Plant Mol Biol
DOI 10.1007/s11103-009-9477-y

Comprehensive sequence and expression profile analysis of *Hsp20* gene family in rice

Yidan Ouyang · Jiongjiong Chen · Weibo Xie · Lei Wang · Qifa Zhang

Received: 23 October 2008 / Accepted: 23 February 2009

Abstract The *Hsp20* genes represent the most abundant small heat shock proteins (sHSPs) in plants. *Hsp20* gene family has been shown to be involved in preventing heat shock and promoting resistance to environmental stress factors, but very little is known about this gene family in rice. Here, we report the identification and characterization of 39 *OsHsp20* genes in rice, describing the gene structure, gene expression, genome localization, and phylogenetic relationship of each member. We have used RT-PCR to perform a characterization of the normal and heat shock-induced expression of selective *OsHsp20* genes. A genome-wide microarray based gene expression analysis involving 25 stages of vegetative and reproductive development in three rice cultivars has revealed that 36 *OsHsp20* genes were expressed in at least one of the experimental stages studied. Among these, transcripts of *OsHsp20* were accumulated differentially during vegetative and reproductive developmental stages and preferentially down-regulated in Shanyou 63. In addition, *OsHsp20* genes were identified as showing prominent heterosis in family-level expression. Our results suggest that the expression patterns of the *OsHsp20* genes are diversified not only in developmental stages but also in variety level.

Keywords *Oryza sativa* · Heat shock · Expression profiles · Microarray · *Hsp20* gene family

Electronic supplementary material The online version of this article (doi:10.1007/s11103-009-9477-y) contains supplementary material, which is available to authorized users.

Y. Ouyang (✉) · J. Chen · W. Xie · L. Wang · Q. Zhang
National Key Laboratory of Crop Genetic Improvement, National Centre of Plant Gene Research, Huazhong Agricultural University, 430070 Wuhan, China
e-mail: diana1983941@webmail.hzau.edu.cn

Abbreviations

HSP	Heat shock protein
sHSP	Small HSP
ACD	α-Crystallin domain
Os	*Oryza sativa*
At	*Arabidopsis thaliana*
ER	Endoplasmic reticulum
HSE	Heat shock response element
Hsf	Heat shock factor
bp	Base pair
aa	Amino acids

Introduction

All organisms respond to high temperature by synthesizing a group of distinct polypeptides called heat shock proteins (HSPs) (Lindquist and Craig 1988). HSPs play a central role not only in the protection against stress damage, but also in the folding, intracellular distribution, and degradation of proteins, as well as in the signal transduction chains. Based on sequence homology, HSPs in eukaryotes are classified into families of Hsp100, Hsp90, Hsp70, Hsp60 and small HSPs (sHSPs) with molecular size ranging from 15 to 30 kDa (Vierling 1991). Small HSPs represent the major family of HSPs induced by heat stress in plants (Waters et al. 1996). Among many higher plants, sHSPs are the most abundant proteins produced under heat stress and have been reported to accumulate up to as much as 1% of the total protein content of soybean (Hsieh et al. 1991). Plant sHSPs appear to have evolved independently after the divergence of plants and animals (de Jong et al. 1998). All plant sHSPs encoded by nuclear genes are divided into at least six classes, based on localization in different cellular compartments, amino acid sequence homology, and

Published online: 10 March 2009

immunological cross-reactivity (Waters et al. 1996). Three classes (Classes C I, C II and C III) of sHSPs are localized in the cytosol or nucleus (Scharf et al. 2001; Vierling 1991), while the C I gene family is generally the most numerous in plants. Other three classes are localized in the plastid (P) (Vierling 1991), endoplasmic reticulum (ER) (Helm et al. 1995) and mitochondria (M), respectively (Lenne et al. 1995; LaFayette et al. 1996). In addition to these six classes, another subfamily (Po) was identified initially based on the presence of a putative peroxisomal Type 1 targeting signal (PTS1; Scharf et al. 2001) and was subsequently shown to be localized to peroxisomes dependent on the functional C-terminal PTS1 (Ma et al. 2006). Recently, five new subfamilies of sHSPs, C IV, C V, C VI, C VII and M II, were defined in addition to the previously known seven subfamilies, including four additional cytoplasmic/nuclear sHSPs and a second unique family of sHSPs targeted to the mitochondria (Siddique et al. 2008). These 12 subfamilies of plant sHSPs (C I, C II, C III, C IV, C V, C VI, C VII, ER, P, M I, M II and Po) are to date the most integrated classification of the plant sHSPs.

Among all of them, the most abundant and complex member in sHSPs in higher plants is the Hsp20 family (Schöffl and Key 1982; Vierling 1991). They are encoded by members of a multi-gene family in eukaryotes and defined by possessing a conserved carboxyl-terminal domain of approximately 90 amino acid residues called α-crystallin domain (ACD) (Scharf et al. 2001; Vierling 1991). The ACD is flanked by an N-terminal domain and a short C-terminal extension (Caspers et al. 1995; Waters et al. 1996; MacRae 2000). This domain is characterized by a compact β-strand structure and forms a conserved secondary structural despite large levels of sequence diversity (Kim et al. 1998; van Montfort et al. 2001a; Stamler et al. 2005). Another characteristic feature of the sHSPs is that, with few exceptions, these proteins function as multimeric complexes of 8–24 or more subunits (van Montfort et al. 2001a; Stamler et al. 2005).

sHSPs are considered to protect thermo-sensitive substrates from irreversible heat-stress induced denaturation and aggregation. Several researches have indicated the importance of sHSPs in protecting against heat-induced denaturation (Takemoto et al. 1993). To date, the presence and accumulation of sHSPs under stress conditions have been shown to have protective functions that confer thermo-tolerance in different cell types (Landry et al. 1989; Yeh et al. 1997). Biochemical and structural studies have demonstrated that the chaperone activity of sHSPs functioned as a reservoir for the intermediates of denatured proteins, preventing proteins from aggregation caused by heat damages (van Montfort et al. 2001b).

While the functions of Hsp20 proteins presumably have their evolutionary roots in chaperones, many have additional functions. It has been recognized that the Hsp20 proteins have protective roles against a variety of stresses besides high temperature. These HSPs are a group of proteins induced to protect plants from the damage caused by stress (Coca et al. 1994; Kiyosue et al. 1994; Swindell et al. 2007). In addition to heat, stresses caused by salt shock (Harrington and Alm 1988), alcohol (Kuo et al. 2000), amino acid analogs (Lee et al. 1996), chilling (van Berkel et al. 1994; Sabehat et al. 1996, 1998; Sato et al. 2001), oxidative injury (Banzet et al. 1998), drought (Sato and Yokoya 2008) and heavy metals (Lin et al. 1984; Edelman et al. 1988; Tseng et al. 1993; Wollgiehn and Neumann 1995; Guan et al. 2004) also induce expression of one subset of sHSPs or Hsp20 proteins. It is thought that Hsp20 proteins play an important role in plant adaptation to various environmental stress conditions. Recent microarray studies in *Arabidopsis* also revealed that a subset of sHSP genes were induced by various stresses such as salt, drought, chilling, oxidative stress, and wounding (Desikan et al. 2001; Cheong et al. 2002). Moreover, members of the sHSP gene families are also developmentally regulated in seeds, storage organs, and vegetative tissues in plants (Wehmeyer and Vierling 2000; Lubaretz and zur Nieden 2002 Jofré et al. 2003; Sachin et al. 2007).

Rice is the main staple food for a large segment of the world population and the productivity of rice is restricted by a broad range of environmental conditions (Zhang 2007). It is sensitive to heat stress at all stages of development (Maestri et al. 2002). Rice is also an ideal model plant to analyze gene expression and function (Zhang et al. 2008). The completed genome sequences of rice (*Oryza sativa*) constitute a valuable resource for comparative genomic analysis, as they are representatives of the major evolutionary lineages within the monocotyledons. Study of the *OsHsp20* gene family is important for understanding the mechanisms by which plants respond to stress and other biological processes. In this study, an in-depth analysis of *OsHsp20* genes in rice has been investigated through the genome-wide scan and a systematic characterization of the entire *OsHsp20* gene family has been performed during different developmental stages. A total of 39 *OsHsp20* genes were identified in rice. We have used RT-PCR to perform a characterization of the normal and heat shock-induced expression of selective *OsHsp20* genes. In addition, we have also used microarray studies to demonstrate the systematic characterization of the expression of the entire gene family in three rice cultivars in order to better obtain the information of each of the family members during development. To our knowledge, this is the first report focusing on the *OsHsp20* genes in identification of family-level expression patterns. These results provide a solid base for future functional genomic studies of the HSP gene family in rice.

Materials and methods

Identification of Hsp20 homologues in rice

Name search and Hidden Markov Model (HMM) profile of Hsp20 domain (PF00011) downloaded from PFam (http://pfam.sanger.ac.uk/) were employed to identify the *OsHsp20* genes from rice genome. The BlastP search was performed using the HMM profile in TIGR (http://www.tigr.org/), KOME (http://cdna01.dna.affrc.go.jp/cDNA/) and NCBI (http://www.ncbi.nlm.nih.gov/) databases, followed by removal of redundant sequences from the three databases. Name search using Hsp20 as a keyword in these databases was applied for identification of more genes which could be missed using HMM profile due to the presence of incomplete Hsp20 domain. The PFam (http://www.sanger.ac.uk/Software/Pfam/) and SMART database (http://smart.embl-heidelberg.de/smart/batch.pl) were finally used to confirm each predicted Hsp20 protein sequence as the Hsp20 protein family member, sharing a common domain.

The same method was used to search against the TAIR database (http://www.arabidopsis.org/index.jsp) to get the putative Hsp20 homologues in *Arabidopsis*.

Chromosomal localization and gene duplication

OsHsp20 genes were mapped on chromosomes by identifying their chromosomal position given in the TIGR rice database.

The duplicated genes were elucidated from the segmental genome duplication of rice (http://www.tigr.org/tdb/e2k1/osa1/segmental_dup/500). The DAGchainer program (Haas et al. 2004) had been used to determine segmental duplications with parameters V = 5 B = 5 E = 1e-10-filter seg and distance = 500 kb. Genes separated by five or fewer genes were considered to be tandem duplicates. The distance between these genes on the chromosomes was calculated and the percentage of sequence similarity between the proteins encoded by these genes was determined by MegAlign software 4.0.

Structural and protein sequence analysis of the *OsHsp20* genes

Information regarding ORF length and intron numbers was downloaded from TIGR release 5.

Protein sequences of putative OsHsp20 members collected from the TIGR, KOME and NCBI were analyzed by EXPASY PROTOPARAM tool (http://www.expasy.org/tools/protparam.html). Information about the number of amino acids, molecular weight, theoretical isoelectric point (pI), amino acid composition, and instability index [instability index of >40 was considered as unstable (Guruprasad et al. 1990)] were obtained by this tool.

The conserved domain of the OsHsp20 protein in rice and *Arabidopsis* was determined by Pfam program. Protein sequences were analyzed in the MEME program (http://meme.sdsc.edu/meme/cgi-bin/meme.cgi) to confirm the conserved motifs. The MEME program was employed using the following parameters: number of repetitions—any, maximum number of motifs—10, optimum motif width set to >6 and <200. The motifs obtained were annotated using InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/) search program.

Phylogenetic analysis of OsHsp20 family

Crystallin domain identified by SMART and PFam from the OsHsp20 sequences were aligned using ClustalX (version 1.83) program. An un-rooted neighbor-joining (Saitou and Nei 1987) phylogenetic tree was constructed in ClustalX based on the full sequence of the proteins with default parameters from rice and *Arabidopsis*. Bootstrap analysis was performed using 1,000 replicates. The trees thus obtained were viewed using MEGA 4 software.

Expression analysis of *OsHsp20* genes under heat shock

Rice seeds (Minghui 63) were imbibed for three days in water, sown in moist paper and germinated in the dark at 25°C. Three-day-old rice seedlings were subjected to 37 and 42°C treatments for 2 h. Control seedlings were exposed to 25°C. Samples were ground in liquid nitrogen using a mortar and pestle. Total RNA (1 μg) was isolated using a RNA extraction kit (TRIzol reagent, Invitrogen) and reverse-transcribed in 20 μl reaction using DNase I and SuperScript™ II (Invitrogen) according to the manufacturer's instruction. Primers were designed according to the sequences within the boundaries defined by the cDNA ends (Supplementary Table 1) for PCR amplification of the reverse-transcription products. The PCR reaction was in a 20 μl volume containing 1 μl template. Reactions were performed with rTaq polymerase (Takara Biotechnology, Japan) on Gene AMP PCR system 9700 (Applied Biosystem, USA), with the following profile: 3 min at 94°C for pre-denaturation, followed by 28 cycles of 45 s at 94°C, 45 s at 58°C, and 45 s at 72°C, and a final 5 min extension at 72°C. Rice β-actin gene cDNA fragment was used as positive controls for each gene investigated. All experiments were repeated for three times independently.

Genome-wide expression analysis of *OsHsp20* family in three rice cultivars

Expression profile data of *OsHsp20* gene family in 25 tissues for Minghui 63, Zhenshan 97 and Shanyou 63 were extracted from CREP database (http://crep.ncpgr.cn). Massively parallel signature sequencing (MPSS) data (http://mpss.udel.edu/rice/) was consulted to determine the expression profiles of genes that were not represented in CREP database. Expression values of each gene were logarithmized and cluster analyses were performed using R with euclidean distances and hierarchical cluster method of "complete linkage".

For data analysis, a student-*t* test was performed to identify differentially expressed genes. The genes that are up- or down-regulated by more than two-fold and with *P* values <0.05 were considered to be differentially expressed. The average expression of more than two biological replicates for each sample was used for analysis. Mid-parent heterosis of the gene expression level was calculated as: $(F_1-MP)/MP \times 100$, in which F_1 was the expression level of the hybrid and *MP* was the average of the two parents. An *h*-statistic was devised to test statistical significance of mid-parent heterosis of gene expression: $[F_1-(P_1 + P_2)/2]/\sqrt{V_{F1} + (V_{P1} + V_{P2})/4}$ (Huang et al. 2006), which was similar to the *t*-statistic with the degrees of freedom contributed by the three genotypes, F_1, P_1 and P_2, V_{F1}, V_{P1} and V_{P2} represented the variances of the hybrid and the two parents, respectively.

Results

Identification of *OsHsp20* homolog in rice

HMM analysis and name search resulted in a total of 50 sequences from TIGR pseudomolecule version 5, out of which 42 were unique. By the same method, 50 and 35 sequences were obtained from KOME and NCBI, of which five and eight did not find any significant matches in the TIGR pseudomolecule database, respectively. The PFam and SMART scans of these sequences confirmed the presence of the Hsp20 domain in 39 of them. For convenience, the 39 *OsHsp20* genes were named *OsHsp20-1* to *OsHsp20-39* according to their positions on pseudomolecules (Table 1). Details on accession number of *OsHsp20* genes were listed in Table 1. Meanwhile, similar searches identified 29 *AtHsp20* members in *Arabidopsis*, and additional two *AtHsp20* genes were identified according to the previously published work (Siddique et al. 2008). These *AtHsp20* genes were named according to their positions on pseudomolecules (Supplementary Table 2).

Intron position and protein analysis of the *OsHsp20* genes

The presence of an existing cDNA with a known corresponding position on the genome was required for the determination of the number and length of introns. This information was available in TIGR release 5 (Table 1). It was found that 19 (48.72%) of the total *OsHsp20* genes were intronless. This is much higher than the average percentage (19.9%) of rice genes predicted to be intronless (Jain et al. 2008). Among those having introns, 17 (43.59%) genes had a single intron while only three genes had more than one intron. On the basis of intron presence, genes could be divided into three patterns by the number of introns. Patterns 1 and 2 were the most prevalent. Pattern 1 had no intron whereas pattern 2 had one intron. Pattern 3 was uncommon and found in three *OsHsp20* genes only. The overall pattern of intron position acts as an index to the phylogenetic relationships in a gene family, and this will be further illustrated in phylogenetic analysis section.

Pfam analysis suggested that protein sequences of the *OsHsp20* family had a typical single OsHsp20 domain. The length of OsHsp20 proteins varied from 146 to 486 amino acids. EXPASY analysis suggested that the OsHsp20 protein sequences had large variations in isoelectric point (pI) values (ranging from 4.52 to 10.17) and molecular weight (ranging from 16.0 kDa to 52.9 kDa) (Supplementary Table 3). Only eight of the 39 *OsHsp20* genes were predicted to be stable proteins, while the rest were unstable. Details on other parameters of protein sequences were shown in Supplementary Table 3.

Structure and phylogenetic analysis of Hsp20 family

The MEME motif search tool was employed to identify the conserved motifs in Hsp20 proteins and ten distinct motifs were identified (Fig. 1a). The details of the ten putative motifs are referred in Table 2. The locations of the ten motifs matched well with the conserved regions revealed by multiple sequence alignment analysis. Motif 1, 2, 3 or 10 specifying the Hsp20 domain were found in most of the members of the *Hsp20* family both in rice and *Arabidopsis*. Other unknown motifs were also revealed by MEME motif search. Motif 4 corresponding to the intervening region was found behind the Hsp20 conserved domain (motif 2), whereas motif 8 was found to be distributed in the C-terminal regions next to Hsp20 conserved domain (motif 1). The conserved motif 9 was found mainly in the N-terminal regions and motifs 5 and 6 were found mainly in the C-terminal regions. The results suggested that these motifs were conserved in rice and *Arabidopsis*. Although the functions of these motifs are not yet clear, the presence of these conserved motifs certainly reflects functional

Table 1 List of 39 *OsHsp20* genes identified in rice and their sequence characteristics

No.	Name	Accession Number		Chr. No.	TIGR v5 pseudomolecule position	ORF (bp)	Introns
		TIGR	KOME				
1	OsHsp20-1	LOC_Os01g04340	AK063681	1	1931049–1931501	453	0
2	OsHsp20-2	LOC_Os01g04350	AK119243, AK119717, AK069547	1	1937951–1938451	501	0
3	OsHsp20-3	LOC_Os01g04360	/	1	1941724–1941275	450	0
4	OsHsp20-4	LOC_Os01g04370	/	1	1946259–1945807	453	0
5	OsHsp20-5	LOC_Os01g04380	AK121025	1	1948849–1949301	453	0
6	OsHsp20-6	LOC_Os01g08860	AK071240	1	4446592–4446092	501	0
7	OsHsp20-7	LOC_Os01g40530	/	1	23225691–23224329	522	1
8	OsHsp20-8	LOC_Os01g40550	AK109917	1	23228962–23228014	810	1
9	OsHsp20-9	LOC_Os02g03570	/	2	1450233–1450766	534	0
10	OsHsp20-10	LOC_Os02g10710	AK106682	2	5640360–5639315	660	1
11	OsHsp20-11	LOC_Os02g12610	/	2	6588065–6587538	528	0
12	OsHsp20-12	LOC_Os02g48140	AK107963	2	29462598–29463092	495	0
13	OsHsp20-13	LOC_Os02g48370	AK107036	2	29610261–29605862	1461	12
14	OsHsp20-14	LOC_Os02g52150	AK105464	2	31915052–31915819	663	1
15	OsHsp20-15	LOC_Os02g54140	AK119261	2	33173673–33174279	519	1
16	OsHsp20-16	LOC_Os03g06170	/	3	3075991–3075141	639	1
17	OsHsp20-17	LOC_Os03g14180	AK063618, AK120048, AK120045, AB020973	3	7676185–7676907	723	0
18	OsHsp20-18	LOC_Os03g15960	AK104129, AK119664, AK119616, AK119239, AK119675, AK073671, AB110191	3	8784837–8785322	486	0
19	OsHsp20-19	LOC_Os03g16020	AK119243, AK119717, AK069547	3	8813430–8812966	465	0
20	OsHsp20-20	LOC_Os03g16030	/	3	8813998–8814483	486	0
21	OsHsp20-21	LOC_Os03g16040	AK120655, AK069547	3	8817563–8817084	480	0
22	OsHsp20-22	LOC_Os03g45330	/	3	25550411–25546770	1149	4
23	OsHsp20-23	LOC_Os03g45340	AK062338	3	25554893–25553918	879	1
24	OsHsp20-24	LOC_Os03g61940	/	3	35053234–35052695	540	0
25	OsHsp20-25	LOC_Os04g36750	AK063700	4	21982033–21982680	648	0
26	OsHsp20-26	LOC_Os05g23140	AK063686	5	13121792–13114968	756	1
27	OsHsp20-27	LOC_Os05g42120	AK110627	5	24544068–24545554	612	1
28	OsHsp20-28	LOC_Os05g51440	/	5	29411340–29412432	945	1
29	OsHsp20-29	LOC_Os06g11610	/	6	6152222–6151339	747	1
30	OsHsp20-30	LOC_Os06g14240	AK105317	6	7940609–7940169	441	0
31	OsHsp20-31	LOC_Os06g41730	AK101550, AK073058, AK099075, AK064549, AK071559.1	6	25013207–25017963	1386	13
32	OsHsp20-32	LOC_Os07g33350	AK071773, AK099296, AK063798	7	19933883–19934489	522	1
33	OsHsp20-33	LOC_Os09g17660	/	9	10800077–10801248	603	1
34	OsHsp20-34	LOC_Os10g07200	AK064267	10	3765580–3763974	1074	1
35	OsHsp20-35	LOC_Os10g07210	AK107162	10	3771437–3770106	594	1
36	OsHsp20-36	LOC_Os10g30162	AK062774	10	15352247–15353949	492	1

Table 1 continued

No.	Name	Accession Number		Chr. No.	TIGR v5 pseudomolecule position	ORF (bp)	Introns
		TIGR	KOME				
37	OsHsp20-37	LOC_Os10g30180	/	10	15359380–15360457	552	1
38	OsHsp20-38	LOC_Os11g13980	AK107883	11	7746533–7745913	621	0
39	OsHsp20-39	/	AK109077	/	/	792	0

similarities among the Hsp20 proteins sharing these common motifs both in rice and *Arabidopsis*. However, the biological significance of these motifs remains to be elucidated.

The aligned crystallin domains (Fig. 1b) can be subdivided into consensus I and consensus II domains separated by a hydrophilic domain of variable length (Vierling 1991; Waters et al. 1996). The residues Pro-X(14)-Gly-Val-Leu in consensus I are a signature typical in almost all Hsp20 proteins, whereas a similar motif, Pro-X(14)-X-Val/Leu/Ile-Val/Leu/Ile, also appears in the consensus II region (Caspers et al. 1995).

To examine the evolutionary relationships of the Hsp20 family in rice and *Arabidopsis*, an unrooted phylogenetic tree was constructed from alignments of their full-length protein sequences, which resulted in the formation of 12 distinct clusters named Class I, II, III, IV, V, VI, VII, M I, M II, P, ER and Po comprising 17, 6, 5, 1, 2, 2, 1, 4, 2, 4, 3 and 2 proteins, respectively (Fig. 1c). Hsp20 proteins from *Arabidopsis* are present in all the classes. However, no OsHsp20 protein was found in Class IV and VII. This classification was consistent with the result suggested by Siddique et al. (2008) that genes homologous to *Arabidopsis* Class IV member were exclusively found in dicotyledonous. Many OsHsp20 proteins were classified to Class I or Class II in cytosol/nucleus, indicating that cytosol may be the main function area for OsHsp20. Notably, the OsHsp20 members were more closely to those in the same class in divergent species than to other Hsp20 proteins in the same species (Fig. 1c), implying that synteny was relatively well conserved between rice and *Arabidopsis* proteins.

The phylogenetic classification was found to be consistent with the pattern of intron type. The cytosolic V and VI subfamilies, together with the mitochondria subfamilies of M I and M II, share a unique intron position (Table 1, Fig. 1c). The C I, ER and Po subfamilies all lack introns, suggesting a close phylogenetic relationship (Table 1, Fig. 1c). In addition, most of the C II and P members show no intron (Table 1, Fig. 1c). Interestingly, it is noteworthy that all two genes belonged to C III subfamily have 12 and 13 introns, respectively (Table 1, Fig. 1c). The presence of multiple introns indicates particular phylogenetic status. In a word, intron patterns shed light on the phylogenetic relationships.

All plant Hsp20 proteins are encoded by nuclear genes. They are divided into 12 classes and the majority of Hsp20 proteins in rice accord with such classification. However, there was a special protein, OsHsp20-28, whose localization was in membrane. The similar membrane localization has been found in GmHSP22.3 protein of soybean (LaFayette et al. 1996), which might be an indication of a potential new class.

Expression analysis of *OsHsp20* genes under heat shock

Because it would be too hard to exhaustively describe the expression of all 39 *OsHsp20* genes in response to heat shock, we used RT-PCR to determine the expression patterns of 11 *OsHsp20* genes, belonging to 11 different classes in *Hsp20* family, to understand the expression patterns in heat shock (Fig. 2). Based on our results, *OsHsp20-6* belonged to Class II expressed extremely low in all conditions. *OsHsp20-4*, *OsHsp20-31*, *OsHsp20-32* and *OsHsp20-33* were detected to be expressed in 25°C condition, while expression level of other genes appeared to be rather low. In response to heat shock, we found that the expression pattern of *OsHsp20* genes selected can be divided into two types. For *OsHsp20-31* and *OsHsp20-32*, the expression was relatively abundant in different temperatures, showing no significant differences according to the heat treatment. Interestingly, *OsHsp20-32* was observed to have two transcripts and sequencing analysis showed that 85 bp deletion was found in the small fragment obtained by RT-PCR compared to the long fragment. It was speculated that the long fragment observed might be the unspliced precursor of the *OsHsp20-32*. On the other hand, we found the expression of all *OsHsp20* genes selected, except *OsHsp20-31* and *OsHsp20-32*, were up-regulated in response to heat shock at 37°C. Not surprisingly, the transcripts of *OsHsp20-27*, *OsHsp20-14*, *OsHsp20-17*, *OsHsp20-38* and *OsHsp20-30* were further accumulated in 42°C when compared with those in 37°C. The expression profiles revealed that these *OsHsp20* genes were up-regulated in response to heat shock although there were remarkable differences in the transcription level.

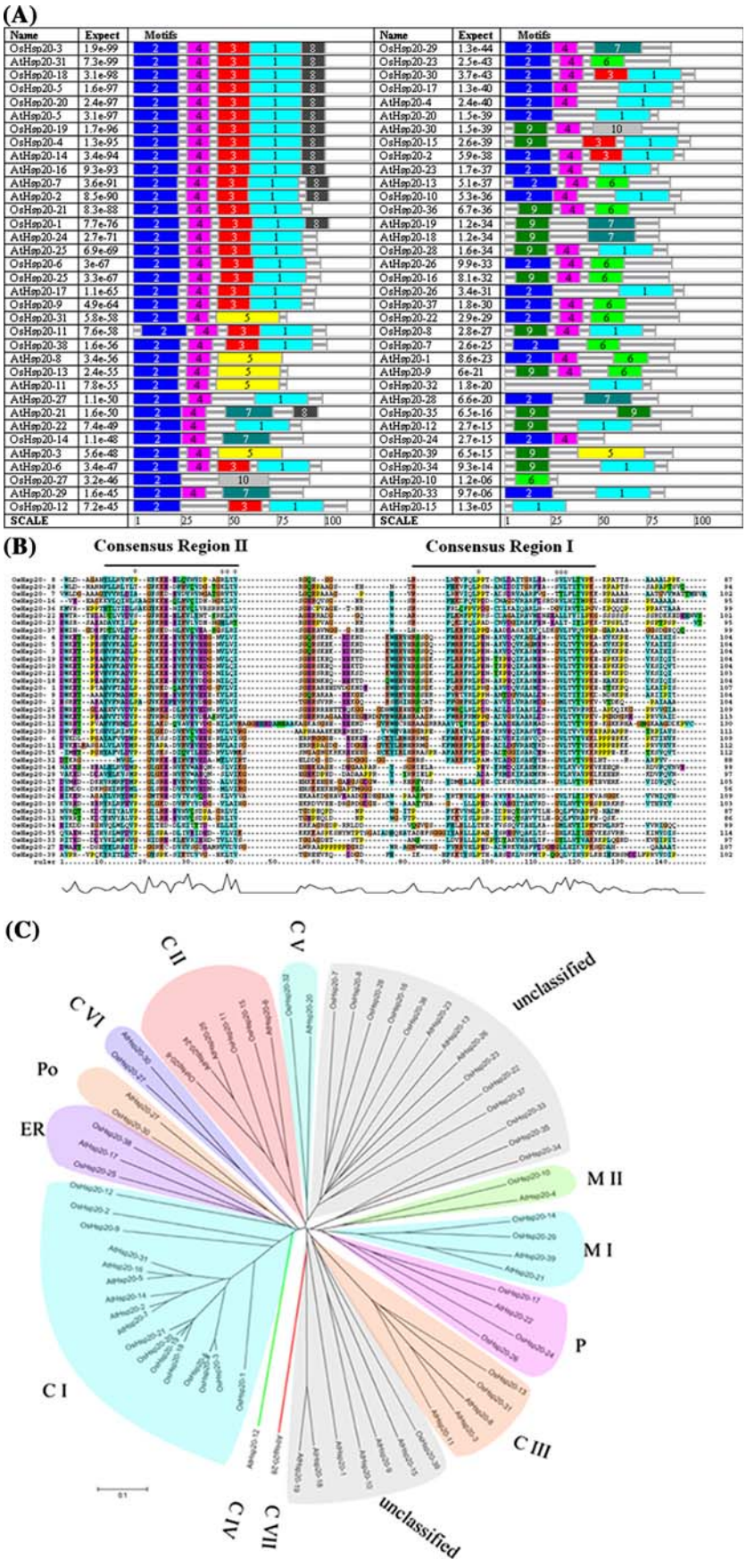

Fig. 1 Structural and phylogenetic analysis of Hsp20 protein. **a** Distribution of conserved motifs in Hsp20 proteins identified by MEME software. Ten putative motifs are identified and each motif is represented by a number in *colored box*. Names of all the members and combined e-values are shown on the left side of the figure and motif sizes are indicated at the bottom of the figure. **b** Alignment of OsHsp20 crystallin domains. Only gene identity numbers are provided on the left. Color shading indicates type of amino acid residue, where these residues are conserved. The amplitude of the wave below alignment represents the degree of conservation at each position. The defined consensus regions I and II are *underlined* on the top and *asterisks* indicate conserved residues within these regions as discussed in the text. **c** Phylogenetic tree of rice and *Arabidopsis* Hsp20 proteins. An unrooted NJ tree of 39 rice and 31 *Arabidopsis* Hsp20 proteins is shown. The 12 classes are marked by *different colors*. The rice proteins which were unclassified on alignment with *Arabidopsis* proteins have been colored in grey. Scale bar represents 0.1 amino acid substitution per site

Table 2 Details of the ten putative motifs

Motif	Motif length (aa)	Sequence
1	33	FRLPENCDMDQIKASMENGVLTVTVPKLPPPKP
2	29	DWKETPECHVFYADMPGLKKEEVKVQVED
3	21	QEEKNDKWHRVERSSGQFMRR
4	11	NVLQISGERSR
5	41	NPWGITPFKKVVNLPARIDPHHTSAVVTLHGQLFVRVPFEQ
6	21	WSRFHKMFQLPPNCDLDAISA
7	29	RRYTSRIGLPCDCYKIDNIKCFMKNGVLW
8	8	QVKSIQIS
9	21	HNYFLWLDCPGYSKESIQVQL
10	30	DWRAGRWWEHGYVRRLELPEDADWKYSEAF

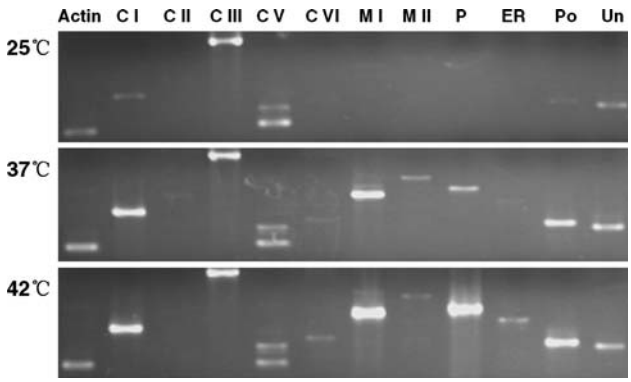

Fig. 2 Expression patterns of *OsHsp20* genes in response to temperature increase. Three-day-old rice seedlings were subjected to 25°C (control), 37°C and 42°C for 2 h. The subfamily type of each gene was annotated above the figure. Gene names in RT-PCR: Actin = β-actin used as an internal PCR control; C I = *OsHsp20-4*; C II = *OsHsp20-6*; C III = *OsHsp20-31*; C V = *OsHsp20-32*; C VI = *OsHsp20-27*; M I = *OsHsp20-14*; M II = *OsHsp20-10*; P = *OsHsp20-17*; ER = *OsHsp20-38*; Po = *OsHsp20-30*; Un indicates the unknown type = *OsHsp20-33*

Chromosomal localization and gene duplication

To determine the genomic distribution of *OsHsp20* genes, their chromosomal positions were identified according to the TIGR rice database. Out of 39 *OsHsp20* genes, 38 could be located on ten of the rice chromosomes by virtue of their presence on the TIGR pseudomolecules (Fig. 3), while *OsHsp20-39* is present as a cDNA in the KOME database only presumably from the unsequenced regions of the rice genome. *OsHsp20* genes were present in all regions of the chromosomes, i.e., at the telomeric ends, near centromere, scattered all over or in clusters (Fig. 3). Their distributions amongst the ten rice chromosomes were highly variable. A maximum number of nine genes was present on chromosome 3 closely followed by eight genes on chromosome 1. On the other hand, no *OsHsp20* gene was present on chromosome 8 and 12. Three chromosomes, 1, 3 and 10, had groups of *OsHsp20* genes in vicinity of each other. Interestingly, these clusters contained *OsHsp20* genes that were transcribed in reverse orientations. *OsHsp20-26* was located very close to the centromere and the rice centromeres have been found to contain actively transcribed genes (Bohm et al. 1997; Cooke 2004; Nagaki et al. 2004).

Gene families are generated through either tandem duplication or large-scale segmental duplication during evolution (Cannon et al. 2004). Analysis of the TIGR rice segmental duplication database revealed only six (three pairs) *OsHsp20* genes could be assigned to TIGR segmental duplication blocks at a maximal length distance permitted between collinear gene pairs of 500 kb. The overall similarity of the cDNA sequences of these genes ranged from 38.2% to 60.2% and all of them were found to have their counterparts on duplicated segments (Supplementary Table 4, Fig. 3). All these genes exhibited high sequence similarity in both the Hsp20 domain and the flanking regions. 17 (43.59%) of the *OsHsp20* genes seemed to be produced from tandem duplications. Six groups of genes were found to be tandemly duplicated (Supplementary Table 5). They were separated by a maximum of five intervening genes. Four groups of the gene pairs were placed juxtaposed with no intervening gene. The distance between these genes ranged from 0.5 kb to 27.6 kb (Supplementary Table 5). There were two genes in tandem in most cases. However, groups I and III (Supplementary Table 5) had more than two genes. These results suggested that both tandem duplication and segmental genome duplication events had contributed largely to the expansion of the *OsHsp20* gene family in rice.

The expression patterns of *OsHsp20* genes presented in segmentally duplicated regions and for tandemly duplicated genes were also examined. Analysis of the TIGR rice segmental duplication database revealed six *OsHsp20* genes (three pairs) were localized on the duplicated segments of the rice genome. A comparison of expression profiles of the duplicated gene pairs from microarray data revealed that the expression pattern was quite similar for all

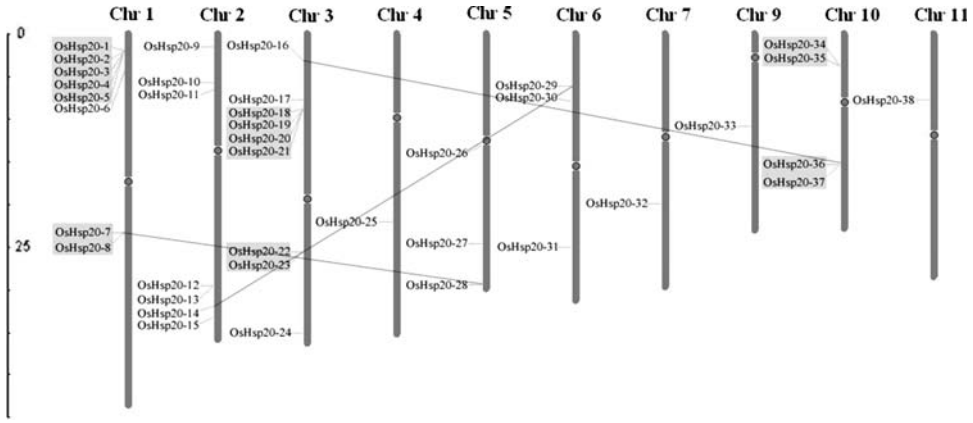

Fig. 3 Genome wide distribution of *OsHsp20* genes on rice chromosomes. Only gene identity numbers are provided. Grey ovals on the chromosomes (*vertical bars*) indicate the position of centromeres. Chromosome numbers are indicated at the top of each bar. The *OsHsp20* genes present on duplicated chromosomal segments are connected by *line connectors* and tandem duplicated genes are marked by *grey shaded boxes*

three pairs of genes (Fig. 4a). This is an indication that the segmentally duplicated genes in *OsHsp20* family maintain their functions during evolution all the time.

The result for tandem duplicated genes was more complicated. Out of six groups of tandemly duplicated genes, probe sets were available for five of them in microarray data. In most of the cases, the expression patterns of the tandem duplicated genes were highly similar in four groups (Fig. 4b, Table 3), even though the expression of one of the duplicated genes was extremely low. This may be due to the fact that the gene with low level of expression will, in the course of evolution, slowly lose its function. Genes in the fourth group showed divergent expression profiles, as *OsHsp20-22* did not express at significant levels in most of the tissues. This is an indication of pseudo-functionalization after duplication. The comparison of their sequences revealed much reduced level of homology at amino acid level, indicating that these genes might have undergone significant diversification after the duplication of the respective genomic segments leading to neo-functionalization for the pair partners.

Genome-wide expression analysis of *OsHsp20* genes in three rice cultivars during vegetative and reproductive development

Gene expression data of *OsHsp20* genes were from CREP database. The results from the database for 25 tissues with two or more replications in three cultivars indicated that the expression of *OsHsp20* genes was detected at different developmental stages. One gene, *OsHsp20-7*, which was not represented in CREP database, was studied in MPSS database. RT-PCR was performed to confirm the expression profile of the microarray data respectively in our lab (Nayidu et al. 2008; Nuruzzaman et al. 2008).

The genes exhibiting high or low expression were segregated based on MAS 5.0 software. Only 36 genes showed a ''present'' detection call at *P* value of 0.05 in at least one of the stages analyzed, whereas two low expressing genes, *OsHsp20-11* and *OsHsp20-24*, were either ''absent'' or ''marginal'' under these conditions. Two separate heat maps (Fig. 5a, b) were generated to emphasize the differential expression profiles of these two groups of genes.

OsHsp20-11 and *OsHsp20-24* transcripts were at almost undetectable levels in all the stages analyzed, but it is possible that these genes might respond to specific stimuli or their expression might be limited to specialized cell types that have not been analyzed in this investigation. Other genes that showed differential expression during various stages of development in comparison to seed were summarized in Fig. 6. Detailed *P* values and fold change values are given in Supplementary Table 6. It can be observed that the numbers of up-regulated genes in all stages shared a similar pattern in three cultivars. Conversely, the number of down-regulated genes was much greater in Shanyou 63 than in Minghui 63 and Zhenshan 97 in all stages of growth development.

During vegetative stages, 19 *OsHsp20* genes were up-regulated and 20 genes were down-regulated by two-fold or more in Minghui 63. Similarly, 17 up- and 17 down-regulated genes were found in Zhenshan 97, while 17 up- and 25 down-regulated in Shanyou 63. It should be noted that of all the up-regulated genes that were found above, 11 genes were differentially expressed in all three cultivars. Similarly, 12 down-regulated genes were differentially expressed in all three cultivars. These data showed that the differentially regulated *OsHsp20* genes did not exhibited high variations, implying that their expression could be quite conservative in vegetative development in different cultivars. It can be observed that most *OsHsp20* genes (77.92%) shared a similar expression pattern in all three cultivars and were either up-or down-regulated in different stages in vegetative development. Only nine genes were found to be up-regulated during several stages but down-regulated during other stages in at least one of the cultivars.

The result for the number of differentially regulated genes in reproductive development was not highly similar to that in vegetative development. Twenty-three *OsHsp20*

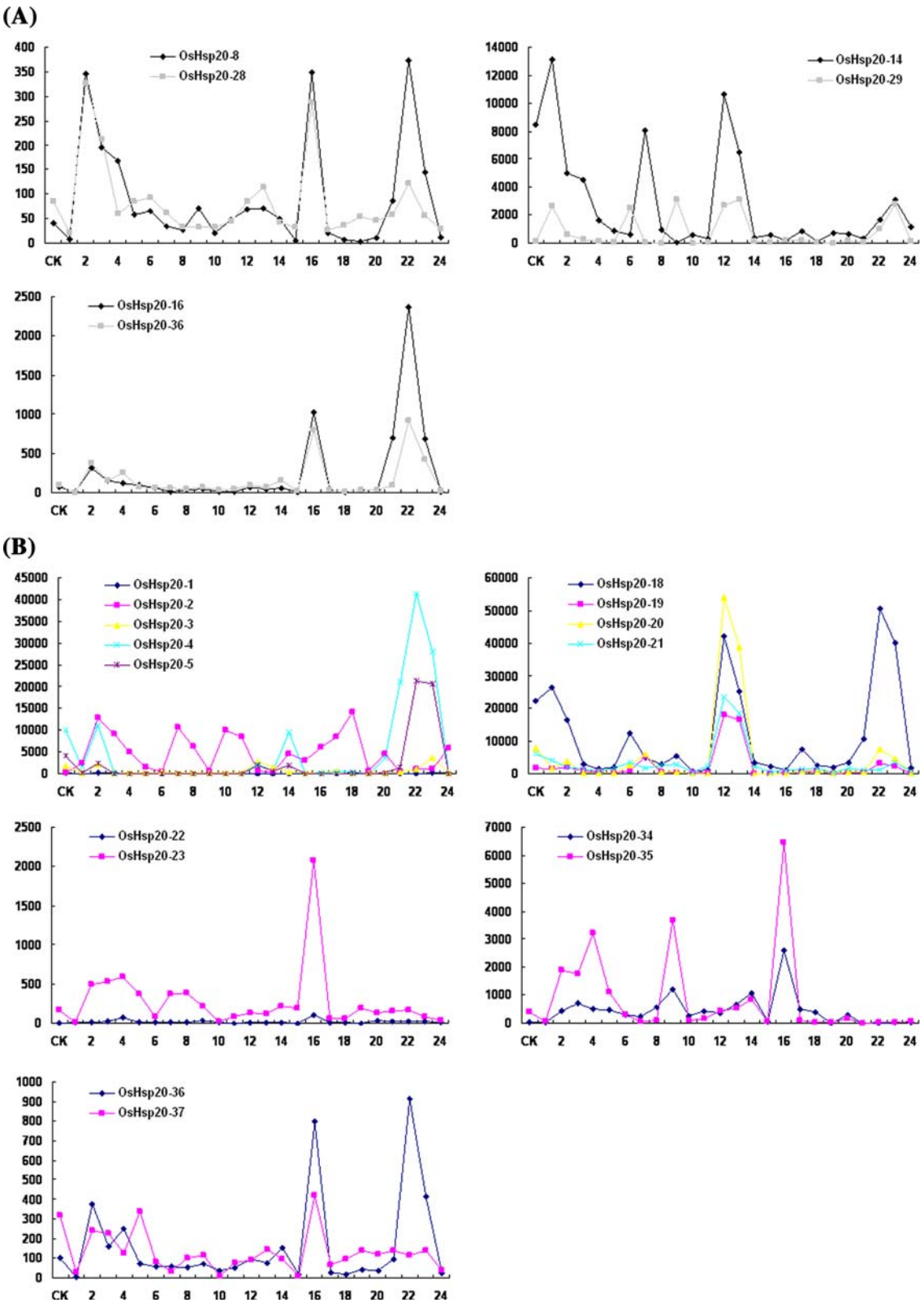

Fig. 4 Expression patterns of *OsHsp20* genes found in segmentally duplicated regions of the rice genome **a** and present as tandem duplicates **b**. *X*-axis represents the developmental stages as given in Table 3. *Y*-axis represents the raw expression values obtained from microarray

Table 3 The developmental stages are represented in *X*-axis

Sample	Developmental stages analyzed
CK	Seed: germination (72 h after imbibition)
1	Calli:15 day after induction
2	Seedling 1: 3 days after sowing
3	Seedling 2: trefoil stage
4	Shoot 1: seedling with 2 tillers
5	Root 1: seedling with 2 tillers
6	Panicle 1: secondary branch primordium differentiation stage
7	Leaf 1: secondary branch primordium differentiation stage
8	Sheath1: secondary branch primordium differentiation stage
9	Panicle 2: 4–5 cm young panicle
10	Leaf 2: 4–5 cm young panicle
11	Sheath 2: 4–5 cm young panicle
12	Panicle 3: pistil/stamen primordium differentiation stage
13	Panicle 4: pollen-mother cell formation stage
14	Stem 1: 5 days before heading
15	Flag leaf 3: 5 days before heading
16	Stem 2: heading stage
17	Panicle 5: heading stage
18	Glume: one day before flowering
19	Stamen: one day before flowering
20	Spikelet: 3 days after pollination
21	Endosperm 1: 7 days after pollination
22	Endosperm 2: 14 days after pollination
23	Endosperm 3: 21 days after pollination
24	Flag leaf 4:14 days after heading

genes were up-regulated equal to or more than two-fold in Minghui 63, whereas 23 were down-regulated. Meanwhile, 31 up- and 14 down-regulated genes were found in Zhenshan 97, and 19 up- and 33 down-regulated in Shanyou 63. It was also found that a majority of up-regulated genes (15 genes) were differentially expressed in all three cultivars, but only eight genes were found to be down-regulated in common. This confirmed the probability that the expression pattern of *OsHsp20* genes in reproductive development was much more complex, implying the activation of different genes in different cultivars in these cases. A biphasic expression pattern was also observed for *OsHsp20* genes in reproductive development, as 18 genes were found to be up-regulated during several stages but down-regulated during other stages in at least one of the cultivars.

Certain genes were up-regulated or down-regulated during vegetative as well as in reproductive development, probably because of the prevalence of similar processes in these stages. Seven *OsHsp20* genes (*OsHsp20-2*, *8*, *13*, *16*, *27*, *34* and *35*) showed up-regulated expression during both vegetative and reproductive development in all three cultivars, whereas another seven down-regulated genes (*OsHsp20-4*, *9*, *15*, *17*, *19*, *25* and *38*) were found.

One gene, *OsHsp20-7*, did not have corresponding probe set in the microarray database used in the present study. An investigation to the expression of *OsHsp20-7* in MPSS database revealed that this gene did not express in any stage.

Heterosis of gene expression

To identify significant mid-parent heterosis of gene expression, an *h*-statistic analysis as well as the *F*-test at $P < 0.05$ probability level were performed between the hybrid and parents. A total of 32 (82.05%) *OsHsp20* genes showed significant expression heterosis in 25 tissues (Supplementary Table 7), of which 15 genes were identified as showing significant positive heterosis in panicle at pollen-mother cell formation stage, and 13 genes showed negative heterosis in seed.

There were obvious differences in different stages in the numbers of heterotic expression sequences and directions of mid-parent heterosis (Fig. 7). Interestingly, all the *OsHsp20* genes exhibited expression heterosis in the same directions in one tissue except in tissue 15, 18 and 23 where opposite directions of expression heterosis were found (Fig. 7a). Among the five genes heterotically expressed in leaf (Fig. 7b), four showed negative heterosis. 66.67% (14) of the 21 heterotically expressed genes showed significant positive heterosis in panicle (Fig. 7c), 23.81% (5) of the genes showed negative heterosis, and two (9.52%) genes showed opposite directions of expression heterosis in this tissue. All the genes heterotically expressed in seedlings and calli (Fig. 7d) showed significant positive heterosis, whereas 88.89% (16) of the 18 heterotically expressed genes showed significant positive heterosis in seed and endosperm (Fig. 7e). This result seems to indicate that genes showing positive heterosis or negative heterosis were relevant to tissue specificity.

Discussion

Heat shock response

This study presents genomic information of the *OsHsp20* gene family in rice. As stated previously, the expression of *OsHsp20* genes was always induced by heat treatment. Our current results were largely consistent with the results in other papers, where nine Class I *OsHsp20* genes were reported to be up-regulated by heat shock in rice (Guan et al. 2004), and a number of *Hsp20* genes were up-regulated by heat shock in zebrafish (Kimberly and Lara 2007) and *Arabidopsis* (Swindell et al. 2007). In our study, four

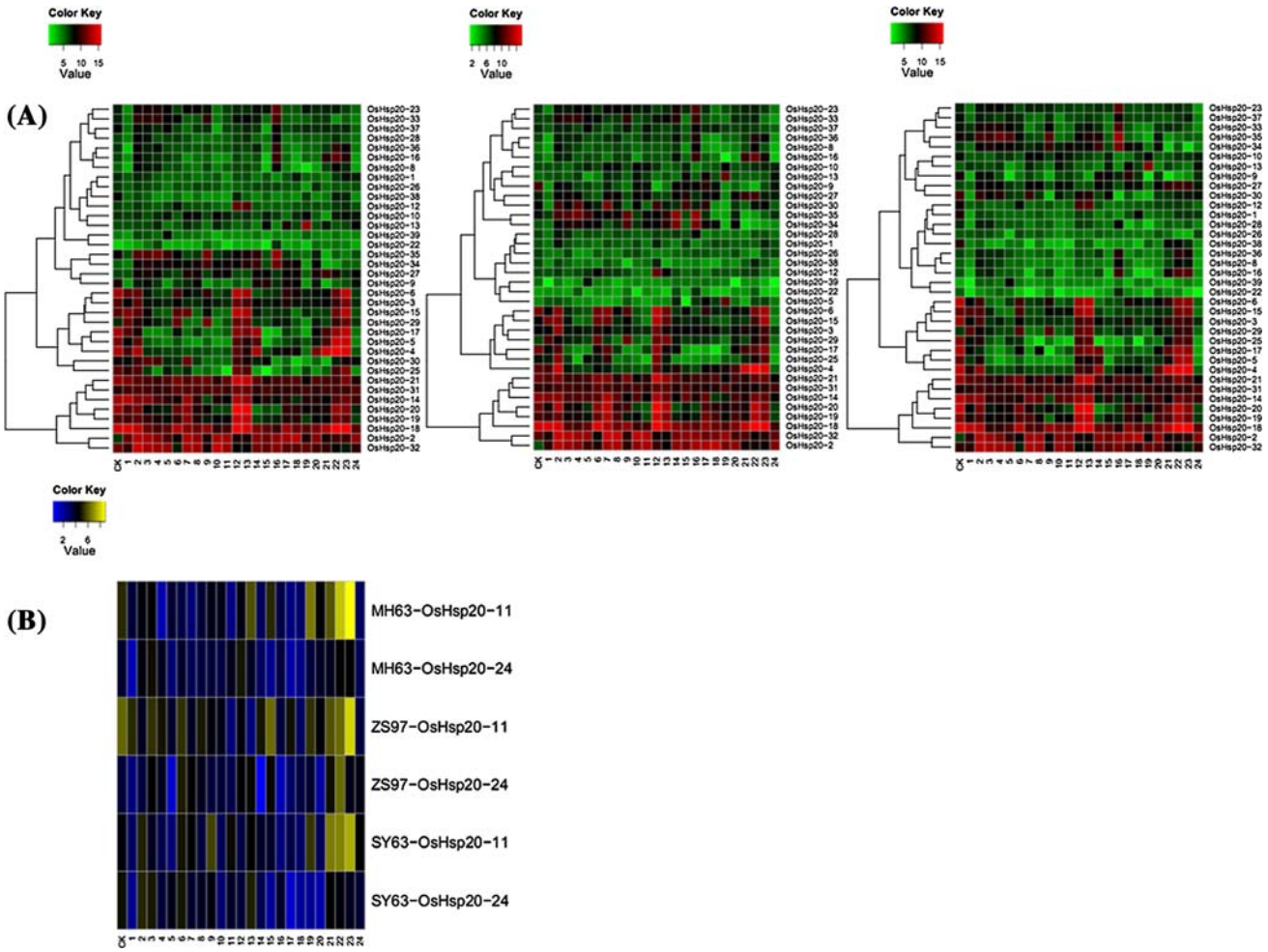

Fig. 5 Expression pattern of *OsHsp20* genes during the life cycle of the rice plant. **a** Hierarchical cluster display of expression profile for 36 *OsHsp20* genes showing high level expression in at least one of the samples in Minghui 63, Zhenshan 97, Shanyou 63, separately. (*Color bar* at the base represents log2 expression values: *green*, representing low expression; *black*, medium expression; *red*, high expression). **b** Hierarchical cluster display of expression profile for two *OsHsp20* genes showing low level expression in all the tissues. (*Color bar* at the base represents log2 expression values: *green*, representing low expression; *black*, medium expression; *red*, high expression). The color scale (representing log signal values) is shown above. All tissues used for expression profiling on bottom of each column are mentioned in Fig. 4. On the left side of expression map, cluster dendrogram is shown

OsHsp20 genes were detected to be expressed in 25°C conditions, these genes might be involved in specific housekeeping functions of rice under normal growth conditions. All *OsHsp20* genes selected, except *OsHsp20-31* and *OsHsp20-32*, were increased in different level under heat shock (Fig. 2). Our results indicated that, to varying extents, heat treatment induced a number of *OsHsp20* genes at different expression level. The similarity of expression patterns among the OsHsp20 proteins may reflect shared induction mechanisms. However, differences in the accumulation of transcript levels during heat treatment indicate that *Hsp20* genes might be specifically controlled by different Hsfs (Schramm et al. 2006).

The heat shock response network of rice involves both HSPs and Hsfs, and the expression of HSPs is mainly attributed to the activation of Hsfs under heat stress. The exact number of Hsf genes differs greatly among various eukaryotic organisms. While animals have only one to four genes encoding Hsf, plants contain more Hsf genes, as the *Arabidopsis* genome contains 21 copies of Hsf (Kotak et al. 2004). Intriguingly, the expression of several but not all of the plant Hsf family is heat-induced, this eukaryotic system is a feature unique to plants (Nover et al. 2001), and this feature suggests a multi-step mechanism of Hsf involvement in the heat response. It appears that the diversity of Hsfs provides redundancy and specialization of stress signals, a means to differentially control the rate of transcription of heat shock genes, and provides novel interactions with other regulatory factors thus expanding the link between cell stress and other genetic networks. Meanwhile, Hsfs are the terminal compounds of the signal transduction pathway to activate the expression of the *Hsp*

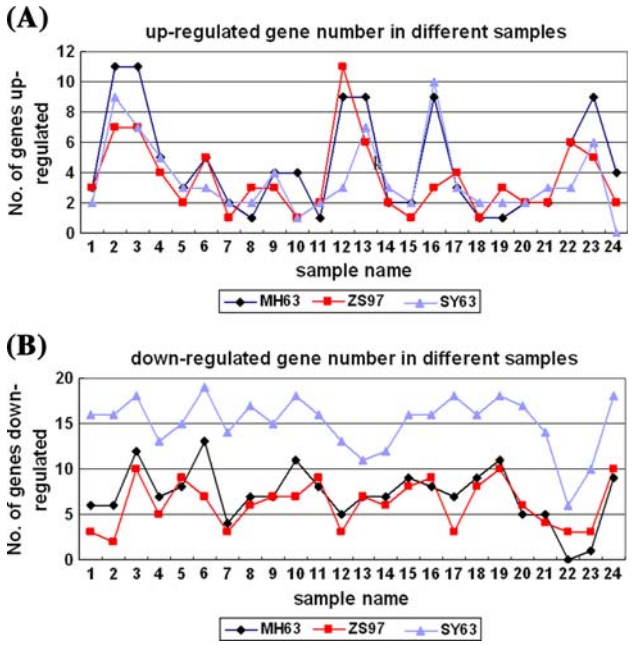

Fig. 6 Differential expression of *OsHsp20* genes in different stages in three cultivars based on microarray analysis. **a** Number of genes up-regulated in various stages. **b** Number of genes down-regulated in various stages. X-axis represents the developmental stages as given in Table 3

genes (Chen et al. 2006). Hsfs participate in regulating the expression of such heat stress-inducible genes by recognizing the conserved binding motifs (heat shock response element, HSE) that exist in their promoters (Bienz and Pelham 1987). Except for *OsHsp20-33*, cluster of HSE modules were found varying from two to nine in the 2 kb promoter regions of *OsHsp20* genes induced by heat treatment in the experiment (Supplementary Table 8) (Higo et al. 1999; Prestridge 1991). Since the HSE modules mediate efficient binding of Hsfs and the regulation of *Hsp20* genes seems to be linked to the network of Hsfs due to the complex patterns of HSE (Scharf et al. 2001), it will be interesting to study the connection between the HSP, Hsfs and HSE in the future studies of the rice heat shock response network.

Evolution of the *OsHsp20* gene family

Gene duplications are one of the primary driving forces in the evolution of genomes and genetic systems (Moore and Purugganan 2003). For most gene families, their dramatic variations in family size and distribution are affected by tandem duplications and segmental duplications (Cannon et al. 2004). Duplication and subsequent expansion of *OsHsp20* genes seemed to have occurred frequently throughout evolution and these results shed light on the evolution process of the rice genome. *OsHsp20* gene family comprises a diverse and rapidly evolving family of proteins. A large number of *OsHsp20* genes were involved in tandem duplications and segmental duplications in rice genome. *Arabidopsis* was found to have 31 *AtHsp20* genes. However, in rice with a genome size three times of *Arabidopsis*, the number of *OsHsp20* genes was found to be only 39. The reason for this could be the variable status of whole genome duplications in *Arabidopsis* and rice (Yu et al. 2005). It is due to less genomic duplication events contributing to the expansion of the *OsHsp20* gene family in rice. A total of 23 *OsHsp20* duplicated genes were present in rice chromosomes. Except for those genes which were unclassified in type, it seems that tandem duplication events have contributed significantly towards evolution of C I genes. Our analysis also revealed only two M I genes were generated because of segmental duplications except those were unclassified. Phylogenetic analyses indicated that C I genes clustering on a chromosome resembled more to each other than to those on other chromosomes in rice. Phylogenetic analysis also suggested that rice and *Arabidopsis Hsp20* genes on individual chromosomes probably had a common ancestor and were duplicated independently after divergence of monocots and dicots. Through comparison of their amino acid sequences, we found their homology ranged from 8.6% to 99.3% (see Supplemental Tables 4 and 5). However, our analysis did not show any correlation between the percentage of homology and the expression profiles of the genes in the duplicated regions of rice. The results indicated functional similarities in all segmentally duplicated genes despite remarkable differences between the amino acid sequences of them. We also found that the expression pattern of the tandem duplicated genes was highly similar in four groups (Fig. 4b). The similarity of expression patterns among the *OsHsp20* genes may reflect shared induction mechanisms and network. However, redundancy, as one of the fates of duplication, was also found in *OsHsp20* family. *OsHsp20-22* in the fourth group in tandem duplicated genes was not expressed at significant levels in most of the tissues. It might be an indication of pseudofunctionalization after duplication.

Heterosis of gene expression

OsHsp20 genes were identified as showing prominent heterosis in family-level expression. However, little is known regarding how OsHsp20 proteins were regulated. HSPs and Hsfs are central components of the rice regulatory network. However, Hsfs participate in regulating the expression of HSPs, which are critical in the protection against stress damage and many other important biological processes. The expression of heat shock genes are mainly attributed to the activation of Hsfs. It has long been recognized that these elements are also involved in response to non-thermal stress treatments (Feder and Hofmann 1999),

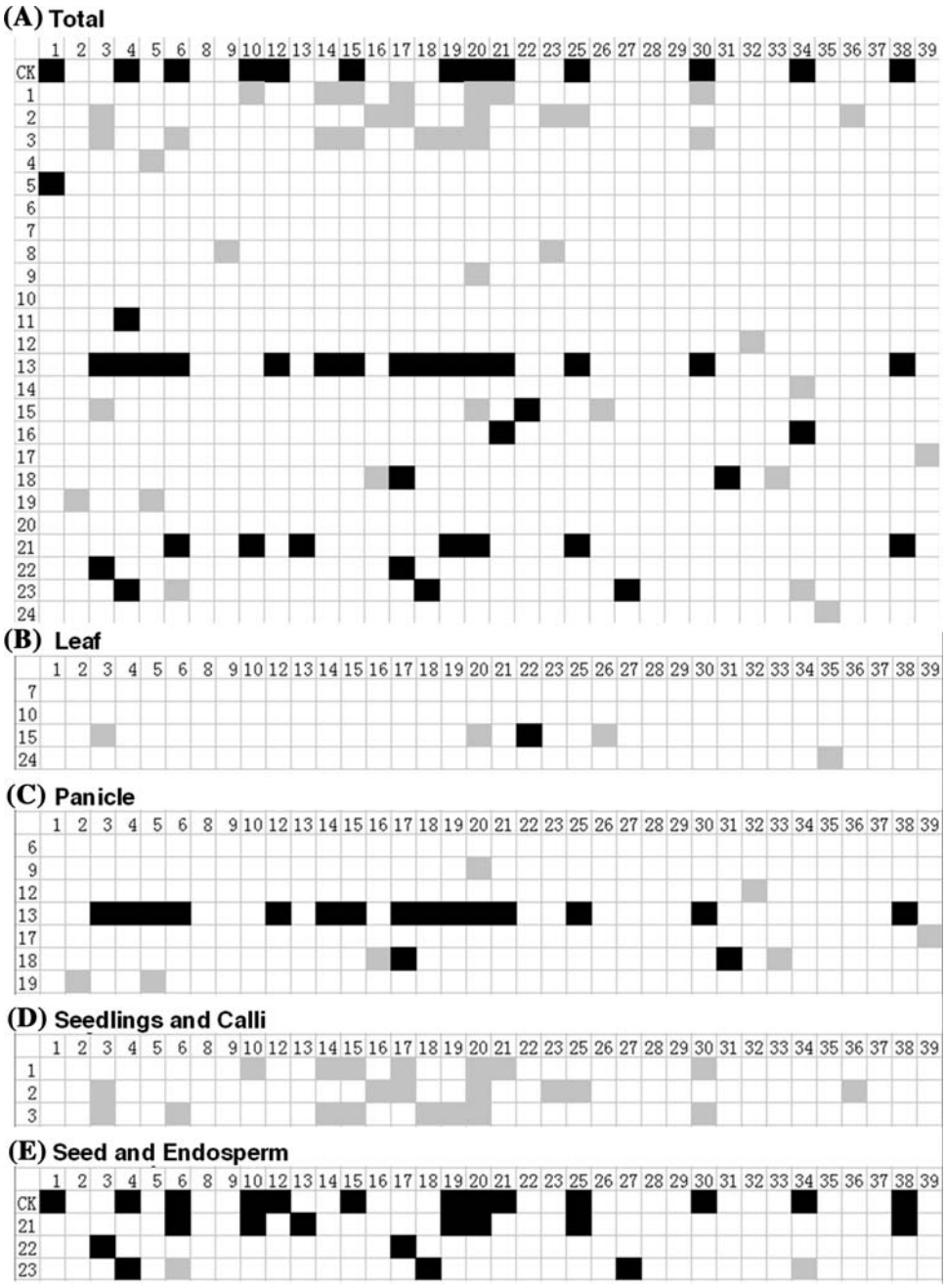

Fig. 7 *OsHsp20* genes exhibited heterosis expression (black, positive heterosis; grey, negative heterosis) in 25 tissues (**a**), four leaf tissues (**b**), seven panicle tissues (**c**), seedlings and calli (**d**) and seed or endosperm (**e**). Numbers in horizontal axis represent *OsHsp20-1* to *OsHsp20-39*. Numbers in vertical axis represent the developmental stages as given in Fig. 4

but the phenomenon that *OsHsp20* genes were identified as showing prominent heterosis in family-level have not been identified, and the mechanism of how these proteins work was unclear. Heat shock transcription factors are of fundamental importance to understanding metabolism networks involved in *OsHsp20* genes, since these proteins coordinate the expression of Hsfs and other factors. Shared induction mechanisms among OsHsp20 proteins may include accumulation of denatured proteins in the cytoplasm (Sung et al. 2003), generation of reactive oxygen species (Miller and Mittler 2006), or changes in membrane lipid composition and fluidity (Tsvetkova et al. 2002). These processes are thought to be upstream signals leading to the activation of critical Hsfs, which are most likely to be the direct inducers of *OsHsp20* expression. We have examined the expression levels of the 24 *OsHsf* members in rice between the hybrid and parents. Surprisingly, a total of 20 (83.33%) *OsHsf* genes showed significant expression heterosis in 15 tissues (Supplementary Table 9), of which 14 genes were identified as showing significant positive heterosis and 14 genes were showed as significant negative heterosis in different stages, as well as eight *OsHsf* genes were identified as showing either positive or negative heterosis. The striking percentage of *OsHsf* genes identified

as showing heterosis supports the hypothesis that HSPs and Hsfs are associated with each other and represent an intersection point in heterosis phenomenon.

In this paper, we have presented a comprehensive expression profiling for all the *OsHsp20* genes in rice along with an account of their phylogenetic relationships with the *Arabidopsis* genes. While stage specific expression in terms of up- or down-regulation was evident, it may be emphasized that groups of the genes that showed similar expression profile during various developmental stages may or may not have similar functions because their expression could be confined to specific cell types. Therefore, these data would be useful in selecting candidate genes for functional validation in relation to stress and various aspects of vegetative and reproductive development in rice and other crops.

Acknowledgements We thank two anonymous reviewers for their careful reading and helpful comments on this manuscript. This research was supported in part by grants from the National Special Key Project of China on Functional Genomics of Major Plants and Animals, and the National Natural Science Foundation of China.

References

Banzet N, Richaud C, Deveaux Y, Kazmaier M, Gagnon J, Triantaphylides C (1998) Accumulation of small heat shock proteins, including mitochondrial HSP22, induced by oxidative stress and adaptive response in tomato cells. Plant J 13:519–527. doi:10.1046/j.1365-313X.1998.00056.x

Bienz M, Pelham HR (1987) Mechanisms of heat-shock gene activation in higher eukaryotes. Adv Genet 24:31–72. doi:10.1016/S0065-2660(08)60006-1

Bohm S, Frishman D, Mewes HW (1997) Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. Nucleic Acids Res 25:2464–2469. doi:10.1093/nar/25.12.2464

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4:10. doi:10.1186/1471-2229-4-10

Caspers GJ, Leunissen JAM, De Jong WW (1995) The expanding small heat-shock protein family, and structure predictions of the conserved "α-crystallin domain". J Mol Evol 40:38–48. doi:10.1007/BF00163229

Chen XJ, Ye CJ, Lu HY (2006) Cloning of GmHSFA1 gene and its overexpression leading to en-hancement of heat tolerance in transgenic soybean. Acta Genet Sin 28:1411–1420

Cheong YH, Chang HS, Gupta R, Wang X, Zhu T, Luan S (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. Plant Physiol 129:661–677. doi:10.1104/pp.002857

Coca MA, Almoguera C, Jordano J (1994) Expression of sunflower low-molecular-weight heat-shock proteins during embryogenesis and persistence after germination: localization and possible functional implications. Plant Mol Biol 25:479–492. doi:10.1007/BF00043876

Cooke HJ (2004) Silence of the centromeres–not. Trends Biotechnol 22:319–321. doi:10.1016/j.tibtech.2004.04.015

de Jong WW, Caspers GJ, Leunissen JAM (1998) Genealogy of the a-crystallin-small heat-stress superfamily. Int J Biol Macromol 22:151–162. doi:10.1016/S0141-8130(98)00013-0

Desikan R, Mackerness SAH, Hancock JT, Neill SJ (2001) Regulation of the *Arabidopsis* transcriptome by oxidative stress. Plant Physiol 127:159–172. doi:10.1104/pp.127.1.159

Edelman L, Czarnecka E, Key JL (1988) Induction and accumulation of heat shock-specific poly (A+) RNAs and proteins in soybean seedlings during arsenite and cadmium treatments. Plant Physiol 86:1048–1056. doi:10.1104/pp.86.4.1048

Feder ME, Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu Rev Physiol 61:243–282. doi:10.1146/annurev.physiol.61.1.243

Guan J-C et al (2004) Characterization of the genomic structures and selective expression profiles of nine class I small heat shock protein genes clustered on two chromosomes in rice (*Oryza sativa* L.). Plant Mol Biol 56:795–809. doi:10.1007/s11103-004-5182-z

Guruprasad K, Reddy BV, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4:155–161. doi:10.1093/protein/4.2.155

Haas BJ, Delcher AL, Wortman JR, Salzberg SL (2004) DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics 20:3643–3646. doi:10.1093/bioinformatics/bth397

Harrington HM, Alm DM (1988) Interaction of heat and salt shock in cultured tobacco cells. Plant Physiol 88:618–625. doi:10.1104/pp.88.3.618

Helm KW, Schmeits J, Vierling E (1995) An endomembrane-localized small heat-shock protein from *Arabidopsis thaliana*. Plant Physiol 107:287–288. doi:10.1104/pp.107.1.287

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database:1999. Nucleic Acids Res 27:297–300. doi:10.1093/nar/27.1.297

Hsieh MS, Chen JT, Jinn TL, Chen YM, Lin CY (1991) A class of soybean low molecular weight heat shock proteins: immunological study and quantitation. Plant Physiol 99:1279–1284. doi:10.1104/pp.99.4.1279

Huang Y, Li LH, Chen Y, Li XH, Xu CG, Wang SP, Zhang QF (2006) Comparative analysis of gene expression at early seedling stage between a rice hybrid and its parents using a cDNA microarray of 9198 uni-sequences. Sci China Ser C 49:519–529. doi:10.1007/s11427-006-2031-0

Jain M, Khurana P, Tyagi AK, Khurana JP (2008) Genome-wide analysis of intronless genes in rice and *Arabidopsis*. Funct Integr Genomics 8:69–78. doi:10.1007/s10142-007-0052-9

Jofré A, Molinas M, Pla M (2003) A 10-kDa class-CI sHsp protects *E. coli* from oxidative and high-temperature stress. Planta 217:813–819. doi:10.1007/s00425-003-1048-x

Kim KK, Kim R, Kim SH (1998) Crystal structure of a small heat-shock protein. Nature 394:595–599. doi:10.1038/29106

Kimberly SE, Lara DH (2007) Genome-wide analysis and expression profiling of the small heat shock proteins in zebrafish. Gene 403:60–69. doi:10.1016/j.gene.2007.08.003

Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994) Cloning of cDNAs for genes that are early responsive to dehydration stress in *Arabidopsis thaliana* L.: identification of three ERDs as HSP cognate genes. Plant Mol Biol 25:791–798. doi:10.1007/BF00028874

Kotak S, Port M, Ganguli A, Bicker F, von Koskull-Doring P (2004) Characterization of C-terminal domains of *Arabidopsis* heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES

motifs essential for activator function and intracellular localization. Plant J 39:98–112. doi:10.1111/j.1365-313X.2004.02111.x

Kuo HF, Tsai YF, Young LS, Lin CY (2000) Ethanol treatment triggers a heat shock-like response but nothermotolerance in soybean (Glycine max cv. KaohsiungNo.8) seedlings. Plant Cell Environ 23:1099–1108. doi:10.1046/j.1365-3040.2000.00621.x

LaFayette PR, Nagao RT, O'Grady K, Vierling E (1996) Molecular characterization of cDNAs encoding low-molecular-weight heat shock proteins of soybean. Plant Mol Biol 30:159–169. doi:10.1007/BF00017810

Landry J, Chretien P, Lambert H, Hickey E, Weber LA (1989) Heat shock resistance conferred by expression of the human *HSP27* gene in rodent cells. J Cell Biol 109:7–15. doi:10.1083/jcb.109.1.7

Lee YRJ, Nagao RT, Lin CY, Key JL (1996) Induction and regulation of heat-shock gene expression by an amino acid analog in soybean seedlings. Plant Physiol 110:241–248

Lenne C, Block MA, Garin J, Douce R (1995) Sequence and expression of the mRNA encoding HSP22, the mitochondrial small heat-shock protein in pea leaves. Biochem J 311:805–813

Lin CY, Roberts JK, Key JL (1984) Acquisition of thermotolerance in soybean seedlings: synthesis and accumulation of heat shock proteins and their cellular localization. Plant Physiol 74:152–160. doi:10.1104/pp.74.1.152

Lindquist S, Craig E (1988) The heat shock proteins. Annu Rev Genet 22:631–677. doi:10.1146/annurev.ge.22.120188.003215

Lubaretz O, zur Nieden U (2002) Accumulation of plant small heat-stress proteins in storage organs. Planta 215:220–228. doi:10.1007/s00425-002-0745-1

Ma C, Haslbeck M, Babujee L, Jahn O, Reumann S (2006) Identification and characterization of a stress-inducible and a constitutive small heat shock protein targeted to the matrix of plant peroxisomes. Plant Physiol 141:47–60. doi:10.1104/pp.105.073841

MacRae TH (2000) Structure and function of small heat shock/alpha-crystallin proteins: established concepts and emerging ideas. Cell Mol Life Sci 57:899–913. doi:10.1007/PL00000733

Maestri E, Klueva N, Perrotta C, Gulli M, Nguyen HT, Marmiroli N (2002) Molecular genetics of heat tolerance and heat shock proteins in cereals. Plant Mol Biol 48:667–681. doi:10.1023/A:1014826730024

Miller G, Mittler R (2006) Could heat shock transcription factors function as hydrogen peroxide sensors in plants? Ann Bot (Lond) 98:279–288. doi:10.1093/aob/mcl107

Moore RC, Purugganan MD (2003) The early stages of duplicate gene evolution. Proc Natl Acad Sci USA 100:15682–15687. doi:10.1073/pnas.2535513100

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145. doi:10.1038/ng1289

Nayidu NK, Wang L, Xie WB, Zhang CJ, Fan CZ, Lian XM, Zhang QF, Xiong LZ (2008) Comprehensive sequence and expression profile analysis of PEX11 gene family in rice. Gene 412:59–70. doi:10.1016/j.gene.2008.01.006

Nover L, Bharti K, Doring 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 Chaperones 6:177–189. doi:10.1379/1466-1268(2001)006<0177:AATHST>2.0.CO;2

Nuruzzaman M, Gupta M, Zhang CJ, Wang L, Xie WB, Xiong LZ, Zhang QF, Lian XM (2008) Sequence and expression analysis of the thioredoxin protein gene family in rice. Mol Genet Genomics 280:139–151. doi:10.1007/s00438-008-0351-4

Prestridge DS (1991) SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. CABIOS 7:203–206

Sabehat A, Weiss D, Lurie S (1996) The correlation between heat shock protein accumulation and persistence and chilling tolerance in tomato fruits. Plant Physiol 110:531–537. doi:10.1104/pp.110.2.531

Sabehat A, Lurie S, Weiss D (1998) Expression of small heat-shock proteins at low temperatures. Plant Physiol 117:651–658. doi:10.1104/pp.117.2.651

Sachin K, Elizabeth V, Helmut B, von Pascal KD (2007) A novel transcriptional cascade regulating expression of heat stress proteins during seed development of *Arabidopsis*. Plant Cell 19:182–195. doi:10.1105/tpc.106.048165

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

Sato Y, Yokoya S (2008) Enhanced tolerance to drought stress in transgenic rice plants overexpressing a small heat-shock protein, sHSP17.7. Plant Cell Rep 27:329–334. doi:10.1007/s00299-007-0470-0

Sato Y, Murakami T, Funatsuki H, Matsuba S, Saruyama H, Tanida M (2001) Heat shock-mediated APX gene expression and protection against chilling injury in rice seedlings. J Exp Bot 52:145–151. doi:10.1093/jexbot/52.354.145

Scharf KD, Siddique M, 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 Chaperones 6:225–237. doi:10.1379/1466-1268(2001)006<0225:TEFOAT>2.0.CO;2

Schöffl F, Key JL (1982) An analysis of mRNAs for a group of heat shock proteins of soybean using cloned cDNAs. J Mol Appl Genet 1:301–314

Schramm F, Ganguli A, Kiehlmann E, Englich G, Walch D, von Koskull-Döring P (2006) The heat stress transcription factor HsfA2 serves as a regulatory amplifier of as subset of genes in the heat stress response in *Arabidopsis*. Plant Mol Biol 60:759–772. doi:10.1007/s11103-005-5750-x

Siddique M, Gernhard S, von Koskull-Döring P, Vierling E, Scharf KD (2008) The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. Cell Stress Chaperones 13:183–197. doi:10.1007/s12192-008-0032-6

Stamler R, Kappe G, Boelens W, Slingsby C (2005) Wrapping the [alpha]-crystallin domain fold in a chaperone assembly. J Mol Biol 353:68–79. doi:10.1016/j.jmb.2005.08.025

Sung DY, Kaplan F, Lee K-J, Guy GL (2003) Acquired tolerance to temperature extremes. Trends Plant Sci 8:179–187. doi:10.1016/S1360-1385(03)00047-5

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. doi:10.1186/1471-2164-8-125

Takemoto L, Emmons T, Horwitz J (1993) The C-terminal region of alpha-crystallin: involvement in protection against heat-induced denaturation. Biochem J 294:435–438

Tseng TS, Tzeng SS, Yeh KW, Yeh CH, Chang FC, Chen YM, Lin CY (1993) The heat-shock response in rice seedlings: isolation and expression of cDNAs that encode class I low-molecular-weight heat-shock proteins. Plant Cell Physiol 34:165–168

Tsvetkova NM, Horvath I, Torok Z, Wolkers WF, Balogi Z, Shigapova N, Crowe LM, Tablin F, Vierling E, Crowe JH, Vigh L (2002) Small heat-shock proteins regulate membrane lipid polymorphism. Proc Natl Acad Sci USA 2002:13504–13509. doi:10.1073/pnas.192468399

van Berkel J, Salamini F, Gebhardt C (1994) Transcripts accumu lating during cold storage of potato (*Solanum tuberosum* L.) tubers are sequence related to stress-responsive genes. Plant Physiol 104:445–452. doi:10.1104/pp.104.2.445

van Montfort RLM, Slingsby C, Vierling E (2001a) Structure and function of the small heat shock protein/a-crystallin family of molecular chaperones. Adv Protein Chem 59:105–156. doi:10.1016/S0065-3233(01)59004-X

van Montfort RLM, Basha E, Friedrich KL, Slingsby C, Vierling E (2001b) Crystal structure and assembly of a eukaryotic small heat shock protein. Nat Struct Biol 8:1025–1030. doi:10.1038/nsb722

Vierling E (1991) The roles of heat shock proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 42:579–620. doi:10.1146/annurev.pp.42.060191.003051

Waters ER, Lee GJ, Vierling E (1996) Evolution structure and function of the small heat shock protein in plants. J Exp Bot 47:325–338. doi:10.1093/jxb/47.3.325

Wehmeyer N, Vierling E (2000) The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests general protective role in desiccation tolerance. Plant Physiol 122:1099–1108. doi:10.1104/pp.122.4.1099

Wollgiehn R, Neumann D (1995) Stress response of tomato cell cultures to toxic metals and heat shock: differences and similarities. J Plant Physiol 146:736–742

Yeh CH, Chang PFL, Yeh KW, Lin WC, Chen YM, Lin CY (1997) Expression of a gene encoding a 16.9-kDa heat-shock protein, *Oshsp16.9*, in *Escherichia coli* enhances thermotolerance. Proc Natl Acad Sci USA 94:10967–10972. doi:10.1073/pnas.94.20.10967

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C et al (2005) The Genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3(2):e38. doi:10.1371/journal.pbio.0030038

Zhang QF (2007) Strategies for developing Green Super Rice. Proc Natl Acad Sci USA 104:16402–16409. doi:10.1073/pnas.0708013104

Zhang QF, Li JY, Xue YB, Han B, Deng XW (2008) Rice 2020: a call for an international coordinated effort in rice functional genomics. Mol Plant 2008:1–5