Complexity of rice Hsp100 gene family: lessons from rice genome sequence data

Gaurav Batra, Vineeta Singh Chauhan, Amanjot Singh, Neelam K Sarkar and Anil Grover*
Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110 021, India

**Corresponding author (Email, anil.anilgrover@gmail.com)*

Elucidation of genome sequence provides an excellent platform to understand detailed complexity of the various gene families. Hsp100 is an important family of chaperones in diverse living systems. There are eight putative gene loci encoding for Hsp100 proteins in *Arabidopsis* genome. In rice, two full-length Hsp100 cDNAs have been isolated and sequenced so far. Analysis of rice genomic sequence by *in silico* approach showed that two isolated rice Hsp100 cDNAs correspond to *Os05g44340* and *Os02g32520* genes in the rice genome database. There appears to be three additional proteins (encoded by *Os03g31300*, *Os04g32560* and *Os04g33210* gene loci) that are variably homologous to Os05g44340 and Os02g32520 throughout the entire amino acid sequence. The above five rice *Hsp100* genes show significant similarities in the signature sequences known to be conserved among Hsp100 proteins. While *Os05g44340* encodes cytoplasmic Hsp100 protein, those encoded by the other four genes are predicted to have chloroplast transit peptides.

[Batra G, Chauhan V S, Singh A, Sarkar N K and Grover A 2007 Complexity of rice Hsp100 gene family: lessons from rice genome sequence data; *J. Biosci*. **32** 611–619]

1. Major plant Hsp families

Molecular chaperones are considered to mitigate stress effects in living cells by protecting the native states of the cellular proteins and removing non-native proteins to prevent them from accumulating to excess levels (Miernyk 1999; Hartl and Hayer-Hartl 2002; Ellis 2006). Hsps/chaperones are known to be expressed in plants not only when they experience high temperature stress but also in response to a wide range of other environmental stresses, such as water stress, salt stress, cold stress and oxidative stress (Vierling 1991; Waters *et al* 1996). In plants, major classes of Hsps that act as chaperones include Hsp100 (Clp family), Hsp90, Hsp70 (DnaK family), Hsp60 (chaperonins; GroEL) and low molecular weight (LMW)-Hsps (in the range of 16-42 kDa; also called sHsps) (Nover 1991; Narberhaus 2002). However, except for the LMW-Hsp family, Hsps in plants have received relatively less attention (Hong *et al* 2003; Wang *et al* 2004). Hsp70 proteins are the predominant forms of chaperones expressed under high temperature stress conditions: Hsp70 chaperones, along with their co-chaperones (e.g. DnaJ/Hsp40 and GrpE) make up cellular machines that interact with a wide range of proteins in almost all cellular compartments. *Arabidopsis* genome contains at least 18 genes encoding members of this family, of which 14 belong to the DnaK subfamily and four to the Hsp110/SSE subfamily (Wang *et al* 2004). Chaperonins are a class of molecular chaperones found in prokaryotes and in the mitochondria and chloroplasts of eukaryotes. Chaperonins are further classified into two subfamilies: the GroE chaperonins (group I) and the CCT chaperonins [chaperonins containing *t*-complex polypeptide 1 (TCP1); Group II]. Seven *Arabidopsis* genomic sequences have been identified as having the potential to encode plastid Cpn60 proteins (Hill and Hemmingsen 2001); while nine *Arabidopsis* sequences are predicted to encode proteins similar to CCT protein subunits (Hill and Hemmingsen 2001). Hsp90 family is distinct from other molecular chaperones in that most of its known substrates are signal

Keywords. Chaperone, gene family, Hsp100, *Oryza sativa*

transduction proteins such as steroid hormone receptors and signaling kinases (Young *et al* 2001a). In *Arabidopsis* genome, the Hsp90 family includes seven members: AtHsp90-1 to AtHsp90-4 constitute the cytoplasmic subfamily while AtHsp90-5, AtHsp90-6 and AtHsp90-7 are predicted to be localized to the plastid, mitochondria and ER, respectively (Miloni and Hatzopoulos 1997; Krishna and Gloor 2001). Plant sHSPs form a more diverse family than other Hsps based on the sequence similarity, cellular location and functions (Vierling 1991). In *Arabidopsis*, 13 different sHSPs are grouped into six classes based on their intracellular localization and sequence relationships (Scharf *et al* 2001). Hsp100 family of proteins has a wide distribution in both prokaryotes and eukaryotes (Singla *et al* 1998a; Katiyar-Agarwal *et al* 2001). Yeast Hsp104 can be considered as a prototype of this chaperone family. Although close homologs of yeast Hsp104 have been identified in bacteria (ClpB) and in mitochondria (Hsp78), and in the cytosol of plants (Hsp101), reports indicate that animal cells do not have this chaperone or its close homologs in the cytosol (Mosser *et al* 2004; Sherman 2004). cDNAs encoding Hsp100 have been isolated from *Arabidopsis*, soybean, tobacco, rice, maize and wheat plants (Lee *et al* 1994; Schirmer *et al* 1994; Nieto-Sotelo *et al* 1999; Young *et al* 2001b; Agarwal *et al* 2003; Shen *et al* 2003).

2. Signature sequences of plant Hsp100 proteins

Detailed analysis of the amino acid sequence has revealed that Hsp100 members contain several conserved signatures. Schirmer *et al* (1996) divided Hsp100 into class I (Hsp100 types A–D; containing two nucleotide-binding domains) and class II (Hsp100 types M, N, X, and Y; containing one nucleotide-binding domain) proteins. The molecular weights of the Hsp100 proteins is considered to vary between 75 and 100 kDa because of the size of the non-conserved spacers between the 2 domains and additional sequences at the C- and N-termini. Basically, Hsp100 proteins are composed of five specific domains as follows: (i) amino (N) - terminal domain, (ii) nucleotide – binding domain 1 (NBD1), (iii) middle domain, (iv) NBD2 and (v) carboxyl (C) - terminal domain. The conserved sequences of the five domains include (i) signature sequence I in the N-terminal domain, (ii) Walker A, Walker B1, and Walker B2 sequences in the NBD1, (iii) signature sequence II and signature sequence III in the middle domain, (iv) Walker A and Walker B sequences in NBD2 and (v) signature sequence IV and V in the C-terminal domain (Schirmer *et al* 1996). Based on more extensive analysis with larger spectrum of plant species, Agarwal *et al* (2002) reported further specific alterations in the amino acid sequence of various motifs of plant Hsp100 members.

3. Expression characteristics of plant Hsp100 members

Hsp100 transcript/protein expression is induced during heat stress in *Arabidopsis* (Schirmer *et al* 1994), soybean (Lee *et al* 1994), rice (Pareek *et al* 1995; Agarwal *et al* 2003, Shen *et al* 2003), maize (Nieto-Sotelo *et al* 1999; Young *et al* 2001b), tobacco and wheat (Wells *et al* 1998; Campbell *et al* 2001). There are also indications that Hsp100 transcript/protein is induced by desiccation (Shen *et al* 2003), low temperature (Shen *et al* 2003) and ABA (Pareek *et al* 1995). Apart from stress-regulation, Hsp100 transcript and protein are developmentally regulated in plants (Queitsch *et al* 2000). In non-stressed maize, Hsp100 is expressed to a high level in the tassel at the pre-meiosis stage, the ear (including silks) and the developing endosperm and embryo (Young *et al* 2001b).

4. Rice Hsp100 members

Singla and Grover (1993, 1994) noted that heat stress in rice cells cause accumulation of ~100 kDa protein that cross-reacted with antibodies raised against yeast Hsp104. It was further noted that high levels of this protein are primarily localized in the embryonal portion of the seeds in rice (Singla *et al* 1998b). Young *et al* (2001b) isolated rice Hsp100 cDNA (accession number AF332981). Northern analysis carried out using probe made from the 5'UTR of this cDNA showed that the corresponding transcript is strictly heat-inducible and the induction is transient in nature (Agarwal *et al* 2003). In the temperature regimes tested, 45°C treatment to intact rice seedlings for 2 h showed maximal levels of *Hsp100* mRNA. This cDNA corresponded to the gene locus *Os05g44340* in the rice genome database. Rice full-length *Hsp100* cDNA complemented yeast mutant disrupted for its own *Hsp104* gene by insertional mutagenesis (Agarwal *et al* 2003). Shen *et al* (2003) subsequently isolated another full-length Hsp100 cDNA from rice (which contained a 2817 base pairs long ORF encoding a 938 amino acid protein) that belongs to the ClpD subfamily of Hsp100/Clp proteins. Transcript of this gene was shown to be induced by heat shock, cold shock and dehydration stresses. Singla *et al* (1998b) showed that rice Hsp100 protein is constitutively expressed in dry rice grains. Within the grains, this protein predominantly accumulated in embryo. This protein was found to be gradually lost from the grains when seeds were placed for germination.

5. Complexity of *Arabidopsis* Hsp100 gene family

Nuclear genome of *A. thaliana* was the first one to be completely sequenced in plants. Based on the genome

sequence analysis of this species, we earlier showed that *Arabidopsis* genome has eight Hsp100-related gene loci that show significant similarities or identities in the signature sequences conserved among Hsp100 proteins (At1g74310, At4g14670, At5g15450, At2g25140, At5g50920, At3g48870, At5g51070, At5g57710; Agarwal *et al* 2001). Importantly, five out of these eight proteins (namely At4g14670, At5g50920, At3g48870, At5g51070 and At5g57710) were shown to possess predicted plastidial localization signals. This analysis for the first time indicated that Hsp100 members are represented in the genome in a complex manner. While the structural features of different *Arabidopsis* Hsp100 protein were shown using *in silico* analysis by Agarwal *et al* (2001), 'wet' experiments for isolation and functional characterization of these clones are now appearing. Lee *et al* (2006) have recently made a detailed analysis of these loci. Their findings suggest that At4g14670 is mis-annotated as Clp/Hsp100. By fusing the putative transit peptides of At5g15450 (ClpB3) and At2g25140 (ClpB4) with GFP, they show that these proteins are targeted to chloroplast and mitochondria, respectively. Their analysis further reveals that there are two lineages of ClpB protein in plants, a eukaryotic, cytosol/nuclear-localized represented by At1g74310 (AtHsp101) and an organelle-localized lineage, containing both At5g15450 (ClpB3) and At2g25140 (ClpB4). The gene expression profiles recently carried out during high temperature response in *A. thaliana* by microarray show that both At5g15450 and At2g25140 genes are transcribed (Lim *et al* 2006).

6. Complexity of rice Hsp100 gene family

Rice (*Oryza sativa* L.) is the first crop plant for which almost the entire genome sequence is now available (IRGSP, 2005). We analyzed NCBI genome annotation initiative database (*http://www.ncbi.nlm.nih.gov/*), TIGR rice genome database (*http://www.tigr.org/*) and KOME database (*http://cdna01.dna.affrc.go.jp/cDNA/*) for analyzing the rice Hsp100 family members. The derived information was analyzed by a host of different softwares, e.g. DNAsys, Vector NTI and DNASTAR; multiple alignments using MegAlign module of DNASTAR; BlastN and BlastP programmes (Altschul *et al* 1997) for homology profiles; domain search in the protein sequences was carried out using SMART, Pfam and ExPasy softwares available on internet. The question that we have addressed herein is how many ORFs in rice genome are significantly homologous to Os05g44340/ Os02g32520 (encoding OsHsp100) and At1g74310 (encoding AtHsp101)?

Analysis of rice genome by *in silico* approach showed that ORFs encoded by *Os03g31300*, *Os04g32560* and *Os04g33210* gene loci are variably homologous to *Os05g44340*/*Os02g32520* and *At1g74310* over their entire length (notably, *Os05g44340* and *At1g74310* are 84% identical and *Os02g32520* and *At1g74310* are 27.4% identical). The extent of identity of these members with respect to *Os05g44340* is as follows: *Os03g31300* - 48%, *Os04g32560* - 37%, *Os02g32520* - 37% and *Os04G33210* - 32%. With respect to *At1g74310*, extent of identity turns is as follows: *Os03g31300* - 48%, *Os04g32560* - 39%, *Os02g32520* - 27% and *Os04g33210* - 31%. The predicted molecular weights, iso-electric points, ORF length, sub-cellular localization of the encoded proteins and the genomic organization of the five rice Hsp100 homologs are presented in table 1. Specific domains (such as AAA domain, UVR domain, KE2 domain, ClpN domain) present in the above rice proteins (Os02g32520, Os03g31300, Os04g32560, Os04g33210 and Os05g44340), as predicted using SMART database (Letunic *et al* 2006), are shown in figure 1. The genetic relationship of these members is shown in the form of phylogenetic tree in figure 2.

Table 1. Survey of the members of rice Hsp100 family.

No.	Protein/ gene ID[a]	Protein/ gene ID[b]	Chromosome[c]	No. of introns	Amino acids	MW (kDa)	pI	Intracellular localization[d]
1	Os05g44340	AF332981	V	4	912	100.896	6.17	Cytosolic
2.	Os03g31300	XP_468773	III	10	978	108.985	6.56	Plastidial
3.	Os04g32560	XP_472335*	IV	8	918	101.80	6.43	Plastidial
4.	Os02g32520	XP_466044	II	11	938	101.883	7.18	Plastidial
5.	Os04g33210	XP_472386[$]	IV	10	858	93.255	8.51	Plastidial

[a]gene locus on TIGR rice genome database; [b]gene/protein ID on NCBI database; [c]chromosomal localization of the genes; [d]predictions for the intracellular localization, derived with the help of TargetP program. [*]XP_472335 is 30 amino acids smaller than Os04g32560 on N-terminus. XP_472335 coding sequence starts with "ATGATG" which is less possible. XP_472335 has predicted mitochondrial targeting sequence but on the basis of whole sequence homology with other proteins, it looks to be chloroplastic precursor protein. When we analyse Os04g32560 on TargetP, this shows predicted chloroplastic targeting sequence so we take TIGR Os04g32560 as final sequence. [$]XP_472386 has 19 amino acids extra in the middle of sequence in comparison with Os04g33210. We take Os04g33210 as final sequence because this sequence shows proper middle signature sequence. MW, molecular weight. pI- isoelectric point

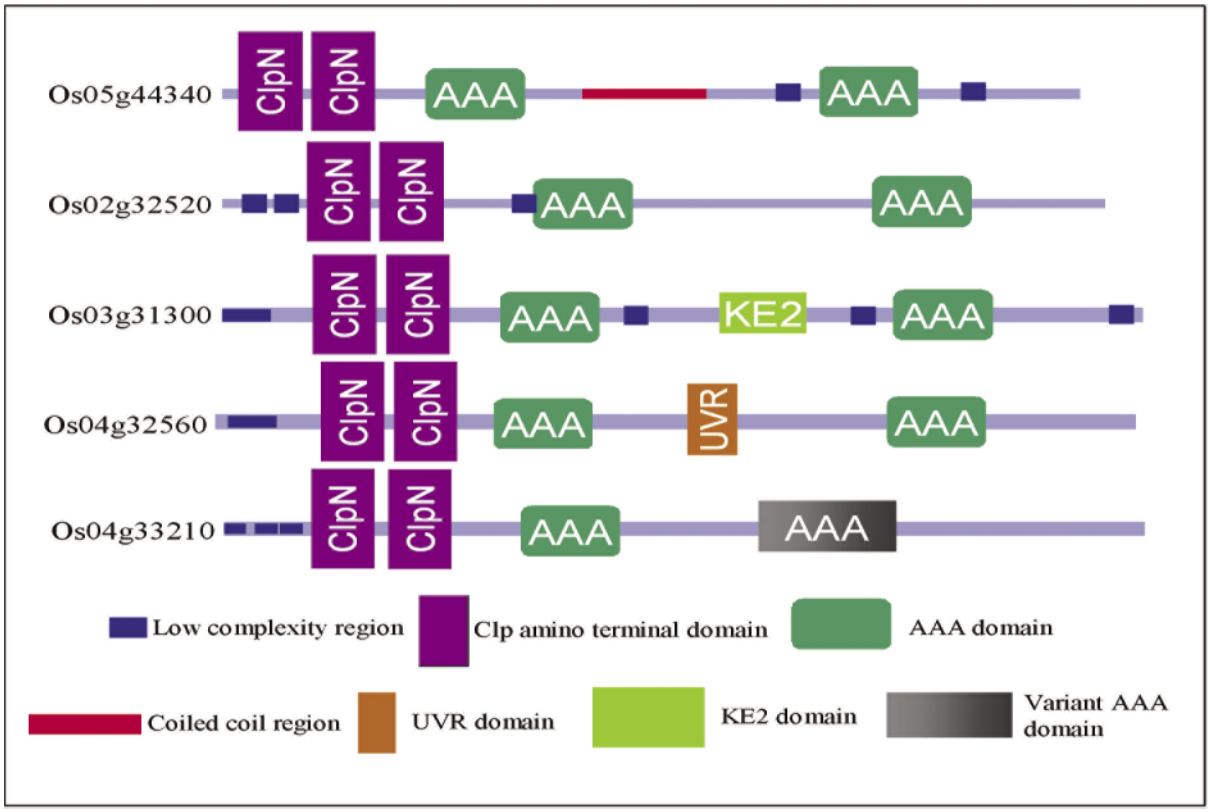

Figure 1. Domain architecture of Os05g44340 homologues as revealed by SMART genomic mode (*http://smart.embl-heidelberg.de*).

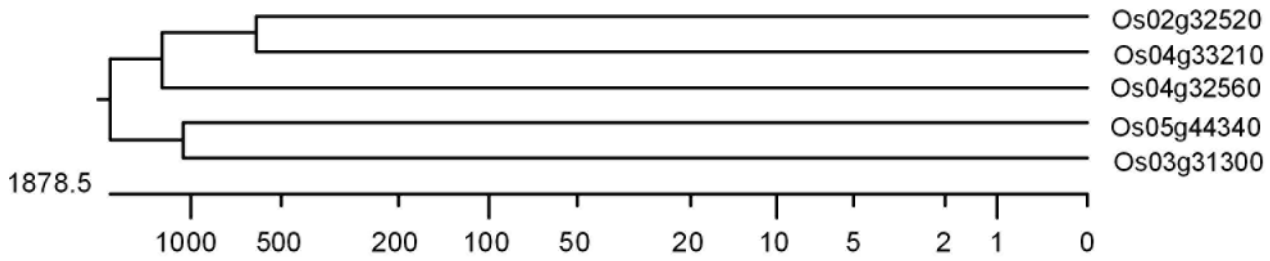

Figure 2. Phylogenetic relationship among various rice Hsp100 homologs. The scale bar shows the correspondence between the unit length and the number of substitution events per position. The phylogenetic tree was constructed using the J Hein method with residue weight set at PAM 250 in the MegAlign module of DNASTAR.

Hsp100 encoded by genomic locus *Os05g44340* is the first species in rice for which the entire cDNA was isolated as well as analyzed experimentally (Young *et al* 2001b, Agarwal *et al* 2003). We therefore coin *OsHsp100-1* term for this gene. Full-length cDNA of Hsp100 reported by Shen *et al* (2003), corresponding to *Os02g32520* gene locus, on the basis of above argument can be considered as *OsHsp100-2* gene. Search in rice EST database showed presence of ESTs corresponding to all five different rice *Hsp100* members (table 2). Most of these ESTs were obtained from libraries made from developing stage or from green portions of the rice plants. This may be because gene libraries sequenced thus far have mainly been made using such tissues. In future course, analysis of ESTs using gene libraries from high temperature stressed rice plants would be a more ideal situation to obtain detailed information on Hsp genes. Nonetheless, it appears from the above account that all five rice *Hsp100* genes may be expressing in rice, and regulation

Table 2. ESTs corresponding to different rice *hsp100* homologs.

Protein/gene ID	C	P	L	IS	WPB	ML	WPT	EL	CS	GS	5-MT
Os05g44340	+(n)	+(n)		+							
Os03g31300	+(n)	+(n)	+(n)	+	+	+(n)	+(n)	+	+		
Os04g32560		+(n)	+(n)		+		+(n)			+(n)	+
Os02g32520	+(n)		+(n)		+(n)	+	+(n)				
Os04g33210		+	+(n)		+	+					

EST libraries at NCBI database were searched. +, one EST found for a given homologue; +(n), more than one ESTs found for a given homologue. EST libraries indicated are as follows: P, panicle; C, callus; L, 3-week-old leaf; IS, immature seed 5 days after pollination; WPB, whole plant (booting); ML, mature leaf; WPT, whole plant (tillering); EL, endosperm library 10 days after anthesis; CS, cold stress germination; GS, green shoot; 5-MT, 5-methyl tryptophan treated leaves, stems and roots

of their expression might be a function of different stress conditions and developmental stages.

For the assignment of an ORF to the Hsp100 family, it is critical to analyse the conservation of Hsp100 signature sequences. Analysis carried out by *in silico* approach showed that the N-terminal signature I is conserved (with some deviation from consensus sequence) in all the five Hsp100 members. NBD1 Walker A and Walker B2 sequences were highly-conserved for all the five members of the rice Hsp100 family (figure 3). The same was found to be true for the NBD1 Walker B1 sequence with the exception of first amino acid residue which was lysine in Os05g44340 and Os03g31300, glutamic acid in Os04g32560 and arginine in Os02g32520 and Os04g33210. Only Os05g44340 and Os03g31300 encoded proteins contained the middle domain spacer signature sequence II (with some deviation from consensus sequence). This signature sequence was lacking in rest of the proteins. The middle domain spacer signature sequence III was found in Os05g44340, Os03g31300, Os04g32560 and Os02g32520 (with some deviation from consensus) but not in Os04g33210. Walker A sequence of NBD2 domain was highly-conserved except that Os04g33210 lacked this consensus. NBD2 Walker B pocket was less conserved than Walker A but was present in all the sequences. C-terminal domain signature sequences IV and V were noted to be present in all the members of the rice Hsp100 family. It thus appears that all five Hsp100 sequences show the typical feature of Clp/Hsp100 proteins. Analysis of Hsp100 orthologs using ScanProsite tool of ExPASy and Pfam HMM database showed that all five Hsp100 members have conserved Clp N-terminus domain, ClpA/B signature I, ClpA/B signature II and ATP/GTP- binding site motif A (P-loop). This analysis further showed that Os04g32560 gene has UVR-B/UVR-C domain which is found in UVR-B and UVR-C proteins which are responsible for repair mechanism of DNA in prokaryotes (Van Houten and Snowden 1993; Moolenaar *et al* 1995). It has been previously shown that a conserved region similar to the UVR domain is also found in the ATP-binding subunit of bacterial and chloroplastic Clp proteases. However, precise role of this domain in Clp proteases remains unknown. As Os04g32560 is predicted to be a chloroplastic protein (table 1), it may be possible that this protein has role in repair mechanism of chloroplastic DNA. Furthermore, all rice Hsp100 homologs were noted to possess AAA family conserved motif. AAA superfamily of ATPases are associated with a wide variety of cellular activities, including membrane fusion, proteolysis and DNA replication (Neuwald *et al* 1999; Sauer *et al* 2004). So far, direct involvement of plant Hsp100 proteins in these biochemical events remains to be shown.

Biology of yeast Hsp104 is extensively worked out (Piper 1993; Lindquist and Kim 1996). According to Parsell *et al* (1994), ScHsp104 plays a key role in acquisition of thermotolerance in yeast primarily by facilitating the dissociation of protein aggregates formed as a result of high temperature. Hsp104 yeast null mutant is viable and defective in induced thermotolerance and shows apparent moderate growth defect on YPD after 20-60 generations, NaCl, lactate and minimal medium. When expressed in the cytosol, Hsp78 can substitute for the homologous Hsp104 in mediating cellular thermotolerance, suggesting a conserved mode of action of the two proteins. Yeast genome database (SGD; Saccharomyces Genome Database) shows single Hsp100 encoding gene (Sanchez and Lindquist 1990). Rice Hsp100 homologs when searched on SGD showed that Hsp104 and Hsp78 of *S. cerevisiae* have reasonably good homology with all the rice Hsp100 members, ranging from 45% observed between Os03g31300 and Hsp78 to 26% observed between Os04g33210 and Hsp104 (table 3).

In our earlier results, Hsp100 protein was seen to be expressed in the crude protein preparations of the heat-shocked rice tissues (Singla and Grover 1993, 1994; Pareek *et al* 1995; Singla *et al* 1998b, Agarwal *et al* 2003). Intracellular localization analysis suggested that Os05g44340 protein appears to be cytosolic as it lacks any distinct targeting sequence. On the other hand, other four Hsp100 homologs are predicted to be chloroplast-localized as they have variable lengths of the plastidial signals. Os03g31300 protein has a predicted plastidial targeting sequence of 76 amino acids. Importantly, this protein

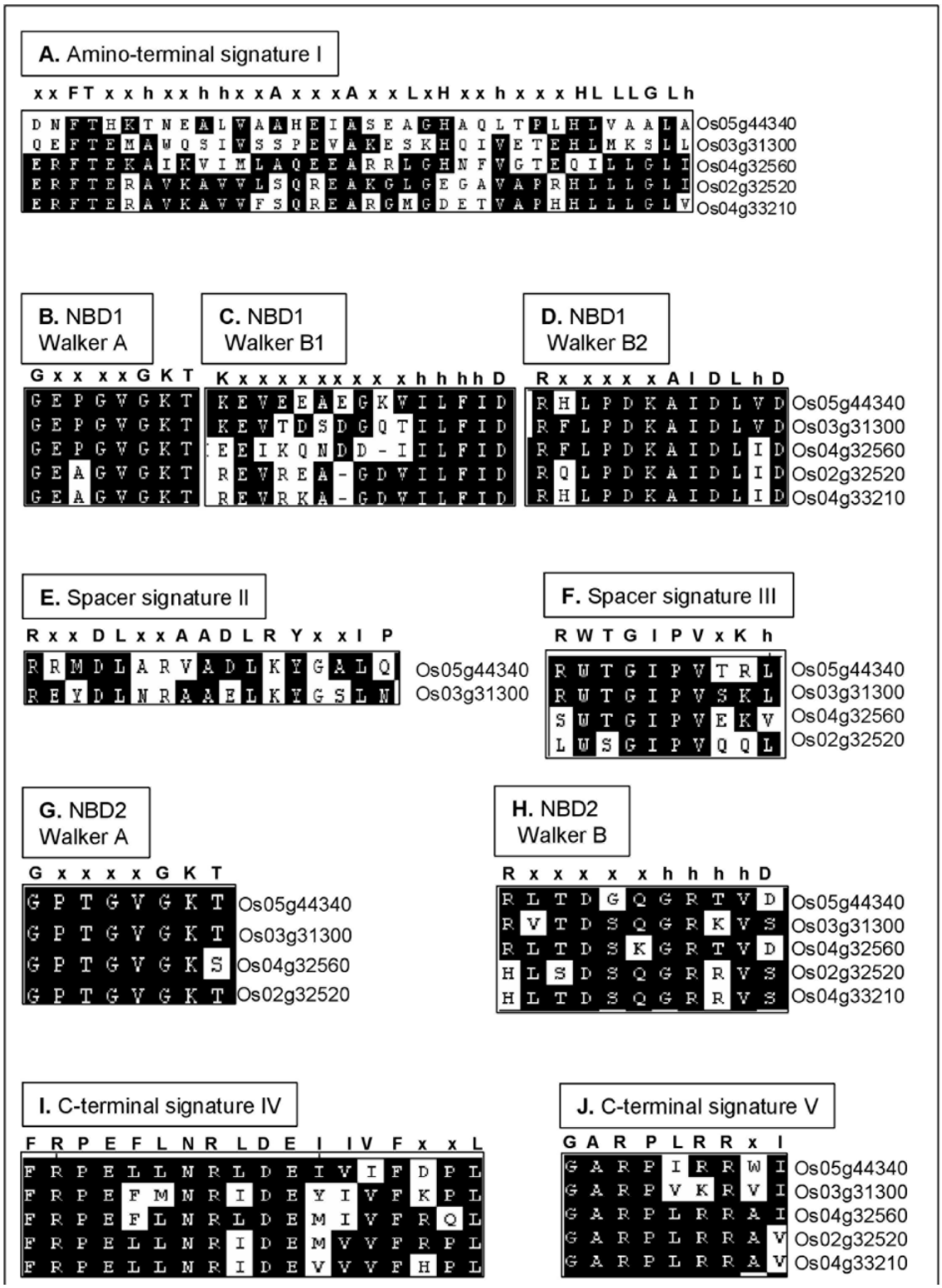

Figure 3. Sequence alignments of the signature sequences present in amino-terminus (A), NBD1 **(B-D)**, spacer domain **(E-F)**, NBD2 **(G-H)** and C-terminus **(I-J)** of rice Hsp100 members. The consensus sequences described by Schirmer *et al* (1996) are indicated at the top of the boxes: large-case letters represent the corresponding amino acid residues in the 1-letter code; x is any amino acid residue; h stands for any of the hydrophobic amino acid residues (i.e. isoleucine, leucine, valine, methionine, phenylalanine, or tyrosine).

Table 3. Rice Hsp100 homolog -corresponding protein(s) in *S. cerevisiae.*

Rice Hsp100 member	Homology to *S. cerevisiae* Hsps
Os05g44340	44% identical with Hsp78, 42%identical with Hsp104
Os03g31300	45% identical with Hsp78, 39% identical with Hsp104
Os04g32560	42% identical with Hsp78, 33% identical with Hsp104
Os02g32520	36% identical with Hsp78, 30% identical with Hsp104
Os04g33210	30% identical with Hsp78, 26% identical with Hsp104

shows 68% identity to the *Phaseolus lunatus* plastidial ClpB (Keeler *et al* 2000). Os02g32520 and Os04g33210 proteins have predicted plastidial targeting sequences of 83 and 80 amino acids respectively (Shen *et al* 2003). The protein corresponding to Os04g32560 has a plastidial targeting sequence of 28 amino acids. BlastP result shows that this protein has significant homology with different chloroplastic proteins such as 84% with tomato chloroplastic endopeptidase (Clp cd4B), 82% with pea chloroplast ClpA homolog, 83% with *A. thaliana* chloroplastic ClpC, 83% with tomato chloroplastic endopeptidase Clp cd4A, 81% with *Brassica napus* chloroplastic ClpA homolog and 82% with *Spinacia oleracea* ClpC. Chloroplasts are dynamic organelles, not only importing ~3000 different nuclear-encoded proteins from cytosol, but also producing ~120 proteins from its own plastome (Leister, 2003). Different Clp proteins are reported to perform important functions in chloroplasts. Plastomic ClpP1 appears to play a role in the degradation of the cytochrome b6/f and PSII complexes in *Chlamydomonas reinhardtii* (Majeran *et al* 2000, Majeran *et al* 2001), and disruption of its expression in tobacco retards chloroplast development and causes ablation of the shoot system (Kuroda and Maliga 2003). Early attempts to repress ClpC expression in tobacco using antisense approach failed to produce viable cell lines with significant decreases in ClpC content (Shanklin *et al* 1995), suggesting ClpC was an essential chloroplastic protein. Similarly, closely-related ClpC in cyanobacteria is also necessary for cell viability and phototrophic growth (Clarke and Eriksson 1996). Majority of ClpC in the stroma is believed to function as a housekeeping enzyme, both in its capacity as an independent molecular chaperone and as the regulatory component of the Clp protease. ClpC has also been implicated in the stromal degradation of aberrant-imported pre-proteins normally targeted to the thylakoid membranes. A small proportion of ClpC associated with the envelope membrane is also thought to act as a molecular chaperone in the import process of cytosolic pre-proteins via the Tic-Toc pathway (Kouranov *et al* 1998). To unveil the importance of ClpC in chloroplasts, Sjogren *et al* (2004) examined the effects of decreased amounts of this protein in *Arabidopsis*. This study demonstrated that the inactivation of the *ClpC1* (*At5g50920*) gene by T-DNA insertion causes a significant drop in total ClpC content in chloroplasts and produces a pleiotropic phenotype that includes retardation in plant growth and development, chlorosis of leaves, impairment to the photosynthetic process, and a specific loss in PSI and PSII content. Constan *et al* (2004) reported that knockout mutant line that contains T-DNA disruption in *AtHSP93-V* (*ClpC1*/ *At5g50920*) in *A. thaliana* is much smaller and paler than wild-type plants. In addition, mutant chloroplasts contain less thylakoid membrane when compared to the wild-type. Os04g32560 is 83% identical with At5g50920, on the basis of homology. It may thus be possible that Os04g32560 also has some role in chloroplast development and import mechanisms.

7. Conclusions

It is shown here that like in *Arabidopsis*, rice Hsp100 family has multiple members. These are the only two plant species that have so far been completely sequenced. It further appears that chaperoning roles of Hsp100 members are in particular important for the functioning of the chloroplasts. As chloroplast is the hub of the cellular activities, Hsp100 members might have some direct/ indirect role in the process of photosynthesis. In future years, it is important to establish the expression characteristics of the different Hsp100 genes under normal growth conditions and under variably stressed conditions. The role of different Hsp100 homolog will become clearer with the possible use of mutational and/or reverse genetics approaches.

Acknowledgements

We are grateful to Department of Science and Technology (SP/SO/A-59/98) and Centre for Plant Molecular Biology, Department of Biotechnology, New Delhi for the financial support. AJS is thankful to Council of Scientific and Industrial Research (CSIR), New Delhi for the fellowship award.

References

Agarwal M, Katiyar-Agarwal S and Grover A 2002 Plant Hsp100 protein: structure, function and regulation; *Plant Sci.* **163** 397–405

Agarwal M, Katiyar-Agarwal S, Sahi C, Gallie D and Grover A 2001 *Arabidopsis thaliana* Hsp100 proteins: kith and kin; *Cell Stress Chap.* **6** 219–24

Agarwal M, Sahi C, Katiyar-Agarwal S, Agarwal S, Young T, Gallie D R, Sharma V M, Ganesan K and Grover A 2003 Molecular characterization of rice Hsp101: complementation of yeast Hsp104 mutation by disaggregation of protein granules and differential expression in *indica* and *japonica* rice types; *Plant Mol. Biol*. **51** 543–553

Altschul S F, Madden T L, Scheffer A A, Zhang J, Zhang Z, Miller W and Lipman D J 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs; *Nucleic Acids Res.* **25** 3389–3402

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

Clarke A K and Eriksson M J 1996 The cyanobacterium *Synechococcus* sp. PCC 7942 possesses a close homologue to the chloroplast ClpC protein of higher plants; *Plant Mol. Biol.* **31** 721–730

Constan D, Froehlich J E, Rangarajan S and Keegstra K 2004 A stromal Hsp100 protein is required for normal chloroplast development and function in *Arabidopsis*; *Plant Physiol*. **136** 3605–3615

Ellis R J 2006 Molecular Chaperones: assisting assembly in addition to folding; *Trends Biochem. Sci.* **31** 395–401

Hartl F U and Hayer-Hartl M 2002 Protein folding- Molecular chaperones in the cytosol: from nascent chain to folded proteins; *Science* **263** 224–227

Hill J E and Hemmingsen S M 2001 *Arabidopsis thaliana* type I and II chaperonins; *Cell Stress Chap.* **6** 190–200

Hong S W, Lee U and Vierling E 2003 *Arabidopsis* hot mutants define multiple functions required for acclimation to high temperatures; *Plant Physiol.* **132** 757–767

International Rice Genome Sequencing Project 2005 The map based sequence of the rice genome; *Nature (London)* **436** 793–800

Katiyar-Agarwal S, Agarwal M, Gallie D R and Grover A 2001 Search for the cellular functions of plant Hsp100/Clp family proteins; *Crit. Rev. Plant Sci.* **20** 277–295

Keeler S J, Boettger C M, Haynes J G, Kuches K A, Johnson M M, Thureen D L, Keeler Jr. C L and Kitto S L 2000 Acquired thermotolerance and expression of the HSP100/ClpB genes of Lima bean; *Plant Physiol.* **123** 1121–1132

Kouranov A, Chen X, Fuks B and Schnell D J 1998 Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane; *J. Cell Biol.* **143** 991–1002

Krishna P and Gloor G 2001 The Hsp90 family of proteins in *Arabidopsis thaliana*. *Cell Stress Chap.* **6** 238–246

Kuroda H and Maliga P 2003 The plastid clpP1 protease gene is essential for plant development; *Nature (London)* **425** 86–89

Lee U, Rioflorido I, Hong S-W, Larkindale J, Waters ER and Vierling E 2007 The Arabidopsis ClpB/Hsp100 family of proteins: Chaperones for stress and chloroplast development; *Plant J.* **49** 115–127

Lee Y J, Nagao R T and Key J L 1994 A soyabean 101-kD heat stress protein complements yeast HSP104 deletion mutant in acquiring thermotolerance; *Plant Cell* **6** 1889–1897

Leister D 2003 Chloroplast research in the genomic age; *Trends Genet.* **19** 47–56

Letunic I, Copley RR, Pils B, Pinkert S, Schultz J and Bork P 2006 SMART5: domains in the context of genomes and networks; *Nucleic Acid Res.* **34** D257–D260

Lim C J, Yang K A, Hong J K, Choi J S, Yun D-J, Hong J C, Chung W S, Lee S Y, Cho M J and Lim C O 2006 Gene expression profiles during heat acclimation in *Arabidopsis thaliana* suspension-culture cells; *J. Plant Res.* **119** 373–383

Lindquist S and Kim G 1996 Heat-shock protein 104 expression is sufficient for thermotolerance in yeast; *Microbiology* **93** 5301–5306

Majeran W, Olive J, Drapier D, Vallon O, Wollman F-A 2001 The light sensitivity of ATP synthase mutants of *Chlamydomonas reinhardtii; Plant Physiol.* **126** 421–433

Majeran W, Wollman F A and Vallon O 2000 Evidence for a role of ClpP in the degradation of the chloroplast cytochrome b(6)f complex; *Plant Cell* **12** 137–149

Miernyk J A 1999 Protein folding in the plant cell; *Plant Physiol*. **121** 695–703

Milioni D and Hatzopoulos P 1997 Genomic organization of Hsp90 gene family in *Arabidopsis*; *Plant Mol. Biol.* **35** 955–961

Moolenaar G F, Franken KL, Dijkstra D M, Thomas-Oates J E, Visse R, van de Putte P, and Goosen N 1995 The C-terminal region of the UvrB protein of *Escherichia coli* contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision; *J. Biol. Chem.* **270** 30508–30515

Mosser D D, Ho S and Glover J R 2004 *Saccharomyces cerevisiae* Hsp104 enhances the chaperone capacity of human cells and inhibits heat stress-induced proapoptotic signaling; *Biochemistry* **43** 8107–8115

Narberhaus F 2002 Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network; *Microbiol. Mol. Biol. Rev.* **66** 64–93

Neuwald A F, Aravind L, Spouge J L, Koonin E V 1999 AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes; *Genome Res.* **9** 27–43

Nieto-Sotelo J, Kannan K B and Segal M C 1999 Characterization of a maize heat-shock protein 101 gene, *HSP101,* encoding a ClpB/ Hsp100 protein homologue; *Gene* **230** 187–195

Nover L 1991 *Heat shock response* (Boca Raton: CRC Press)

Pareek A, Singla S L and Grover A 1995 Immunological evidence for accumulation of two high-molecular-weight (104 and 90 kDa) HSPs in response to different stresses in rice and in response to high temperature stress in diverse plant genera; *Plant Mol. Biol.* **29** 293–301

Parsell D A, Kowal A S, Singer M A and Lindquist S 1994 Protein disaggregation mediated by heat-shock protein Hsp104; *Nature (London)* **375** 475–478

Piper P W 1993 Molecular events associated with acquisition of heat tolerance by the yeast *Saccharomyces cerevisiae*; *FEMS Microbiol. Rev.* **11** 339–359

Queitsch C, Hong S W, Vierling E and Lindquist S 2000 HSP101 plays a crucial role in thermotolerance in *Arabidopsis*; *Plant Cell* **12** 479–492

Sanchez Y and Lindquist S L 1990 HSP104 required for induced thermotolerance; *Science* **248** 1112–1115

Sauer R T, Bolon D N , Burton B M, Burton R E, Flynn J M, Grant R A, Hersch G L, Joshi S A, Kenniston J A, Levchenko I, Neher S B, Oakes E S C, Siddiqui S M, Wah D A and Baker T A 2004 Sculpting the proteome with AAA^+ proteases and disassembly machines; *Cell* **119** 9–18

Scharf K D, Siddique M and Vierling E 2001 The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing α-crystallin domains (ACD proteins); *Cell Stress Chap.* **6** 225–237

Schirmer E C, Glover J R, Singer M A and Lindquist S 1996 HSP100/Clp proteins: a common mechanism explains diverse functions; *Trends Biochem. Sci.* **21** 289–295

Schirmer E C, Lindquist S and Vierling E 1994 An *Arabidopsis* heat stress protein complements a thermotolerance defect in yeast; *Plant Cell* **6** 1899–1909

Shanklin J, Dewitt N D and Flanagan J M 1995 The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: an archetypal two-component ATP-dependent protease; *Plant Cell* **7** 1713–1722

Shen G A, Pang Y Z, Lin C F, Wei C, Qian X Y, Jiang L Z, Du X L, Li K G, Attia K and Yang J S 2003 Cloning and characterization of a novel Hsp100/Clp gene (osClpD) from *Oryza sativa*; *DNA Seq*. **14** 285–293

Sherman M Y 2004 Yeast prions: protein aggregation is not enough; *PLoS Biol.* **4** 436–439

Singla S L and Grover A 1993 Antibodies raised against yeast HSP104 cross-react with a heat- and abscisic acid-regulated polypeptide in rice; *Plant Mol. Biol.* **22** 1177–1180

Singla S L and Grover A 1994 Detection and quantification of a rapidly accumulating and predominant 104 kDa heat stress polypeptide in rice; *Plant Sci.* **97** 23–30

Singla S L, Pareek A and Grover A 1998a Plant Hsp 100 family with special reference to rice; *J. Biosci.* **23** 337–345

Singla S L, Pareek A, Kush A K and Grover A 1998b Distribution patterns of the 104 kDa stress-associated protein of rice reveal its constitutive accumulation in seeds and disappearance from the just-emerged seedlings; *Plant Mol. Biol.* **37** 911–919

Sjogren L L, MacDonald T M, Sutinen S and Clarke A K 2004 Inactivation of the clpC1 gene encoding a chloroplast Hsp100 molecular chaperone causes growth retardation, leaf chlorosis, lower photosynthetic activity, and a specific reduction in photosystem content; *Plant Physiol.* **136** 4114–4126

Van Houten B and Snowden A 1993 Mechanism of action of the Escherichia coli UvrABC nuclease: clues to the damage recognition problem; *Bioessays* **15** 51–59

Vierling E 1991 The roles of heat shock proteins in plants; *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42** 579–620

Wang W, Vincour B, Shoseyov O and Altman A 2004 Role of plant heat shock proteins and molecular chaperones in the abiotic stress response; *Trends Plant Sci* **9** 244–252

Waters E R, Lee G J and Vierling E 1996 Evolution, structure and function of the small heat shock proteins in plants; *J. Exp. Bot.* **47** 325–338

Wells D R, Tanguay R L, Le H and Gallie D R 1998 HSP101 functions as a specific translational regulatory protein whose activity is regulated by nutrient status; *Genes Dev.* **12** 3236–3251

Young J C, Moarefi I and Hartl F U 2001a Hsp90: a specialized but essential protein folding tool; *J. Cell Biol.* **154** 267–273

Young T E, Ling J, Geisler-Lee C J, Tanguay R L, Caldwell C and Gallie D R 2001b Developmental and thermal regulation of the maize heat shock protein, HSP101; *Plant Physiol.* **127** 777–791

*e*Publication: 29 March 2007