Running Title: Mitochondrial GrpE2 confers tolerance to chronic heat stress

Name and address for corresponding author:

Yee-yung Charng

128, Academia Rd. Sec. 2
Agricultural Biotechnology Research Center
Academia Sinica
Taipei, Taiwan 11529
R.O.C.

Tel: 886-2-2651-2238
Fax 886-2-2651-5600

E-mail: yycharng@gate.sinica.edu.tw
Recent Gene Duplication and Subfunctionalization Produced a Mitochondrial GrpE, the Nucleotide Exchange Factor of Hsp70 Complex, Specialized in Thermotolerance to Chronic Heat Stress in Arabidopsis

Catherine Hu¹, Siou-ying Lin², Wen-tzu Chi¹#, and Yee-yung Charng¹,²*

¹Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan, R.O.C.
²Department of Biochemical Sciences and Technology, National Taiwan University, Taipei, Taiwan, R.O.C.
Footnotes:

This work was funded by the Foresight Project (AS-97-FP-L19-1) of Academia Sinica.

# Current address: Agriculture Bureau, Kaohsiung City Government,
No.132, Sec. 2, Guangfu Rd., Fengshan Dist., Kaohsiung City, Taiwan, R.O.C.

*To whom correspondence should be addressed, and the author responsible for
distribution of materials integral to the findings presented in this article in accordance
with the Journal policy described in the Instructions for Authors

(http://www.plantphysiol.org).

E-mail: yycharn@gate.sinica.edu.tw; Fax +886-2-2651-5600
ABSTRACT

Duplication and divergence of heat stress (HS) response genes might help plants adapt to varied HS conditions, but little is known on the topic. Here, we examined the evolution and function of Arabidopsis mitochondrial GrpE (Mge) proteins. GrpE acts as a nucleotide exchange factor in the Hsp70/DnaK chaperone machinery. Genomic data show that AtMge1 and AtMge2 arose from a recent whole-genome duplication event. Phylogenetic analysis indicated that duplication and preservation of Mges occurred independently in many plant species, which suggests a common tendency in the evolution of the genes. Intron-retention contributed to the divergence of the protein structure of Mge paralogs in higher plants. In both Arabidopsis and tomato, Mge1 is induced by UV-B and Mge2 by heat, which suggests regulatory divergence of the genes. Consistently, AtMge2 but not AtMge1 is under the control of HsfA1, the master regulator of the HS response. Heterologous expression of AtMge2 but not AtMge1 in the temperature-sensitive E. coli grpE mutant restored its growth at 43°C. Arabidopsis T-DNA knockout lines under different HS regimes revealed Mge2 specifically required for tolerating prolonged exposure to moderately high temperature, as compared with the need of Hsp101 and Hsa32 for short-term extreme heat. Therefore, with duplication and subfunctionalization, one copy of the Arabidopsis Mge genes became specialized in a distinct type of HS. We provide direct evidence supporting the connection between gene duplication and adaptation to environmental stress.
INTRODUCTION

Temperature is an environmental factor that affects growth and distribution of plants. To be successful in the survival game of nature, plants need to cope with adverse temperatures caused by daily or seasonal temperature fluctuation and by climate change. Understanding how plants adapted to temperature stress can help improve tolerance in crops (Katiyar-Agarwal et al., 2003; Shou et al., 2004).

Plants are often challenged by elevated temperature, which causes heat stress (HS). Knowledge of the plant response to HS is accumulating quickly because of the progress of genomic and functional genomic studies (Kotak et al., 2007). Many genetic components form a complex signaling transduction network and chaperone machineries. Some of these components belong to multi-gene families, most notably heat shock proteins (HSPs) (Vierling, 1991) and heat shock transcription factors (HSFs) (von Koskull-Doring et al., 2007). The HSP and HSF paralogs apparently arose from gene duplications and subsequent diversification. Gene duplication and divergence increased the capacity of organisms to adapt to environmental change (Conant and Wolfe, 2008). In addition to responding to elevated temperature, HS genes are responsive to other environmental changes. Thus, duplication and divergence of HS genes might have allowed plants to adapt to different environmental stresses or varied HS conditions.

However, linking the adaptation and duplication of HS genes is difficult because loss-of-function of one of the duplicates may not lead to an alteration in phenotype such as thermotolerance, which was often attributed to gene redundancy. Mutant phenotypes may not be revealed because the assays are not performed under the right conditions, as was shown with several Arabidopsis HS genes (Charng et al., 2006; Charng et al., 2007; Meiri and Breiman, 2009; Meiri et al., 2010). Thus, determining an effective phenotyping method is critical for revealing the function of HS genes per se and for shedding light on their evolution.
To determine the biological functions of HS response genes, we have used reverse genetic approaches by phenotyping Arabidopsis T-DNA knockout (KO) mutants under varied HS conditions (Charng et al., 2007; Liu et al., 2011). Here, we focused on the function and evolution of the mitochondrial GrpE genes. GrpE is a highly conserved and ubiquitous protein in prokaryotes and eukaryotes. It functions as a co-chaperone in the DnaK/Hsp70 complex, consisting of DnaK (Hsp70), DnaJ (Hsp40), and GrpE. The complex is a chaperone machine involved in the folding of the nascent protein peptides, organellar protein translocation, and disaggregation and refolding of the denatured proteins (Mayer and Bukau, 2005). DnaK and Hsp70 homologs consist of an N-terminal ATPase domain and a C-terminal substrate domain. Current models suggest that the machine operates by propelling DnaK/Hsp70 in cycles of alternate low-affinity (ATP bound) and high-affinity (ADP bound) states for substrates by interaction with co-chaperones, in which nucleotide dissociation catalyzed by GrpE is the rate-limiting step for substrate release (Mayer and Bukau, 2005). In E. coli, both GrpE and DnaK are inducible by heat treatment. Mutants of GrpE and DnaK are sensitive to adverse high temperature (Paek and Walker, 1987; Ang and Georgopoulos, 1989).

In eukaryotes, multiple sets of the Hsp70 chaperone machineries can be found in cytosol and other organellar compartments (Sung et al., 2001). However, only those found in mitochondria and chloroplasts closely resemble their prokaryotic kin. Only one mitochondrial GrpE, Mge1, was found in yeast. Mge1 is not induced by heat shock (Ikeda et al., 1994) but is essential for yeast viability (Bolliger et al., 1994). Mge1 acts in concert with mitochondrial Hsp70s in translocation and maturation of preproteins (Laloraya et al., 1995). Our knowledge of plant GrpEs and their biological role is limited. Two types of GrpE cDNAs were cloned from tobacco BY2 cells and found to encode mitochondrial proteins that interact with mtHsp70 when reintroduced into BY2 protoplasts (Padidam et al., 1999). In Chlamydomonas reinhardtii, the chloroplastic GrpE...
homolog Cge1 is strongly induced by heat shock and forms ATP-sensitive complexes with Hsp70B, the counterpart of bacterial DnaK (Schroda et al., 2001). *In vitro* studies showed that the Hsp70B-Cdj2-Cge1 chaperones facilitate both the assembly and disassembly of the vesicle-inducing protein in plastid oligomers (VIPP1) involved in biogenesis of thylakoid membranes (Liu et al., 2007).

In this paper, we investigated the phylogenetic, regulatory, and functional aspects of duplicated copies of Arabidopsis *Mges*. Subfunctionalization at the expression and functional levels may have allowed Arabidopsis *Mges* to respond more efficiently to distinct HS regimes. We provide direct evidence to support a connection between gene duplication and adaptation to HS.
RESULTS

Arabidopsis mitochondrial GrpE genes were duplicated in a recent whole-genome duplication (WGD) event

The Arabidopsis genome contains 2 homologous genes putatively encoding for mitochondrial GrpEs, which we named Mge1 (At5g55200) and Mge2 (At4g26780). The At4g26780 locus was previously annotated to have unknown function. Corresponding to these co-chaperones are 2 copies of Hsp70 homologs, mtHsc70-1 (At4g37910) and mtHsc70-2 (At5g09590), encoding for mitochondrial DnaK/Hsp70 proteins. The homology between Mge1 and Mge2 suggests that they are derived from gene duplication. The Arabidopsis genome had experienced at least 2 rounds of WGD events that led to massive gene duplication (Blanc et al., 2003). To determine whether the Mges arose from either one of these events, we used the on-line analytic tool Paralogons to examine the location of the Mges in the duplicated regions of the Arabidopsis genome (Blanc et al., 2003). Mge1 and Mge2 each locate in 2 corresponding duplicated blocks formed after the most recent WGD event (Fig. 1). However, their putatively interacting partners, mtHsc70s, came from a more ancient WGD event. Unlike the duplication time for the Mge pair, that for the 2 chloroplastic GrpE genes, Cge1 (At1g36390) and Cge2 (At5g17710), could not be determined, which suggests that they arose from a time before the traceable WGD events. To our surprise, the chloroplastic DnaK/Hsp70 genes, cpHsc70-1 (At5g49910) and cpHsc70-2 (At4g24280), came from the same WGD as the Mges because they locate on the same duplicated blocks (Fig. 1). To confirm this observation, we used molecular clock theory to estimate the time of divergence of each homologous pair by using Arabidopsis genomic synonymous substitution rates (Ks) in a described formula (Lynch and Conery, 2000; Raes et al., 2003). Table 1 shows the Ks and time of divergence of Arabidopsis Mges, Cges, mtHsc70s, and cpHsc70s. The data are in good agreement with the chronicle of the WGD events described (Blanc et al., 2003).
Duplication of Mge genes occurred independently after divergence of many distantly related plant species

With knowledge that Arabidopsis Mge genes arose from a recent WGD event, we investigated a phylogenetic relationship among plant Mges. With a BLAST search, we retrieved and aligned the amino acid sequences homologous to Arabidopsis Mges from various species (Fig. S1). The aligned sequences were used to generate a phylogenetic tree based on the neighbor-joining method (Fig. 2). The tree showed 2 major clades of GrpE proteins, one from mitochondrion and E. coli and another from chloroplast and cyanobacterium, which is in good agreement with the endosymbiotic theory. Two Mge homologs exist in most plants investigated in this study, except soybean, which has 3. The Mges of most plant species, except monocots or close relatives such as tobacco and tomato, clustered together by species, which implies that the Mges duplicated after divergence of the distantly related species and before the divergence of monocots and closely related species. In contrast, their chloroplastic homologs showed a different clustering pattern. Regardless of plant species, except for moss, Cge1 and Cge2 clustered together by protein types (Fig. 2), which suggests that these proteins were derived from duplication before speciation of various plants.

Retention of an in-frame intron is common in higher-plant Mges

Arabidopsis Mge1 contains 6 exons and Mge2 has 5 (Fig. 3). From comparisons of the genomic DNA structures and protein sequences, exon 2 of Mge2 likely contains a retained intron that presumably corresponds to intron 2 of Mge1 (Fig. 3). This putatively retained intron in Mge2 apparently can no longer be processed by spliceosome because of loss of the essential features for RNA splicing and therefore becomes part of the coding sequence. By contrast, intron 2 of Mge1 does not contain an in-frame sequence and would
disrupt the open reading frame if retained. Interestingly, we found retention of the in-frame intron 2 common in alternatively spliced transcripts for 1 of the 2 Mge genes in other higher-plant species (Table 2). As most of the Mge genes listed here are not annotated, we propose that the Mge genes without intron 2 retention to be annotated as Mge1 (or Mge1 and 3 in soybean), and the Mge genes with intron 2 retention to be annotated as Mge2. Notably, the sequences derived from the retained introns are conserved with a motif of 4 to 6 consecutive arginine or lysine residues (Fig. S1). Therefore, retention of the in-frame intron 2 may be under selection pressure in higher plants. By contrast, neither the Mge1 nor Mge2 of the moss Physcomitrella patens contain an in-frame intron 2.

**Mge2 is localized in mitochondria**

Previously, mtHsc70-1, mtHsc70-2, and Mge1 were identified in the mitochondrial proteome of Arabidopsis (Millar et al., 2001; Heazlewood et al., 2004), so these proteins are localized in mitochondria. However, the localization of Mge2 was not known. The presence of a predicted N-terminal mitochondrial targeting peptide (mTP) in Mge2 (Fig. S1) suggests that this protein is also localized in mitochondria. This prediction was further confirmed by transient expression of the Mge2:GFP fusion protein in Arabidopsis protoplasts. We found fusion proteins localized in the punctate foci overlapped those stained by the mitochondria-specific dye, MitoTracker Orange, with the GFP control localized in the cytosol (Fig. 4).

**Mges are differentially expressed**

Duplicate genes usually undergo subfunctionalization at the regulatory and/or protein level (Zhang, 2003). Subfunctionalization at the regulatory level often leads to differential expression of homologous genes. A search of the public microarray database revealed the paralogs Mge1 and Mge2 responsive to UV-B or heat treatment. Mge1 was
upregulated by UV-B and Mge2 preferentially induced by heat treatment at 38°C (Fig. S2).

Analysis of the cis-elements in the promoters showed the Mge1 promoter containing putative UV-B responsive elements, such as UV boxes and MREANAC13’s (Safrany et al., 2008), and the Mge2 promoter containing 1 canonical heat shock element (HSE) (Fig. 3). HSE is the conserved cis-element binding with trimerized HSF for transcription activation of HS genes in eukaryotes (Kroeger et al., 1993). We then tested whether the heat induction of Mge2 requires HsfA1s, the master regulators of HS response in Arabidopsis (Liu et al., 2011; Yoshida et al., 2011). Semi-quantitative RT-PCR showed no induction of Mge2 with heat treatment in the null mutant of HsfA1 genes, which indicates that Mge2 is under the control of HsfA1 (Fig. 5A). However, Mge1 still maintained the basal expression level regardless of HsfA1s. These results agree with the absence or presence of HSE in the promoter regions of the Mges.

To determine whether Mges also differentially express in species other than Arabidopsis, we examined the transcript levels of Mge1 and Mge2 in tomato seedlings with heat and UV-B treatment. The mRNA expression of Mge1 was upregulated in tomato by UV-B treatment, whereas that of Mge2 was induced by heat, as in Arabidopsis (Fig. 5B and 5C). Therefore, the 2 paralogs had undergone a similar subfunctionalization process at the regulation level after the divergence of Arabidopsis and tomato.

**Heterologously expressed Arabidopsis Mge2 confers thermotolerance in heat-sensitive E. coli grpE mutant at higher temperature than does Mge1**

Eukaryotic GrpE homologs, such as the distantly related Cge, can replace the function of E. coli GrpE (Deloche et al., 1997; Schroda et al., 2001). Because Arabidopsis Mges are differentially expressed under HS, we wondered whether Mge1 and Mge2 could replace the function of bacterial GrpE in different capacities. We performed a functional complementation assay with the E. coli grpE mutant (DA16) (Ang and Georgopoulos,
The mutant strain carries a point mutation that substitutes a conserved glycine at position 122 with aspartate in GrpE (Harrison et al., 1997; Grimshaw et al., 2005). The mutant line exhibits a temperature-sensitive phenotype, with growth abolished at 40°C (Fig. 6). We ligated Arabidopsis Mge1 and Mge2 cDNAs without the coding sequence of the predicted mTPs onto an expression vector for transformation into the mutant line to determine whether the temperature-sensitive phenotype could be rescued. The mTPs predicted as putative targeting peptides differed significantly in length in Mge1 and Mge2, with N-terminal 58 and 38 amino acids, respectively, by both TargetP (Emanuelsson et al., 2007) and MitoProt (Claros and Vincens, 1996). However, because the 20 amino acid difference is in a relatively conserved region, part of the functional region of the protein might be removed along with the stretch of peptide sequence. Therefore, we created constructs for each gene, with one set of constructs having 38 amino acids from the N terminus removed from both Mge1 and Mge2 (designated the “long form,” Fig. S1), and another set with 54 or 58 amino acids removed from the N terminus of both proteins (designated the “short form”).

At control temperature (30°C), all lines, whether Mge-transformed or empty vector-transformed DA16 lines, grew similar to the wild-type strain DA15. When grown at a higher temperature (40°C), DA16 with the empty vector showed the temperature-sensitive phenotype and failed to grow. However, except for the Mge transformed with the long-form Mge1, the Mge-transformed lines were able to grow as normally as the wild type, which suggests that the Arabidopsis proteins and the prokaryotic homologs could function similarly. Nevertheless, at 43°C, the 2 Mge-transformed lines showed different behavior. DA16 transformed with Mge2, whether the short or long form, could at least in part complement the mutation, whereas Mge1 could not, and it showed abolished growth, like DA16 transformed with the empty vector (Fig. 6). The difference between Mge1 and Mge2 was not due to the level of the
recombinant proteins expressed in *E. coli*, because the thermotolerance level was not significantly affected with or without the expression inducer anhydrotetracycline (data not shown). Therefore, Arabidopsis Mge2 could function at higher temperatures than Mge1.

**Mge2 specifically confers thermotolerance to chronic HS in Arabidopsis**

Because *Mge2* was highly induced by elevated temperature, we wondered whether it was required for thermotolerance in Arabidopsis. We characterized 2 homozygous T-DNA KO lines for *Mge2*; the T-DNA KO line for *Mge1* was not available during this work. The locations of the T-DNA insertion are in Figure 3 and were confirmed by PCR analysis of the mutants’ genomic DNA (Fig. S3A). RT-PCR analysis confirmed that these mutants no longer generated transcripts like the wild type (Fig. S3B), so the lines were null mutants. Under normal conditions, the mutants did not differ in phenotype from the wild type, which suggests that Mge2 is not essential for growth and development.

Seedlings of the mutants then underwent treatment with 4 different HS regimes to measure the capacity for basal thermotolerance (BT), short-term acquired thermotolerance (SAT), long-term acquired thermotolerance (LAT), and thermotolerance to moderately high temperature (TMHT), and their survival rates were measured. The first 3 assays involved an extremely high temperature, at 44°C, and the condition for TMHT mimicked chronic HS. The assay findings for SAT differed from that of LAT in recovery time between the acclimation treatment and acute HS challenge (Fig. 7A). Previously, we showed that Hsa32 is required for LAT but not SAT, whereas Hsp101 is required for both (Charng et al., 2006; Charng et al., 2007) as controls. Interestingly, the two KO mutants for *Mge2*, *mge2-1* and *mge2-2*, were defective in TMHT but not the other 3 types of thermotolerance (Fig. 7A and 7B). By contrast, the *hsp101* and *hsa32* mutants showed defective phenotypes in at least 2 of the assay conditions for BT, SAT, and LAT, without substantial alteration in TMHT (Fig. 7A and 7B).
To further examine the function of Mge2 in TMHT, we assessed mRNA and protein expression patterns in seedling samples collected at different times during heat treatment. The mRNA expression of Mge2 and other HSP genes, including Hsp101 and Hsa32, was upregulated throughout the chronic HS (Fig. 8A). To detect Arabidopsis Mge2 protein, we raised rabbit antiserum against Mge2 recombinant protein. The antibody recognized a protein band of about 38 kD in the wild type but not the mge2-2 mutant (Fig. 8B). The size determined by migration on SDS-PAGE was the same as for the short-form Mge2 expressed in E. coli described earlier, which suggests that the N-terminal 54 amino acids in the preprotein sequence of Mge2 is the mTP (Fig. S1). With the TMHT assay, the Mge2 protein level gradually increased and remained high after 9 d (Fig. 8B). By contrast, the other HSP levels were high at d 1 but gradually decreased during treatment and was barely detectable at d 9.

**DISCUSSION**

In this study, we show that in Arabidopsis, the 2 mitochondrial GrpE proteins arose from a recent WGD event, estimated to occur about 24 to 75 million years ago (mya) (Lynch and Conery, 2000; Simillion et al., 2002; Blanc et al., 2003). This finding is consistent with our estimated divergence time for the Mges, about 57 mya (Table 1). Intriguingly, our phylogenetic analysis revealed similar independent duplication events for the Mge genes in other higher plants (Fig. 2). Therefore, these Mge paralogs also likely emerged from recent WGD events, which were found to occur in many plant species around the Cretaceous-Tertiary (KT) boundary about 65 mya (Fawcett et al., 2009). However, we could not exclude that segmental duplication instead of WGD is the cause for some species. Nevertheless, this coincidence implies that selection acted to preserve both copies of the Mge genes in these distantly related species. Preservation of the 2 Mge genes could have conferred tolerance to a wide range of...
environmental conditions. However, retaining both Mge copies might cause dosage imbalance because overexpression of GrpE caused a defect in the DnaK chaperone system in E. coli (Sugimoto et al., 2008). Subfunctionalization at the regulatory level could alleviate the problem of dosage imbalance (Veitia et al., 2008). The RT-PCR and microarray data revealed Arabidopsis Mges differentially expressed under UV-B and HS, which is consistent with the presence of associated cis-elements in the promoters of corresponding genes (Fig. 3). We suspected that for certain eudicots, the promoter of the ancestral Mge contained all of the cis-elements required for both types of stress responses, because the tomato Mges were differentially expressed similar to the Arabidopsis genes (Fig. 5B). As time elapsed from gene duplication, the cis-elements were differentially lost in the 2 promoters, probably to optimize response to HS or UV-B. This inference agrees well with the degenerative complementation model of evolution in regulatory elements after gene duplication (Papp et al., 2003; Li et al., 2005) and with the rapid expression divergence of duplicate genes in response to external changes in Arabidopsis (Ha et al., 2007). Here, we provide genetic evidence that Mge2 is involved in thermotolerance, specifically TMHT. However, the role of Mge1 in UV-B stress tolerance remains to be elucidated.

Besides expression divergence, the functions of Mge1 and Mge2 proteins were also subfunctionalized. The Mge1 but not Mge2 is present in the mitochondrial proteome under a non-HS condition in cultured cells (Heazlewood et al., 2004), which suggests that Mge1 is responsible for crucial housekeeping functions such as protein import (Truscott et al., 2003). We cannot exclude the role of Mge1 in HS tolerance. Mge1 may be required for basal or acquired thermotolerance, in that it could be involved in protein folding during recovery at ambient temperature. Unfortunately, the T-DNA KO line for Mge1 was not available at the time of the study, which might be due to it being essential for viability like its counterpart in yeast and E. coli (Ang and Georgopoulos, 1989; Laloraya et al., 1994).
The functional complementation assay of the *E. coli grpE* mutant showed the 2
Arabidopsis *Mge* with different temperature ranges required for their functions, which is
consistent with results *in planta*. However, that *Mge1* can complement the heat-sensitive
phenotype at 40°C in the *E. coli* DA16 mutant (Fig. 6) but is not sufficient to sustain
growth at 35°C in Arabidopsis *Mge2* KO mutants is intriguing. One reason may be that the
thermostability or function of the protein is greatly affected by the milieu, that is, *Mge1*
might be less stable or less efficient at 35°C in mitochondria than in bacteria cells.
Alternatively, the interaction between *Mge1* and bacterial DnaK might be more efficient
than that between *Mge1* and mitochondrial Hsp70s at 35°C. *In vitro* experiments may help
examine the latter possibility.

A closer examination of the genomic sequences of the duplicated *Mges* in
Arabidopsis and comparison with their homologs in other plant species strongly suggests
that the two *Mges* in Arabidopsis might have emerged from a single ancestral *Mge* that
could produce transcripts with 2 alternative splice (AS) forms (Table 2). However,
following duplication, both Arabidopsis *Mge* genes lost the ability to generate AS
transcripts. Sequence alignment (Fig. S1) revealed that the most distinctive difference
between the Arabidopsis *Mge1* and *Mge2* protein sequences is a 30 amino acid stretch
encoded by the putatively retained *intron* 2 of *Mge2* (Fig. 3). This stretch is located in a
highly diverse region near the N-terminal end of the long α helix (Fig. S1) from the crystal
structure of *E. coli* GrpE (Harrison et al., 1997). The N-terminal 33 amino acids of *E. coli*
GrpE corresponding to this region were found to be unstructured but involved in substrate
release from the DnaK (Harrison et al., 1997; Brehmer et al., 2004). Given the
conservation of these intron-derived peptide sequences, the intron-retention may affect the
co-chaperone function of Mge. Further studies of the effect of the absence or presence of
the in-frame *intron* 2 in Arabidopsis and other plant species with use of the Arabidopsis
*Mge2* KO lines, TMHT assay, and *E. coli* complementation assay established in this study
should provide better insights. Of note, temperature-dependent alternative splicing generates Cge isoforms in *Chlamydomonas* (Willmund et al., 2007). However, the biological function of the AS forms is also not clear.

The observation that Mge2 is involved in thermotolerance of chronic HS but not extreme HS is the first evidence for HSPs characterized to date. The role of Mge2 in TMHT agrees well with the expression induced by heat treatment at transcript and protein levels. Mge2 might be required for import or folding mitochondrial proteins under moderately high temperature. Of note, one recently identified mutant of Arabidopsis, hit2, which encodes a nuclear transport receptor, was found impaired in growth at moderately high temperature (37°C 4d) and defective in BT but not SAT or LAT (Wu et al., 2010). However, Hit2 is not responsive to heat treatment. By contrast, Hsp101 is essential for basal and acquired thermotolerance but not TMHT, which is consistent with previous findings (Queitsch et al., 2000). These observations indicate that plants use different components for diverse HS conditions. Interestingly, despite induction at the transcription level for all HSP genes tested under the TMHT condition, several HSPs, including Hsp101, could not be sustained at the protein level, whereas the level of Mge2 substantially increased with time (Fig. 8B). This phenomenon suggests that the HSPs were under translational or posttranslational control during prolonged exposure to moderately high temperature. Thus, inferring the function of a gene under different HS conditions based on its transcript level is difficult. Investigating differential regulation of HSPs at the translational or post-translational level is needed to better understand the HS response.

The differential evolution of mitochondrial and chloroplastic GrpEs and Hsp70s in Arabidopsis is of interest. Mges arose from a recent WGD and mtHsc70s from an ancient duplication event, and vice versa for their chloroplastic counterparts (Table 1). With the increasing number of plant genomes deciphered, comparison of the evolution of the organellar Hsp70 complexes would help determine how common this phenomenon is in
plants. Results from this study can provide basic information regarding the evolution of the chaperone machineries. Of note, 2 copies of Mges were found in rodents and human (Naylor et al., 1998; Oliveira et al., 2006). How and why 2 Mges evolved in these organisms is unknown.

Recent progress in knowledge of chloroplast Hsp70s and co-chaperones have shown that the Hsp70 chaperone system is involved in protein translocation into chloroplasts (Shi and Theg, 2010; Su and Li, 2010), development and thermotolerance (Su and Li, 2008), and biogenesis of thylakoid membranes by regulating the assembly state of VIPP1 (Liu et al., 2007). Studies of mitochondrial Hsp70 systems are beginning to reveal the importance of the system in plants. We are interested to know whether mtHsc70-1 and mtHsc70-2 also have different roles in thermotolerance as do their co-chaperones, Mges.

CONCLUSIONS

The duplication of Mges occurred independently in many plant species, which suggests a common tendency in evolution of the genes. Subfunctionalization of Arabidopsis Mges at both protein and regulatory levels allowed Mge2 to specialize in tolerance of long-term exposure to moderately high temperature. Our data suggest that duplication and subfunctionalization of a highly conserved HSP such as Mge constituted a critical evolutionary adjustment in plants to cope with distinct types of HS and to better adapt to a changed environment.
MATERIALS AND METHODS

Phylogenetic analysis

Sequences of Mges and Cges homologs were found by using the Arabidopsis Mge2 as query sequence in a TBLASTN search of the databases Dana-Farber Cancer Institute—Plant Gene Indices (DFCI; http://compbio.dfci.harvard.edu/tgi/plant.html), National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/), and Phytozome (http://www.phytozome.net/). The sequences were aligned by use of the ClustalW program in MEGA5 (Tamura et al., 2011). The aligned sequences (see Fig. S1 for the section of sequences used) were then used to generate a phylogenetic tree by the neighbor-joining method.

The determination of Arabidopsis paralogous genes and their duplication events involved the online program Paralogons (http://wolfe.gen.tcd.ie/athal/dup). To estimate the time of divergence of the 2 duplicate genes, DnaSP was used to first determine the synonymous substitution rate, $K_s$, of each gene pair (Librado and Rozas, 2009). Then the resulting $K_s$ was used in the formula $T=K_s/2\lambda$, with $\lambda =6.1\times10^{-9}$ as the average whole-genome $K_s$ of Arabidopsis (Lynch and Conery, 2000; Raes et al., 2003).

In silico analyses of gene expression and cis-elements of promoter

Transcript expression data for Arabidopsis $Mge1$ and $Mge2$ under abiotic stress conditions were obtained from the public microarray database maintained in Arabidopsis eFP Browser, http://www.bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi (Winter et al., 2007). Details for the data of HS and UV-B can be found in the browser by choosing the data source “abiotic stress.” The consensus sequences for UV-B regulatory elements and HSEs were obtained from (Safrany et al., 2008) and (Nover et al., 2001), respectively. Sequences about 1,100 bp upstream of ATG site was considered the promoter region. The elements were found by using the search function of Vector NTI 9.1.0 (Invitrogen).
Plant materials and growth condition

The Mge2 T-DNA insertion lines SALK_075614 (mge2-1) and SALK_082197 (mge2-2) were obtained from the Arabidopsis Biological Resource Center (ABRC; Ohio State University). The mutants are derived from ecotype Columbia-0 (Alonso et al., 2003). Homozygous lines of the mutant allele were identified by PCR analysis as described (Charng et al., 2007). The T-DNA KO mutants of Hsp101, Hsa32, and HsfA1s were described previously (Liu et al., 2011). Seeds were sown on 0.5x Murashige and Skoog (MS) medium plates containing 0.1% or 1% sucrose. The sown seeds were imbibed for 3 d at 4°C in the dark before being allowed to germinate and grow at 22°C with 16 h of light (120 μmol m⁻² s⁻¹).

Determining subcellular localization of Mge2

The full-length coding sequence (CDS) of Mge2 (GenBank Accession: NM_118812) was amplified by PCR from cDNA template and cloned onto the entry vector pCR8/GW/TOPO (Invitrogen), then sequenced to confirm no missense or nonsense mutation. The cDNA was cloned onto pMDC83 (Curtis and Grossniklaus, 2003) by use of LR clonase (Invitrogen) to generate a recombinant DNA encoding a recombinant protein of green fluorescence protein (GFP) fused to the C terminus of Mge2. The finished construct, driven by the CaMV35S promoter, was transiently expressed in Arabidopsis mesophyll cell protoplasts as described (Wu et al., 2009). The transformed protoplasts were then stained with the mitochondria-specific probe MitoTracker Orange (Invitrogen). The GFP and MitoTracker Orange signals were observed under a Zeiss LSM 510 Meta confocal microscope.

Heat and UV-B treatments
Five-d-old Arabidopsis seedlings were subjected to heat treatment at 37°C for 1 h as described (Charng et al., 2006). Seven-d-old *Solanum lycopersicum* cv. *Microtom* seedlings sown in Magenta boxes containing 0.5x MS medium, 1% sucrose were subjected to heat treatment at 40°C for 2 h.

For UV-B treatment, 7-d-old Arabidopsis or tomato seedlings were exposed to 150 mJ/cm² UV-B by using Spectronics X1000 UV-B crosslinker (Spectronics Corp.).

**Semi-quantitative analysis of mRNA**

Seedling samples were collected and immediately incubated in liquid nitrogen after treatment. Total RNA was harvested by the TriZol method (Invitrogen) followed by chloroform purification. RNA was reversely transcribed into cDNA by use of oligo-dT primers and M-MLV reverse transcriptase (Promega). PCR involved AmpliqonIII Taq polymerase mix with specific primers as described (Charng et al., 2007). Primer sequences are in Table S1.

**Functional complementation of heat-sensitive *E. coli* mutant containing defective GrpE**

Primers were designed to remove the putative mTPs predicted by TargetP v1.0 (http://www.cbs.dtu.dk/services/TargetP/) and MitoProt (http://ihg.gsf.de/ihg/mitoprot.html) from Arabidopsis Mge1 and Mge2 protein sequences. The CDS corresponding to the first 54-58 or 38 amino acids from the N terminus was removed from both Mge1 and Mge2 as mitochondrial targeting sequences. The Mge1 and Mge2 CDS for both species were amplified by PCR, cloned onto pASK3 plus expression vector (IBA BioTAGnology) and transformed into DA16 cells (Coli Genetic Stock Center, Yale University). The transformed DA16 lines were grown in Luria-Bertani (LB) medium containing ampicillin (100 μg/mL) at 30°C to about OD₆₀₀ 0.8. The liquid cultures then
were 10-fold serially diluted, and 4 μL of each dilution was dropped onto LB plate containing ampicillin. The plates were grown in growth chamber at 30°C or at 40, 43, or 45°C in a waterbath for overnight. To confirm the identities of the expressed recombinant proteins, the protein bands induced by anhydrotetracycline underwent LC-ESI-MS/MS analysis after trypsin digestion, performed by the Proteomics Core Lab of the Institute of Plant and Microbial Biology/Agricultural Biotechnology Research Center, Academia Sinica.

**Thermotolerance assays**

SAT, LAT, BT, and TMHT assays were as described (Liu et al., 2011) with slight modification. For SAT assay, 5-d-old Arabidopsis seedlings were first acclimated with 37°C heat treatment for 1 h. After a 2-h recovery at 22°C, the seedlings then were challenged with acute HS at 44°C for 150 min. For LAT assay, 5-d-old Arabidopsis seedlings were acclimated to 37°C heat treatment for 1 h, then underwent 2-d recovery at 22°C. The seedlings then were challenged at 44°C for 50 min. For BT assay, 4-d-old seedlings were challenged with 44°C for 23 min without prior acclimation. All phenotypes were assessed after 10-d recovery at 22°C post-HS. For TMHT assay, Arabidopsis seeds were sown on 0.5x MS medium plates containing 0.1% sucrose. Five-d-old seedlings were then transferred to a growth chamber with temperature 35±0.3°C during 16 h of light (110 μmol m² s⁻¹) and 33.5±0.3°C during 8 h of darkness. The seedlings then recovered at 22°C for 10 d before phenotypes were assessed.

**Immunoblotting**

The full-length Arabidopsis Mge2 recombinant protein with a C-terminal tag of 6 histidine residues was produced and purified from *E. coli* cells as described (Charng et al., 2006). Immunization and serum collection were performed by a commercial service (LTK
Biotechnology, Taoyuan, Taiwan). The antibodies against Hsp101, Hsa32, sHsp-CI and tubulin were as described (Charng et al., 2006; Chi et al., 2009). The total protein of plant samples was extracted as described (Charng et al., 2006). The mitochondria enriched protein fraction (MEF) was extracted with 30 mM MOPS (pH7.5) buffer with 0.35 M mannitol and 2 mM EDTA. The lysate was first centrifuged at 4,500 x g for 5 min at 4°C to pellet the undesired crude fraction containing chloroplasts and other cell debris, then the supernatant was centrifuged at 16,000 x g for 15 min at 4°C. The resulting MEF pellet was resuspended with Tris-HCl extraction buffer as for extraction of total protein. The protein amount was measured by use of DC Protein Assay reagents (Bio-Rad) with bovine serum albumin used as a standard. Immunoblot analysis was performed as described (Charng et al., 2006; Chi et al., 2009).

ACKNOWLEDGEMENTS

We thank the Scientific Instrument Center of Academia Sinica for use of the LSM 510 Meta NLO Duo Scan inverted confocal microscope and the assistance of Ms. Shu-Chen Shen; the DNA Sequencing Core Facility of Institute of Biomedical Sciences of Academia Sinica for providing DNA sequencing services; Dr. Duan-nan Wen for protein identification by LC-ESI-MS/MS; the Plant Tech Core Facility of Agricultural Biotechnology Research Center of Academia Sinica for performing the transient expression technique; Dr. Daryi Wang for instructions and assistance in the calculations pertaining to estimation of gene divergence time; Ms. Hsiu-ting Liao for assistance in experiments; Drs. Tzyy-jen Chiou and Kuo-chen Yeh for assisting in molecular cloning with pMDC83 and pASK3 plus, respectively; Drs. Kuo-chen Yeh and Daryi Wang for critically reading the manuscript and comments; Dr. Hong-Yong Fu for lending a UV-B crosslinker; and Ms. Laura Smales for English editing.
REFERENCES


Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci USA 106: 5737-5742


nuclear export receptor EXPORTIN1A (XPO1A) in plant heat tolerance. New Phytol 186: 833-842


Table I  Approximate time of divergence of Arabidopsis GrpE and Hsp70 paralogs.

The duplication events that led to the duplication of indicated gene pairs were determined by use of the bioinformatic program Paralogs. Ks, the synonymous substitution rate; mya, million years ago; n/a, undeterminable time of divergence.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Duplication Event</th>
<th>Ks</th>
<th>Time of Divergence (mya)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mge1</td>
<td>Mge2</td>
<td>Recent</td>
<td>0.7</td>
</tr>
<tr>
<td>mTHsc70-1</td>
<td>mTHsc70-2</td>
<td>Ancient</td>
<td>1.63</td>
</tr>
<tr>
<td>Cge1</td>
<td>Cge2</td>
<td>Undefined</td>
<td>n/a</td>
</tr>
<tr>
<td>cPHsc70-1</td>
<td>cPHsc70-2</td>
<td>Recent</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table II  The alternative splice (AS) form of Mge transcripts from algae, moss, and higher plant species. Mges from several higher plants have AS forms in one or both of the Mges. Note that AtMges do not have any AS form. However, AtMge2 highly resembles an intron 2-retained AS form.

<table>
<thead>
<tr>
<th>Species</th>
<th>Gene</th>
<th>Exons</th>
<th>Description of AS Form</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>CmMge1</td>
<td>6</td>
<td>AS form not detected</td>
<td>Au9.Cre08.q370450</td>
</tr>
<tr>
<td>Physcomitrella patens</td>
<td>PpMge1</td>
<td>6</td>
<td>AS form not detected</td>
<td>Pp1s194_146V2</td>
</tr>
<tr>
<td></td>
<td>PpMge2</td>
<td>6</td>
<td>AS form not detected</td>
<td>Pp1s162_160V2</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>AtMge1</td>
<td>6</td>
<td>AS form not detected, resembles intron 2 retention</td>
<td>At5g65200</td>
</tr>
<tr>
<td></td>
<td>AtMge2</td>
<td>5</td>
<td>AS form not detected, resembles intron 2 retention</td>
<td>At4g26780</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>PoMge1</td>
<td>6</td>
<td>AS form with exon 5 spliced out</td>
<td>POPTR_0011s09130</td>
</tr>
<tr>
<td></td>
<td>PoMge2</td>
<td>6</td>
<td>AS form with in-frame intron 2 retention</td>
<td>POPTR_0013s07680</td>
</tr>
<tr>
<td>Glycine max</td>
<td>GmMge1</td>
<td>6</td>
<td>AS form not detected</td>
<td>Glyma18g10120</td>
</tr>
<tr>
<td></td>
<td>GmMge2</td>
<td>6</td>
<td>AS form with in-frame intron 2 retention</td>
<td>Glyma02g46390</td>
</tr>
<tr>
<td></td>
<td>GmMge3</td>
<td>6</td>
<td>AS form not detected</td>
<td>Glyma08g3430</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>OsMge1</td>
<td>6</td>
<td>AS form not detected</td>
<td>LOC_Os09g11250</td>
</tr>
<tr>
<td></td>
<td>OsMge2</td>
<td>6</td>
<td>AS form with in-frame intron 2 retention</td>
<td>LOC_Os08g25090</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>SbMge1</td>
<td>6</td>
<td>AS form not detected</td>
<td>Sb02g019690</td>
</tr>
<tr>
<td></td>
<td>SbMge2</td>
<td>6</td>
<td>AS form with in-frame intron 2 retention</td>
<td>Sb07g017190</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Fig. 1  The location map of the mitochondrial and chloroplastic GrpE and Hsp70 duplicates in Arabidopsis genome. The map was drawn with use of the Chromosome Map Tool of TAIR (the Arabidopsis Information Resource). The black segments on chromosome IV and V indicate the duplicated regions derived from the recent WGD.

Fig. 2  Phylogenetic tree of Mges and Cges of various species. The tree was generated by use of the neighbor-joining method in MEGA5. Bootstrap values (as a percentage of 1,000 replicates) are provided at the branches. The scale bar represents 10 amino acid replacements/100 positions. At: Arabidopsis thaliana, Cr: Chlamydomonas reinhardtii, Ec: Escherichia coli, Gm: Glycine max, Nt: Nicotiana tabacum, Os: Oryza sativa, Po: Populus sp., Pp: Physcomitrella patens, Syn: Synechocystis sp., Sc: Saccharomyces cerevisiae, Sb: Sorghum bicolor, Sl: Solanum lycopersicum, Vv: Vitis vinifera.

Fig. 3  Schematic structures of genomic DNA of Mge genes. Putative cis-elements are shown in the promoter regions of the genes in ovals. Exons are shown in rectangular boxes with numbers. Approximate T-DNA insertion sites in mge2-1 and mge2-2 are indicated by triangles.

Fig. 4  Subcellular localization of AtMge2. The plasmid containing Arabidopsis Mge2 fused to the 5’end of GFP was introduced into Arabidopsis protoplasts to transiently express the fusion protein Mge2:GFP. Mitochondria of the transformed cells were stained with Mitotracker Orange. Transformation with plasmid containing only GFP (Control) was a control.
Fig. 5  RT-PCR analyses of the differential expression of Mge genes. (A) Mge1 and Mge2 were differentially regulated by HsfA1s in Arabidopsis. The RT-PCR products of Mge1 and Mge2 were derived from the wild type (W) and quadruple KO mutant of HsfA1s (Q). RNA samples were purified from 5-d-old seedlings without (Con) or with (HS) heat treatment at 37°C for 1 h. (B) RT-PCR products of Mge1 and Mge2 derived from Arabidopsis (At) and tomato (Sl) seedlings without (C) or with (H) heat treatment at 37°C for 1 h or 40°C for 2 h, respectively. (C) RT-PCR products of Mge1 and Mge2 derived from Arabidopsis (At) and tomato (Sl) seedlings without (C) or with (U) UV-B treatment. Actin expression was a loading control.

Fig. 6  Differential complementation of the E. coli grpE mutant strain with Arabidopsis Mge1 and Mge2. Heat-sensitive E. coli mutant DA16 grpE was transformed with Arabidopsis Mges encoding long- and short-form Mge proteins with 38 and 58 putative mTPs removed from the N terminus, respectively. DA16 transformed with short Mge1, long Mge1, short Mge2, and long Mge2 are labeled 1S, 1L, 2S, and 2L, respectively. DA16 transformed with empty vector (V) was a negative control. The wild-type strain DA15 (WT) with an isogenic background to DA16 was transformed with the same empty vector to confer the same antibiotic resistance as the others and included as a comparison. Cell cultures were serially diluted and spotted onto LB plates and incubated overnight at 30, 40, 43 or 45°C.

Fig. 7  Thermotolerance assays of Mge2 T-DNA KO lines. (A) The phenotypes of Arabidopsis seedlings of the wild type (WT) and various T-DNA KO mutants after 10-d recovery from the assays under 4 different HS regimes as schematically shown on the right of each row. Seedlings of the same row were grown on the same plate. (B) The survival rates of seedlings after the assay treatments shown in (A). The labels for mutants
are \textit{mge2-1} (m2-1), \textit{mge2-2} (m2-2), \textit{hsp101} (h101), \textit{hsa32} (h32). The bars present mean ± SD of 3 replicates for each treatment (n \geq 35 each). *P < 0.01 (versus wild-type plants, Student’s t-test). BT: basal thermotolerance; SAT: short-term acquired thermotolerance; LAT: long-term acquired thermotolerance; TMHT: thermotolerance to moderate high temperature.

**Fig. 8** Transcript and protein profiles of Mge2 and other HSPs during TMHT treatment. (A) Transcript levels of Mge2 and other HSPs during day (d) 0, 1, 4, 9 of TMHT treatment detected by semi-quantitative RT-PCR. \textit{Actin} expression was a loading control. (B) Total or mitochondrial-enriched fraction proteins were extracted from the TMHT seedling samples in (A). Western blot analyses of proteins. For Mge2, 65 \mu g of proteins from mitochondrial-enriched fraction (MEF) was loaded; for all other HSPs, 60 \mu g of total proteins (TP) was loaded. VDAC, voltage-dependent anion channel, was a loading control for MEF protein. The band intensity of Mge2 was normalized to that of VDAC, with d 0 assigned as 1 and indicated below the corresponding Mge2 bands. The black arrowhead indicates the protein band of Mge2 and the white arrowhead the non-specific band. The molecular weight of Mge2 was estimated by molecular weight markers (not shown). Tubulin was a loading control for total proteins. WT: wild type; m2: \textit{mge2-2}. 


<table>
<thead>
<tr>
<th>Genes</th>
<th>Duplication Event</th>
<th>Ks</th>
<th>Time of Divergence (mya)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mge1, Mge2</td>
<td>Recent</td>
<td>0.7</td>
<td>57</td>
</tr>
<tr>
<td>mtHsc70-1, mtHsc70-2</td>
<td>Ancient</td>
<td>1.63</td>
<td>0.2</td>
</tr>
<tr>
<td>Cge1, Cge2</td>
<td>Undefined</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>cpHsc70-1, cpHsc70-2</td>
<td>Recent</td>
<td>0.71</td>
<td>59</td>
</tr>
<tr>
<td>Species</td>
<td>Gene</td>
<td>Exons</td>
<td>Description of AS Form</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>CrMge1</td>
<td>6</td>
<td>AS form not detected</td>
</tr>
<tr>
<td></td>
<td>PpMge1</td>
<td>6</td>
<td>AS form not detected</td>
</tr>
<tr>
<td></td>
<td>PpMge2</td>
<td>6</td>
<td>AS form not detected</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>AtMge1</td>
<td>6</td>
<td>AS form not detected; resembles intron 2 retention</td>
</tr>
<tr>
<td></td>
<td>AtMge2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td><em>Populus trichocarpa</em></td>
<td>PoMge1</td>
<td>6</td>
<td>AS form with exon 5 spliced out</td>
</tr>
<tr>
<td></td>
<td>PoMge2</td>
<td>6</td>
<td>AS form with in-framed intron 2 retention</td>
</tr>
<tr>
<td><em>Glycine max</em></td>
<td>GmMge1</td>
<td>6</td>
<td>AS form not detected</td>
</tr>
<tr>
<td></td>
<td>GmMge2</td>
<td>6</td>
<td>AS form not detected</td>
</tr>
<tr>
<td></td>
<td>GmMge3</td>
<td>6</td>
<td>AS form not detected</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>OsMge1</td>
<td>6</td>
<td>AS form not detected</td>
</tr>
<tr>
<td></td>
<td>OsMge2</td>
<td></td>
<td>AS form not detected</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em></td>
<td>SbMge1</td>
<td>6</td>
<td>AS form with in-framed intron 2 retention</td>
</tr>
<tr>
<td></td>
<td>SbMge2</td>
<td>6</td>
<td>AS form with in-framed intron 2 retention</td>
</tr>
</tbody>
</table>