br17.2 III*, *Br16.2*, and *Br16.4*, whose expression levels were the highest in the heart embryo stage. In general, more sHSP genes were up-regulated in early cotyledon embryos than in globular or heart-shaped embryos. *Br17.7B I* showed distinct up-regulation during embryogenesis, and the expression levels in early cotyledon embryos were 128 and 283 times higher than that observed in globular and heart embryos, respectively.

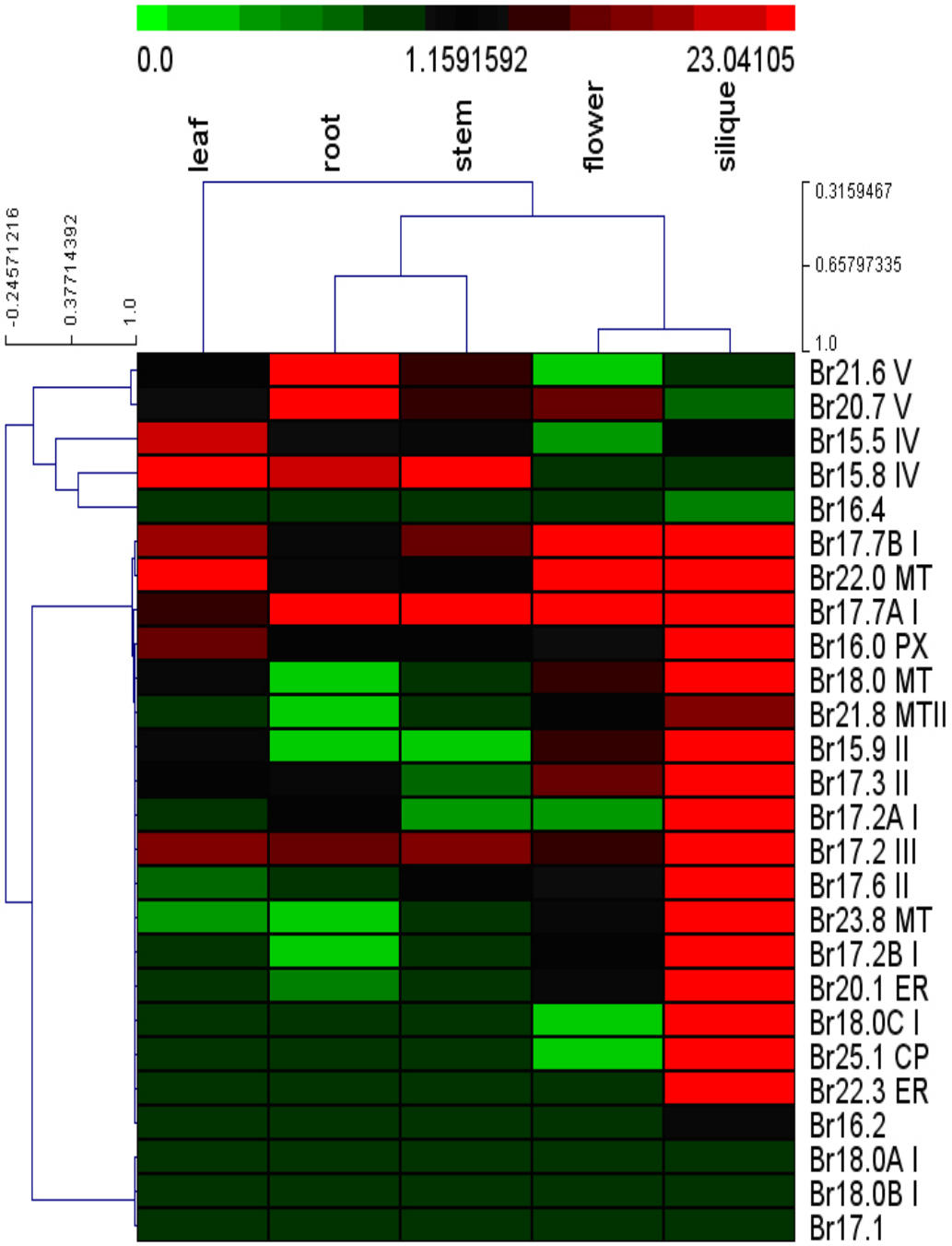


Figure 7. Expression profile cluster analysis of the 26 sHSP genes of Chinese cabbage. The expression values of each sHSP gene were downloaded from NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession No. GSE43245 for five organs: root, stem, leaf, flower, and silique.

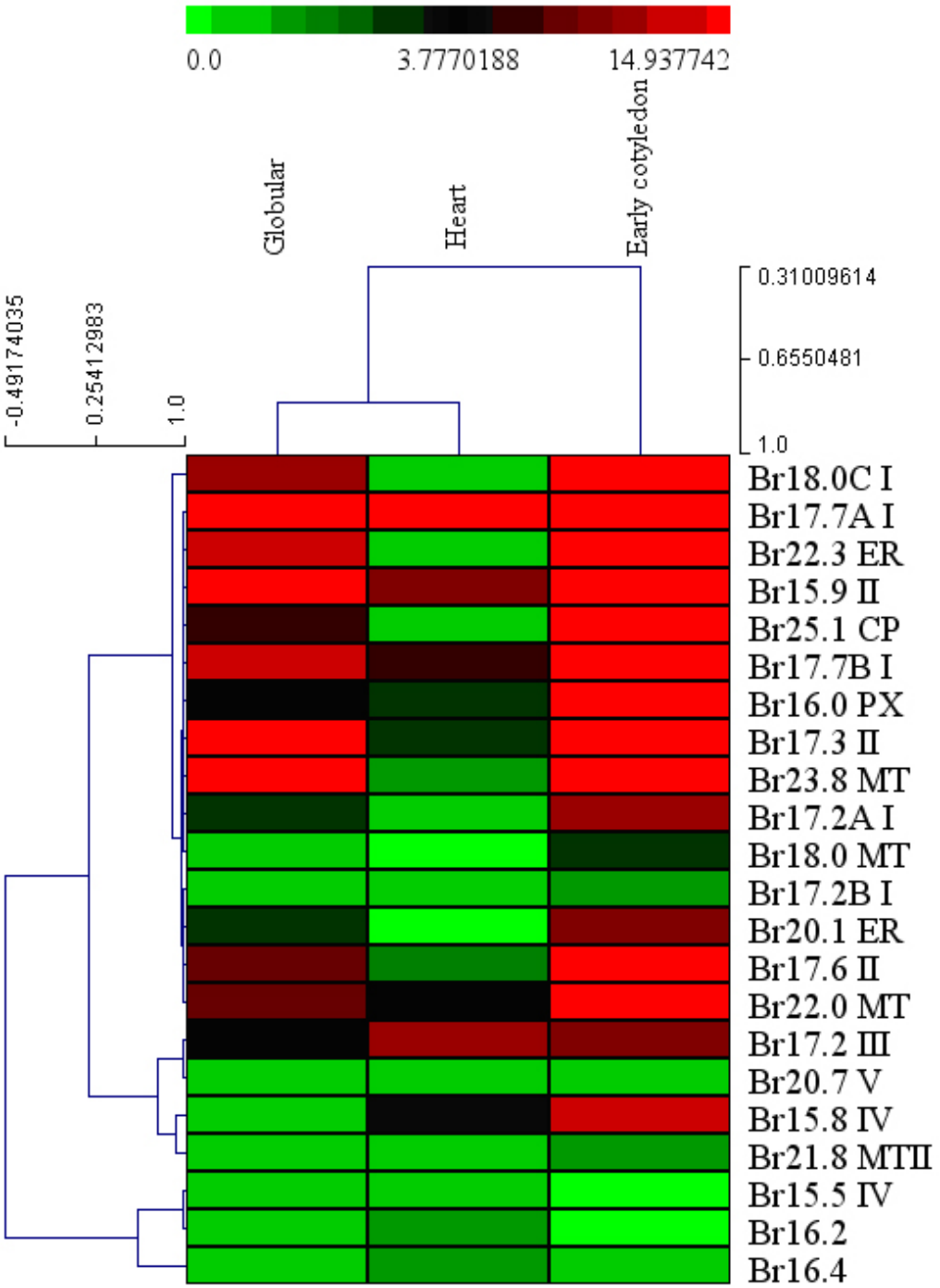


Figure 8. Expression profile cluster analysis of the 22 sHSP genes of Chinese cabbage. The expression values for each sHSP gene in globular, heart, and early cotyledon embryo stages were downloaded from Zhang et al. (2014).

DISCUSSION

In this study, we identified 27 sHSP genes in Chinese cabbage; in contrast, 19 sHSP genes are present in *Arabidopsis*. Thus, the number of sHSP family and subfamily members is higher in Chinese cabbage than in *Arabidopsis*, indicating that the Chinese cabbage genome has experienced gene duplication events during its evolution. A previous study has indicated that *B. rapa* is a paleohexaploid species containing three subgenomes that share the same diploid ancestor as *A. thaliana* (Wang et al., 2011; Cheng et al., 2013). However, the number of sHSP family members in *B. rapa* is less than three-fold than that in *Arabidopsis*. In addition, our colinearity comparison between *B. rapa* and *A. thaliana* genes revealed that most *A. thaliana* sHSP genes had one to three syntenic orthologs in *B. rapa* genomes (Table 2), and more sHSP genes were located on LF than on MF1 and MF2. These results are possible evidence of biased gene loss in the sHSP family of Chinese cabbage, with a whole-genome triplication of two-step origin. Similar results have been observed in other gene families, e.g., AQP (Tao et al., 2014). However, *Br18.0B I*, *Br20.1 ER*, and *Br16.2* had no syntenic orthologs in *A. thaliana*, while no syntenic orthologs of *At17.6A I*, *At17.6C I*, or *At17.8 I* were found in Chinese cabbage (Table 2). These results suggest that gene loss has occurred independently during the evolution of *A. thaliana* and Chinese cabbage. Genomic comparisons in a previous study have revealed that polyploid speciation has been accompanied by gene loss in many eukaryotes (Wang et al., 2011; Cheng et al., 2013).

Although the organization of sHSP genes from model plants such as *Arabidopsis*, soybean, *Populus*, and rice have been recently analyzed (Waters et al., 2008; Sarkar et al., 2009; Lopes-Caitar et al., 2013), the same has not been done for Chinese cabbage. Hence, we compared the structure of sHSP genes from the same sHSP subfamily between Chinese cabbage and *Arabidopsis* (Figure 6). Most sHSP genes within each subfamily were generally conserved, both in the number and position of introns across the two species. Nevertheless, a slight discrepancy was observed in the structure of sHSP genes between *Arabidopsis* and Chinese cabbage. No intron existed in *At15.7 PX* of *Arabidopsis* or *Os17.6 PX* of rice (Waters et al., 2008; Sarkar et al., 2009), whereas the corresponding orthologous *Br16.0 PX* gene of Chinese cabbage had an intron between β3 and β4 strands. This suggests that either *At15.7 PX* of *Arabidopsis* and rice has lost an intron, or that intron gain has occurred independently in *Br16.0 PX* genes of Chinese cabbage. Similarly, CP sHSP genes of *Arabidopsis*, *P. trichocarpa*, and Chinese cabbage had one intron, whereas the CP gene (Os26 CP) of rice had none (Sarkar et al., 2009). A previous study has indicated that low intron gain rates and intron number reduction are common features of plant evolution (Roy and Penny, 2007).

With regard to sHSP gene expression in organs and during embryonic development, most sHSP genes in Chinese cabbage exhibited the highest expression in siliques (Figure 7), and in early cotyledon embryos (Figure 8). Comparative analysis of the expression pattern of paralogous genes in Chinese cabbage indicated that most of these genes were clustered together, suggesting that the paralogous genes in each group play similar roles in Chinese cabbage. However, paralogous sHSP genes showed different expression levels; for example, the transcriptional level of *Br22.0 MT* in different organs and across embryonic stages was at least three times higher than that of *Br18.0 MT*. Tao et al. (2012) compared the expression of two paralogous C-II sHSP genes (*CbHSPCIIa* and *CbHSPCIIb*) and likewise found that the expression of *CbHSPCIIa* was approximately five times higher than that of *CbHSPCIIb* in different organs and across embryonic stages. *Br18.0A I*, *Br18.0B I*, and *Br17.1* were not

expressed in any organs. Similarly, the corresponding orthologous sHSP genes of *Arabidopsis* (AT1G59860, At2G29500, and At5G47600) showed low expression in organs. With respect to organ expression, Chinese cabbage C-I and C-II sHSP genes (*Br17.2A I*, *Br17.2B I*, *Br17.7A I*, *Br17.7B I*, and *Br18.0C I*) showed the highest expression in siliques; these genes were more or less expressed during the globular embryo stage and reached their highest expression in early cotyledons. Dafny-Yelin et al. (2008) have suggested that C-I sHSP proteins have a redundant function in *Arabidopsis* early embryogenesis. *Br17.7B I*, corresponding to *At17.4 I* (AT3G46230), showed strong up-regulation, and its expression in early cotyledon embryos was 120-fold higher than that in globular or heart-shaped embryos. In *Arabidopsis*, *At17.4 I* shows strong up-regulation during embryo development via regulation by HSFA9 (Kotak et al., 2007). The *C-II 17.7* sHSP gene may increase salt and drought tolerance in *Arabidopsis* (Sun et al., 2001). *NnHSP17.5*, a C-II sHSP of *Nelumbo nucifera*, has been found to be expressed in seeds under normal conditions, and shows strong up-regulation in germinating seeds subjected to heat and oxidative stresses. In contrast, we found C-IV and C-V sHSP genes of Chinese cabbage to be dominantly expressed in vegetative organs, while showing little or no expression in flowers and siliques. During embryo development, Br*15.5 IV* and *Br20.7 V* were weakly expressed at the three-embryo stage, whereas *Br15.8 IV* was up-regulated during embryo development. The corresponding sHSP genes of *Arabidopsis* (*At21.7 V*) are expressed constitutively in vegetative organs (Siddique et al., 2008). All organellar sHSP genes of Chinese cabbage showed dominant expression in siliques and in early cotyledon embryos. In *Chenopodium album*, CP-sHSPs play an essential role in the protection of photosystem II and thylakoid membranes under a variety of abiotic stresses (Haq et al., 2013). These proteins can also interact with plastid nucleoid protein pTAC5 and are essential for chloroplast development in *Arabidopsis* under heat stress (Zhong et al., 2013). A heat-inducible ER-sHSP has been demonstrated to perform a molecular chaperone function *in vitro* (Mamedov and Shono, 2008). Finally, Huther et al. (2013) have verified that MT-sHSP23.6 is involved in heat tolerance. In their study, transgenic plants suffered greater physiological damage during heat stress as a consequence of MT-sHSP23.6 suppression.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31301788 and #31372058), the Grand Science and Technology Special Project of Zhejiang Province (#2012C12903-3-8), and the China Postdoctoral Science Foundation (#2013M540500).

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Bailey TL, Williams N, Misleh C and Li WW (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34: 369-373.

Basha E, O'Neill H and Vierling E (2012). Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends Biochem. Sci.* 37: 106-117.

Cheng F, Wu J, Fang L and Wang X (2012). Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Front. Plant Sci.* 3: 198.

Cheng F, Mandáková T, Wu J, Xie Q, et al. (2013). Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell* 25: 1541-1554.

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

Haq NU, Ammar M, Bano A, Luthe DS, et al. (2013). Molecular characterization of *Chenopodium album* chloroplast small heat shock protein and its expression in response to different abiotic stresses. *Plant Mol. Biol. Rep.* 31: 1230-1241.

Heidarvand L and Amiri RM (2010). What happens in plant molecular responses to cold stress? *Acta Physiol. Plant.* 32: 419-431.

Huang GT, Ma SL, Bai LP, Zhang L, et al. (2012). Signal transduction during cold, salt, and drought stresses in plants. *Mol. Biol. Rep.* 39: 969-987.

Huther CM, Ramm A, Rombaldi CV and Bacarin MA (2013). Physiological response to heat stress of tomato 'Micro-Tom' plants expressing high and low levels of mitochondrial sHSP23.6 protein. *Plant Growth Regul.* 70: 175-185.

Kim R, Kim KK, Yokota H and Kim SH (1998). Small heat shock protein of *Methanococcus jannaschii*, a hyperthermophile. *Proc. Natl. Acad. Sci. U. S. A.* 95: 9129-9133.

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

Lopes-Caitar VS, de Carvalho MC, Darben LM, Kuwahara MK, et al. (2013). Genome-wide analysis of the Hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. *BMC Genomics* 14: 577.

Mamedov TG and Shono M (2008). Molecular chaperone activity of tomato (*Lycopersicon esculentum*) endoplasmic reticulum-located small heat shock protein. *J. Plant Res.* 121: 235-243.

Nakashima K, Fujita Y, Katsura K, Maruyama K, et al. (2006). Transcriptional regulation of ABI3- and ABA-responsive genes including *RD29B* and *RD29A* in seeds, germinating embryos, and seedlings of *Arabidopsis*. *Plant Mol. Biol.* 60: 51-68.

Roy SW and Penny D (2007). Patterns of intron loss and gain in plants: intron loss-dominated evolution and genome-wide comparison of *O. Sativa* and *A. thaliana*. *Mol. Biol. Evol.* 24: 171-181.

Saeed AI, Sharov V, White J, Li J, et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34: 374-378.

Sarkar NK, Kim YK and Grover A (2009). Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics* 10: 393.

Scharf KD, Siddique M and Vierling E (2001). 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.

Schöffl F, Prandl R and Reindl A (1998). Regulation of the heat-shock response. *Plant Physiol.* 117: 1135-1141.

Siddique M, Gernhard S, von Koskull-Döring P, Vierling E, et al. (2008). The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. *Cell Stress Chaperon.* 13: 183-197.

Sun W, Bernard C, van de Cotte B, Montagu MV, et al. (2001). *At-HSP17.6A*, encoding a small heat-shock protein in *Arabidopsis*, can enhance osmotolerance upon overexpression. *Plant J.* 27: 407-415.

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

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

Sundby C, Härndahl U, Gustavsson N, Åhrman E, et al. (2005). Conserved methionines in chloroplasts. *Biochim. Biophys. Acta* 1703: 191-202.

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.

Tang H, Woodhouse MR, Cheng F, Schnable JC, et al. (2012). Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190: 1563-1574.

Tao P, Liu L and Wang J (2012). Characterization of eight cytosolic sHSP genes and their expression in *Capsella bursa-pastoris*. *Biol. Plant.* 56: 648-656.

Tao P, Zhong X, Li B, Wang W, et al. (2014). Genome-wide identification and characterization of aquaporin genes (*AQP*s) in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Mol. Genet. Genomics* 289: 1131-1145.

Tong C, Wang X, Yu J, Wu J, et al. (2013). Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. *BMC Genomics* 14: 689.

van Montfort RL, Basha E, Friedrich KL, Slingsby C, et al. (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat. Struct. Biol.* 8: 1025-1030.

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

Wang X, Wang H, Wang J, Sun R, et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43: 1035-1039.

Waters ER (2013). The evolution, function, structure, and expression of the plant sHSPs. *J. Exp. Bot.* 64: 391-403.

Waters ER, Aevermann BD and Sanders-Reed Z (2008). 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.

Yi S and Liu J (2009). Combinatorial interactions of two cis-acting elements, AT-Rich regions and HSEs, in the expression of tomato *Lehsp23.8* upon heat and non-heat stresses. *J. Plant Biol.* 52: 560-568.

Yi SH, Sun AQ, Sun Y, Yang JY, et al. (2006). Differential regulation of *Lehsp23.8* in tomato plants: Analysis of a multiple stress-inducible promoter. *Plant Sci.* 171: 398-407.

Zhang Y, Peng L, Wu Y, Shen Y, et al. (2014). Analysis of global gene expression profiles to identify differentially expressed genes critical for embryo development in *Brassica rapa*. *Plant Mol. Biol.* 86: 425-442.

Zhong L, Zhou W, Wang H, Ding S, et al. (2013). Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in *Arabidopsis* under heat stress. *Plant Cell* 25: 2925-2943.