GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Adaptive evolution of duplicated *hsp17* genes in wild barley from microclimatically divergent sites of Israel

T. Zhang[1]*, G.R. Li[1]*, Z.J. Yang[1] and E. Nevo[2]

[1]School of Life Science and Technology,
University of Electronic Science and Technology of China, Chengdu, Sichuan, China
[2]Institute of Evolution, University of Haifa, Mount Carmel, Haifa, Israel

*These authors contributed equally to this study.
Corresponding author: Z.J. Yang
E-mail: yangzujun@uestc.edu.cn

Genet. Mol. Res. 13 (1): 1220-1232 (2014)
Received December 19, 2012
Accepted July 23, 2013
Published February 27, 2014
DOI http://dx.doi.org/10.4238/2014.February.27.7

ABSTRACT. Gene duplication is a major evolutionary driving force for establishing new gene functions. However, very little is known about the involvement of the structural divergence of recent duplicated genes in local adaptation. We evaluated the nucleotide variation of recent duplicated pair copies of the *hsp17* locus for small heat shock proteins, namely, the older copy *hsp17a* and the younger copy *hsp17b*. Forty wild barley (*Hordeum spontaneum*) genotypes from 7 climatically divergent sites of "Evolution Canyon" I, Mount Carmel, Israel, were investigated for the sequence diversity of their *hsp17* copies. We identified single nucleotide polymorphisms (SNPs) of the *hsp17* locus and found 150 SNPs of *hsp17a* and 135 SNPs of *hsp17b* in the coding region. Total amino acid substitutions were present at 43 positions of *hsp17a* and 35 positions of *hsp17b* among the 7 populations. The genetic divergence and haplotype diversity were significantly different among the populations. The diversity was also correlated with microclimatic interslope

divergence factors in the canyon. The natural microclimatic selection appears to adaptively diversify the duplicated copies of *hsp17* of wild barley between the opposite warmer and cooler slopes, suggesting that the recent duplication of stress-responsive genes were subjected to adaptive evolution. The gene duplication and the resulting functional divergence enable plants to have better fitness in local climate change.

Key words: Evolution Canyon; *hsp17*; Molecular adaptive divergence; Wild barley

INTRODUCTION

Gene duplications are one of the primary forces driving the evolution of eukaryotic genomes. The duplicate genes are a major mechanism for the establishment of new gene functions (Ohnu, 1970; Taylor and Raes, 2004). The generation of new evolutionary novelties will lead to environmental adaptations. Most studies on the evolution of gene duplications focus on examining whole-genome duplications of model organisms (Zou et al., 2009) and the macroevolutionary patterns of gene diversification (Hanada et al., 2008). In contrast, we know very little about the relative importance of the two evolutionary forces, neutral genetic drift and positive selection, which shape the ultimate functional fate of the duplicate gene pair (Moore and Purugganan, 2005). Importantly, the involvement of structural and functional divergence of recent duplicated genes in relation to the local adaptation in economically important plant species has been recently recorded (Akhunov et al., 2007).

Many important cereal species appear in the evolutionary tree of the grass family (Poaceae). Barley (*Hordeum vulgare* L.) belongs to the tribe Triticeae, which includes important crop species such as wheat (*Triticum aestivum* L.) and rye (*Secale cereale* L.). As a special Triticeae plant species, barley was one of the earliest crops to be domesticated in the Near East because it is well adapted to semi-arid conditions (Poukheirandish and Komatsuda, 2007). *Hordeum spontaneum* is the wild progenitor of barley and was therefore an important resource to study the evolution, adaptation, and domestication of cultivated barley (Nevo, 1992).

Heat shock (HS) genes play a major role in helping plants cope with stress. The most representative class of proteins with molecular weights of 16-42 kDa is low molecular weight or small HS proteins (sHSPs). Based on their location in different cell compartments, plant sHSPs are targeted and classified into cytosolic and nuclear classes (Waters, 1995; Sun et al., 2002). The inducible expression of plant cytosolic sHSPs is identified generally in response to developmental or environmental signals such as osmotic stress, heavy metal ions, and oxidative stress (Waters et al., 2008).

Most HSP gene families including sHSPs are highly conserved, and all sHSPs exhibit the α-crystallin domain (ACD) of about 100 residues (Waters et al., 2008) and shared a compact β-sheet sandwich structure based on the analysis of the crystal structure of 2 sHSPs, including HSP16.9 from *T. aestivum* (van Montfort et al., 2001) and HSP16.5 from *Methanococcus jannaschii* (Kim et al., 1998). Various observations have suggested that the ACD is involved in subunit interactions of sHSPs and in binding partially unfolded proteins (Ganea, 2001). The amino acid residues in the ACD played an important role in maintaining both the structural stability and chaperone-like activity. The sHSPs usually form large oligomeric complexes and provide a means for rapidly exposing subunits, which offer hydrophobic surfaces

that protect them from aggregation (Sun et al., 2002).

Environmental stress leads to the expression of small HS genes, which can function as molecular chaperones and are involved with the folding, trafficking, and protection of cellular proteins. Therefore, genetic variation in these genes may be important for the adaptation to changes in the environment. With the help of the complete plant genome project, *Arabidopsis* and rice genomes were analyzed, revealing 19 and 24 sHSP genes, respectively (Waters et al., 2008). The comparative sequence analysis of the sHSPs indicated the extensive gene duplication and conversion in the grass genome (Waters et al., 2008). However, the evolutionary force of the sHSP needs to be further investigated both across the species and within the stressed population on the microevolution level.

The "Evolution Canyon" (EC) model has ecologically divergent environments on a microscale level, which reinforces studies that explore the interaction between the organisms and their environment (Nevo, 2001, 2009). The EC model highlights speciation and adaptation at a microscale level ("Evolution Canyon" at http://evolution.haifa.ac.il). Wild barley, *H. spontaneum*, was selected as a major model organism at EC because it is a highly inbreed species (Nevo, 1992). Based on the physiological and molecular population genetic studies of wild barley (Nevo et al., 1997; Cronin et al., 2007), we found that *H. spontaneum* at EC displays dramatic interslope adaptive genomic divergence, slope-specific fitness components, and incipient sympatric speciation on the opposite slopes (Nevo, 2009; Yang et al., 2009).

The primary objective of this study was to use the EC model to determine the structural variation in promoter and coding regions in duplicated copies of the wild barley *hsp17* gene that may be associated with different environments. The second objective was to examine whether selective forces determine evolution and are related to the functional divergence of the *hsp17* gene in microclimate adaptation.

MATERIAL AND METHODS

EC model

Lower Nahal Oren, Mt. Carmel, Israel (32°°43'N; 34°°58'E), dubbed "Evolution Canyon" I (ECI), is a seasonally dry valley in Mount Carmel, draining from east to west into the Mediterranean Sea. The opposing "African" (AS) south-facing (SFS) and "European" (ES) north-facing (NFS) slopes are 100 m apart at the valley bottom, 400 m apart at the valley top, and 200 m apart at mid-slope (Nevo et al., 2005). The slopes share an identical Plio-Pleistocene evolutionary history, presumably 3-5 million years ago, geology (upper Cenomanian limestone), pedology (Nevo et al., 1997), and regional Mediterranean climate and vegetation (Nevo et al., 1999). The slopes diverge in topology (the SFS dips 35°° and the NFS dips 25°°) and in geographic orientation, resulting in different inputs of interslope solar radiation, temperature, humidity, and biotic differences. Microclimatic data were recorded at 2 stations on each slope: the xeric "African" slope (stations AS1 and AS2) and the mesic "European" slope (stations ES6 and ES7), which were described in detail by Pavlicek et al. (2003).

Sampling of wild barley

Populations of *H. spontaneum* were sampled in ECI at 7 collection sites (populations)

and 3 elevations that were 30 m apart (60, 90, and 120 m above sea level) on each slope: AS = SFS: 3, 2, and 1; and ES = NFS: 5, 6, and 7, and 1 site (station 4) at the valley bottom (VB) (see 7 stations). Seeds from 40 genotypes, including 4 to 9 genotypes from each population, were used for the study. The individual genotypes were collected in each station at least 1-2 m apart.

DNA extraction, amplification, cloning, and sequencing

Seeds from the 40 genotypes were germinated; 1-2 g leaves were taken from a single plant (about 4 weeks old) for DNA extraction based on the method of Yang et al. (2009). Based on the *H. vulgare Hvhsp17* sequence (X64560) that was published by Marmiroli et al. (1993), a pair of primers, P1 (5'-CTTGAATCTAGACTAGTCGCCC-3') and P2 (5'-TCGATCACCAATATTTCAGACC-3'), were designed specifically to amplify the 396-bp promoter region, full-coding region, and 100-bp 3'-untranslated region (UTR) of the *Hvhsp17* gene. The polymerase chain reaction (PCR) of *hsp17* genes for cloning and sequencing were performed as by Yang et al. (2009).

Sequence analyses

Sequence analyses were based on the complete sequence, excluding sites with gaps, that was obtained from the 40 genotypes. We used *DnaSP* version 4.2 (Rozas et al., 2003) to compute several population level measurements of DNA polymorphisms and their variances. Comparisons of haplotype polymorphisms (i.e., haplotype is the combination of several nucleotide sequences) were conducted between all populations within the 2 slopes (intraslope), including populations from the VB, as well as between slopes (interslope). Polymorphisms were included in determining the total and unique number of haplotypes for each station or slope, the probability of sampling 2 different haplotypes simultaneously (haplotype diversity, H_D), the number of variable sites (S), and the average number of nucleotide differences per site between 2 sequences (nucleotide diversity, π; Nei, 1987).

Frequently used statistical tests of neutrality based on DNA polymorphism data include the following: Tajima's D (theta), Fu and Li's D, and D*. Tajima's test (Tajima, 1989) is based on estimates of the number of segregating sites and the average number of nucleotide differences that are correlated under the neutral model. If the value of D is too large or too small, the neutral "null" hypothesis is rejected. Tajima's test is not based on coalescence, but Fu and Li's (1993) tests are. These tests compare the number of mutations in internal and external branches of the genealogical sequence with expectations under selective neutrality. The result from Fu and Li's D test required data from intraspecific polymorphisms and from an outgroup (a sequence from a related species); however, D* required only intraspecific data.

Functional divergence analyses

Molecular adaptive evolution at individual sites along the *hsp17* sequences was investigated using the codeml program implemented in PAML version 3.15 (Yang, 2007). A range of codon models for measuring selection was evaluated by the likelihood ratio test (LRT) statistic, and the different models that were implemented in the codeml program of the PAML package were applied. To investigate different selective pressures along specific lineages, a

model was tested that allows only a single ω for all branches in the tree (M0) against a two-ratio model that allows an additional ω for specific branches in the tree. Site models (M0, M1, M2, M3, M7, and M8; Yang and Nielsen, 2002) that allow for heterogeneous nonsynonymous/synonymous substitution rate ratios (ω = dN/dS) among sites were used to test for diversifying selection at individual sites. The LRT was used to evaluate whether allowing sites to have ω > 1 significantly improved the fit of the model to the data (M1-M2, M0-M3, and M7-M8, where M2, M3, and M7 can accommodate positively selected sites).

RESULTS

Nucleotide sequence variation of *hsp17*

The amplification of the PCR primers of the *H. spontaneum* genotypes from ECI gives rise to 2 distinct bands. Sequencing of the *hsp17* yielded 2 types of products. The duplicated sequences of *hsp17* include a shorter type (*hsp17a*) and a longer type (*hsp17b*). They were 954 and 1345 bp, respectively, and both of them included the 456 bp coding sequence and 397 bp 5'-UTR. The *hsp17a* contained a 3'-UTR of 99 bp, while the *hsp17b* had a 3'-UTR of 492 bp.

The obtained *hsp17a* and *hsp17b* sequences were used to search the National Center for Biotechnology Information (NCBI) GenBank database. We found that there were 4 full-length *hsp17* sequences in cultivated barley, *H. vulgare* (Sato et al., 2009), 3 duplicated sequences in rice, *O. sativa* (Waters et al., 2008), and single cDNA sequence in *Brachypodium distachyon* at high similarity (over 80%). The phylogenetic tree based on the amino acid sequences of these *hsp17* homologs was generated (Figure 1). It is estimated that the divergence of the Triticeae lineage from rice was at 50 Mya, and barley or wheat and *Brachypodium* diverged about 35-40 Mya (Bossolini et al., 2007). Considering that new duplicated loci have been evolving with an average evolutionary rate of 2.9 x 10^{-3} genes/Myr in diploid Triticeae lineages (Akhunov et al., 2007), the duplication of *hsp17a* and *hsp17b* sequences likely occurred about 2-3 Mya.

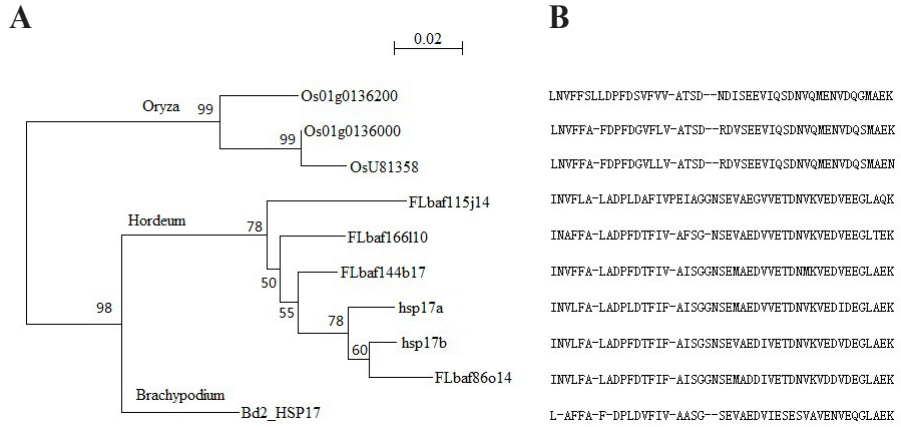

Figure 1. A. Maximum parsimony tree of the *hsp17* homology families based on nucleotide sequences of coding regions. Bootstrap values based on 1000 replicates are indicated above the branches. The lengths of tree branches are proportional to the number of mutations. **B.** Amino acid sequence alignment of the *hsp17* homology families. Only variable sites are shown and N- and C-terminals are indicated above the amino acid alignments. The scale bar is 10-nt substitutions.

The alignment for 38 *hsp17a* sequences showed a total number of 989 sites; 35 of them included alignment gaps. There were 954 sites, excluding sites with gaps. Of the 954 sites, 816 (85.5%) were monomorphic, whereas 138 sites (14.5%) were variable (polymorphic) (Table 1). Based on these variable sites, we defined 35 haplotypes with a high value of haplotype diversity (H_D = 0.991) that presented a 0.0118 nucleotide diversity (π) (Table 1). Of the *hsp17a* haplotypes, 63 of the 138 polymorphic sites were in the UTR and 75 represented insertion/deletion events in the coding region. These haplotypes were then used as the basis for frequency polymorphism and divergent calculations.

Table 1. Estimates of nucleotide diversity and test statistic for selection at full sequences of *hsp17a* and *hsp17b* of wild barley at ECI.

Total region	*Hsp17a*			*Hsp17b*		
	Total	AS	ES	Total	AS	ES
No. of sequences	38	19	16	36	14	18
No. of total sites	958	958	958	1359	1359	1359
Polymorphic sites	138	81	85	150	81	97
Singleton	117	79	83	123	79	93
Haplotypes	35	18	15	35	14	18
Haplotype diversity	0.991	0.994	0.992	0.998	1	1
Nucleotide variation						
Diversity (π/bp)	0.0118	0.0110	0.0131	0.0205	0.0162	0.0203
Polymorphism (θ/bp)	0.0369	0.0253	0.0269	0.0272	0.0191	0.0211
Nucleotide divergence						
Synonymous	0.0261	0.0250	0.0212	0.1135	0.0246	0.1005
Nonsynonymous	0.0048	0.0064	0.0079	0.0154	0.0089	0.0165
Neutrality tests						
D	-2.5281**	-2.3059*	-2.2419*	-1.0662	-0.6054	-0.0932
D*	-4.6903**	-3.2813*	-2.9975*	-2.4679	-0.9559	-1.0616
F	-4.6717**	-3.4827*	-3.2159*	-2.3233	-0.9878	-0.9029

θ = Watterson's estimate; π = Tajima's estimate; D = Tajima's D-test; D* = Fu and Li's D-test without outgroup information; F = Fu and Li's F-test. AS = South-facing slopes; ES = North-facing slopes. *0.01 < P < 0.05; **P < 0.01.

The *hsp17b* alignment sequence of 36 samples showed a total number of 1359 sites; 14 of them included alignment gaps. There were 1345 sites, excluding sites with gaps. Of these 1345 sites, 1195 (88.9%) were invariable (monomorphic), whereas 150 sites (11.2%) were variable (polymorphic) (Table 1). Based on these variable sites, we defined 35 haplotypes with a high value of haplotype diversity (H_D = 0.998), which presented a 0.0205 nucleotide diversity (π) (Table 1). Of the *hsp17b* haplotypes, 89 of 150 polymorphic sites were in the UTR and 61 represented insertion/deletion events in the coding region.

Haplotype polymorphism

The alignment of full-length *hsp17* sequences in wild barley varied considerably among the 7 populations. Most of the haplotypes (*hsp17a* H_D = 0.991, *hsp17b* H_D = 0.998) were represented by 1 genotype and were thus unique to 1 population. As for the coding region of *hsp17a*, we found that the highest values of nucleotide diversity were obtained on the "European" slope at ES7 (π = 0.03728), which had more polymorphisms than all 3 populations of the AS, the VB population, and the other 2 ES populations, while the lowest values of nucleotide diversity (π = 0.00585) were obtained from the AS2 population. For the *hsp17b* coding region, the highest values of nucleotide diversity were obtained from ES5 (π

= 0.05117), while the lowest values of nucleotide diversity (π = 0.00439) were also obtained from the AS2 population.

For *hsp17a*, 33 different haplotypes were found on the AS and ES, including 18 haplotypes from the AS and 15 haplotypes from the ES. Higher values of haplotype diversity were obtained on the AS (H_D = 0.994) than on the ES (H_D = 0.992); however, nucleotide diversity was significantly higher on the ES (π = 0.01312, or average number of nucleotide differences, k = 12.28) than on the AS (π = 0.0110, k =10.29) using the Wilcoxon two-sample nonparametric test (P < 0.001). Although the haplotypes of *hsp17b* were H_D = 1.0 between the AS and ES stations, nucleotide diversity was significantly higher at the ES (π = 0.02628, or average number of nucleotide differences, k = 27.5) than at the AS (π = 0.1614, k = 22) using the Wilcoxon two-sample nonparametric test (P < 0.0001).

We reached the following results using the 2 neutrality tests: 1) Tajima's D test: negative values of D were observed for each of the 2 slopes, indicated by large insertions/deletions (Tajima, 1989); 2) Fu and Li's test: negative values of D* were obtained for each of the 2 slopes. Tajima's D was performed for *hsp17a* (AS: D = -2.3059, 0.01 < P < 0.05; ES: D = -2.2419, 0.01 < P < 0.05); the neutrality null hypothesis is rejected. Tajima's D test was also performed for *hsp17b* (AS: D = -0.6054, P > 0.1; ES: D = -0.0932, P > 0.1); the neutrality null hypothesis is accepted.

DNA divergence and genetic diversity among populations

The average number of nucleotide substitutions per site (D_{xy}; Nei, 1987) was calculated for paired comparisons of the 7 populations (Table 2) as a measure of the extent of DNA divergence between populations. We obtained the D_{xy} values between pairs of populations from the opposite slopes (interslope) as compared them to D_{xy} values between pairs of populations from the same slope (intraslope). The effect of the inter/intraslope divergence was estimated using the Wilcoxon two-sample nonparametric test. For the coding region of *hsp17a*, the mean value of D_{xy} (0.02030) between ES populations was significantly higher than the mean value of D_{xy} (0.01030) between AS populations. The D_{xy} value of interslope pairs between AS1 and ES7 was the highest (0.02545), while the D_{xy} value of interslope pairs between AS3 and ES5 was the lowest (0.00992). A similar result was also observed in the inter/intraslope D_{xy} values in the coding region of *hsp17b*. The D_{xy} value of interslope pairs between AS1 and ES7 was the highest (0.05863), while the D_{xy} value of interslope pairs between AS3 and ES5 was the lowest (0.01640). Based on the comparison of the D_{xy} value between the *hsp17a* and *hsp17b*, we can conclude that the evolutionary rate of the coding region of *hsp17b* is significantly faster than that of *hsp17a*.

Table 2. DNA divergence (D_{xy}) based on the coding region of *hsp17a* (below diagonal) and *hsp17b* (above diagonal) among seven populations of wild barley at ECI.

	AS1	AS2	AS3	VB	ES5	ES6	ES7
AS1	-	0.0123	0.0177	0.0312	0.0462	0.0128	0.0586
AS2	0.0114	-	0.0106	0.0265	0.0391	0.0164	0.0536
AS3	0.0112	0.0084	-	0.0280	0.0413	0.0199	0.0545
VB4	0.0119	0.0101	0.0099	-	0.0428	0.0327	0.0469
ES5	0.0128	0.0106	0.0099	0.0112	-	0.0486	0.0407
ES6	0.0132	0.0110	0.0098	0.0126	0.0117	-	0.0604
ES7	0.0255	0.0228	0.0221	0.0243	0.0243	0.0241	-

AS = South-facing slopes; VB = Valley bottom; ES = North-facing slopes.

Regulatory element and insertion at the UTR region

The *hsp17* gene was first isolated from a barley genomic library (Marmiroli et al., 1993). The presence of 2 HS elements (HSEs) in the promoter was necessary for the heat stress induction of *Hvhsp17* (Gulli et al., 2005). The number of putative regulatory elements located in the 5'-upstream regions of the *hsp17* genes, including two HSEs located at positions -261 to -296 and -179 to -191 and the TATA box at -204 to -213 bp upstream of the ATG, were analyzed among the populations. The nucleotide changes of HSEs in 5'-upstream regions of the 2 slopes have no clear differences (*hsp17a*, 4 in AS and 4 in ES; *hsp17b*, 5 in AS and 5 in ES), indicating its random distribution.

For the 341 bp insertion in the *hsp17b* alignment at position 861-1254 in the 3'-UTR, we compared the interslope distribution pattern between the AS and ES stations. The total number of nucleotide changes in the populations from the 2 slopes (36 in AS *vs* 16 in the ES) is significant ($P < 0.001$ by Wilcoxon two-sample nonparametric test). This fact suggests that the variation in the *hsp17b* 3'-UTR insertion in the AS was significantly higher than variations in the ES, indicating a nonrandom distribution.

Pearson correlation with climatic variables

Analysis of Pearson correlation coefficients between climatic parameters with haplotype diversity (H_D) and nucleotide diversity (π) are presented in Table 3. Remarkable correlation values were obtained between Hudd (daily relative humidity difference, r = 0.9462), Tdd (daily temperature difference, r = 0.8187), Tm (average ambient temperature, r = 0.7857) (P = 0.395) with H_D of *hsp17a*; a negative correlation (r = -0.9701) was observed for *hsp17a Hd* and Pc (general plant cover); and correlations were not significant between climatic parameters of *hsp17b*.

Table 3. Pearson correlation coefficient (rp) between haplotype polymorphism indices at the *hsp17a* and *hsp17b* alignment sequence of wild barley, *Hordeum spontaneum*, and climatic variables at four stations in "Evolution Canyon" I, Israel.

	Hsp17a		*Hsp17b*	
	π	H_D	π	H_D
Hu	0.1011	-0.5551	-0.0450	-0.3361
Hudd	0.0162	0.9463	0.1920	-0.3419
Pc	-0.1311	-0.9709	-0.3042	0.2792
Tdd	-0.2754	0.8187	-0.1019	-0.4248
Tm	-0.3346	0.7857	-0.1646	-0.4683

Hu = average relative humidity (%), Hudd = daily relative humidity difference (%), Tdd = daily temperature difference (°C), and Tm = average ambient temperature (°C) are means for entire days (Pavlicek et al., 2003); Pc = general plant cover (%) is the mean for 2 years (Nevo et al., 1999). Note that perennial species are divided into trees, shrubs, and other life forms in different plant layers, hence, the coverage can exceed 100% (from Pavlicek et al., 2003).

Amino acid sequence variation and positive selection

The amino acid sequences of *hsp17a* and *hsp17b* have 43 and 35 substitutions, respectively (Figure 2). Based on amino acid sequence alignments, the prediction of the secondary structure of *hsp17* was generated, and all sequences included the conserved structural

features of α1-3 and β2-10 domains. Analysis of the alignment found that amino acid variations are absent in the α1 and β10 regions. More important finds were the unique variations that were present in AS- and ES-specific amino acid substitutions. As illustrated in Figure 2, the amino acid changes in *hsp17a* were found in the α2 domain (L to F) only on the ES, while the 6 amino acid changes, including changes in domains β5 and β8, occurred only on the ES at *hsp17b*.

When the dN/dS ratios were calculated, that of *hsp17a* (ω = 0.27942) was higher than that of *hsp17b* (ω = 0.13326), indicating that unequal rates of evolution were also detected for *hsp17a* and *hsp17b* at the amino acid level.

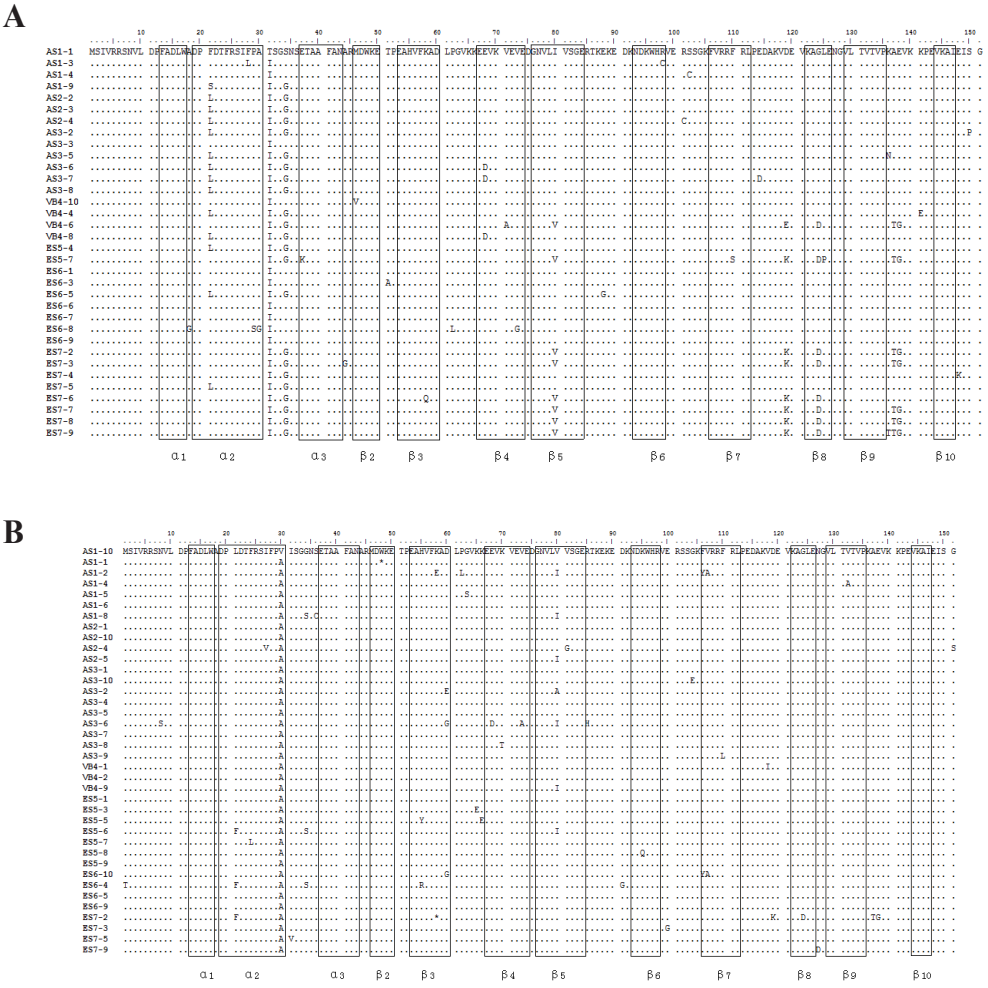

Figure 2. Alignment of amino acid sequences translated from hsp17a (**A**) and hsp17b (**B**) coding regions by Clustal X. The locations of predicted α and β strand domains are indicated by squares.

The LRT for positive selection compares the fit of the 2 nested models to the sequence data: a null model without adaptive evolution and an alternative model with adaptive evolution. Both models may invoke variation in ω among codons, but the null model is restricted to $\omega \leq 1$, whereas the alternative model allows adaptive evolution with $\omega \geq 1$. If the alternative model provides a significantly better fit to the data, then adaptive evolution is inferred. The two-ratio model that allows for a different ω for *hsp17a* compared with *hsp17b* was not a significantly better fit than the one-ratio model (M0). The results of the site models indicated that the selective pressure on the protein greatly varied among amino acid sites. For *hsp17b*, all models allowing for positively selected sites (M2, M3, and M8) provided a significantly better fit to the data than their neutral counterparts (M1, M0, and M7, respectively). Three amino acid sites (21L, 119D, and 136K) were consistently detected by the M2, M3, and M8 models as being positively selected sites. Site 21L was detected as being a positively selected site by M2 and M3 (significant by NEB). As for *hsp17a*, the two amino acid sites 30G and 60D were positively selected by the M2, M3, and M8 models.

DISCUSSION

Evolutionary forces for *hsp17* duplication copies

Under the classic model of duplicate gene evolution, one of the duplicated genes is free to accumulate mutations, which results in either the inactivation of transcription and/or a function (pseudogenization or nonfunctionalization) or the gain of a new function (neofunctionalization) as long as another copy retains the requisite physiological functions (Lynch and Conery, 2000). An evolutionary consequence for duplicated loci was referred to as subfunctionalization (Moore and Purugganan, 2005). Recently, the models involving epigenetic silencing of duplicate genes or purifying selection for gene balance have also been proposed (Zou et al. 2009).

There are several different sHSP subfamilies in higher plants, although the exact timing of the origin of the plant sHSP subfamilies is not yet known (Waters, 1995; Waters and Vierling, 1999). Based on the reported phylogenetic tree of sHSPs in several cereal plant species (Waters et al., 2008), it was found that *hsp17* gene homologs, such as *Oshsp16.9*, displayed extensive duplication on the chromosomes. All members of *Oshsp16.9* were tandemly located on rice chromosome 1. Thus, it is assumed that the present *hsp17* of wild barley was very likely orthologous to the locus on chromosome 3H because the distal end of the barley chromosomes' homoeologous group 3 is collinear with rice chromosome 1 (Bossolini et al., 2007). A maximum parsimony tree of *hsp17a* and *hsp17b* based on nucleotide sequences indicated that they were recently duplicated genes.

In this study, the duplicated copies of *hsp17a* and *hsp17b* were clearly divergent because of the recent insertion at the 3'-UTR region. We identified 150 single nucleotide polymorphisms (SNPs) of *hsp17a* and 135 SNPs of *hsp17b* in the coding region of the *hsp17* locus. The Tajima D test revealed that the SNP diversity of hsp17a rejects selective neutrality, and the genetic divergence and haplotype diversity were significantly different among AS and ES populations. The nucleotide diversity among the populations of *hsp17b* suggested the divergent evolutionary rate of different selected *hsp17* copies. The results demonstrated that an ancient copy of *hsp17a* has accumulated more genetic variations than the more recent copy of *hsp17b* in the history of duplication of the sHSP gene *hsp17*. The level of variation of the *hsp17* coding

region also exhibited similar results. It was suggested that the selective constraints are usually higher in the coding sequences than UTRs (Frydenberg et al., 2003), and the duplicates of *hsp17* copies might not have experienced similar selective forces. Significant correlation values were obtained between daily relative humidity differences, daily temperature difference, and average ambient temperature with the haplotype diversity of *hsp17a*, while there was a negative correlation with general plant cover (Table 3). The result provides evidence of the ecological and environmental properties that apparently drive the evolution of stress-inducible sHSP genes both in animal and plant kingdoms.

Functional divergence of the duplicated *hsp17* genes for stress adaptation

The structure/function studies of α-crystallin have shown that the hydrophobic N-terminus is important for the binding of unfolded proteins with the ACD, while the polar C-terminal extension is likely to be important to keep complexes of chaperones and bound proteins in solution (de Jong et al., 1998). In this study, we found that none of the amino acid substitutions were located in the domains of α1 in the N-terminus and β10 in the C-terminus. Because the hydrophobicity of the conserved domains was responsible for the solubility of the protein, it is also important for chaperone function. Moreover, the *hsp17a* gene sequences showed a considerably higher evolutionary rate of amino acid substitutions than that of the *hsp17b* sequences. The 5 positively selected amino acid sites may have resulted in higher hydrophobicity in the region of the ACD of *hsp17*. The study indicated that the shift in hydrophobicity in the ACD region might have played an important role in molecular chaperone activity (Wu et al., 2007).

Thus, the alternative models of functional diversification following gene duplication might explain the divergence of *hsp17b* from *hsp17a*. First, the function of one copy might be selectively maintained, while the other copy evolves without selective constraints, obtaining a new function (Ohnu, 1970; Zou et al., 2009). This model is supported by evidence that nucleotide substitution rates in *hsp17a* differ significantly from those in *hsp17b*. Second, one copy might evolve a new function via positive selection. This model would be supported by evidence that the dramatic divergence in amino acid sequences of the ACD indicates functional divergence. It is likely that the recent copy of the barley sHSP probably evolved as a more efficient chaperone molecule by the positive selection of the amino acid residue located in the ACD, which needs to be investigated further.

Model of the ECI is valuable to study the local gene adaptation

Plants are often exposed to drastically changing temperatures, daily fluctuations of low and high light intensities, and many other biotic and abiotic stresses (Heckathorn et al., 2004). A single phyletic line may diverge into a series of lineages that can adapt to rather different niches. Adaptive evolution can be a rapid process through which lineages derived from a recent common ancestor occur simultaneously. During this process, gene duplication and the resulting functional divergence enable lineages to have better fitness in rather different environments (Hanada et al., 2008; Zou et al., 2009).

The EC model has proven to be very fruitful in studying the biodiversity of genes, genomes, individuals, species, ecosystems, and biota at a microscale level that shares rocks

and soil, but diverging sharply and locally macroclimatically on the AS and ES slopes (Nevo, 1992, 2001, 2009). Comparing the recent studies of the stress-induced genes *Dhn1* (Yang et al., 2009) and *Dhn6* (Yang et al., 2011), *hsp17* has higher divergences, while the disease resistance-related gene *Isa* had fewer polymorphisms (Cronin et al., 2007). The ECI model provides clues to future evolutionary studies on both stress genes and speciation genes. It is interesting to further illustrate the evolutionary forces of gene families that are affected by environmental stress in relation to evolution, adaptation, speciation, and application.

ACKNOWLEDGMENTS

Z.J. Yang and G.R. Li were supported by grants #31171542 and #31101143 from the National Natural Science Foundation of China (NSFC) and Fundamental Research Funds for the Central Universities of China (#ZYGX2010J099). We thank Mrs. Robin Barasch-Permut for editing the manuscript.

REFERENCES

Akhunov ED, Akhunova AR and Dvorak J (2007). Mechanisms and rates of birth and death of dispersed duplicated genes during the evolution of a multigene family in diploid and tetraploid wheats. *Mol. Biol. Evol.* 24: 539-550.

Bossolini E, Wicker T, Knobel PA and Keller B (2007). Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. *Plant J.* 49: 704-717.

Cronin JK, Bundock PC, Henry RJ and Nevo E (2007). Adaptive climatic molecular evolution in wild barley at the Isa defense locus. *Proc. Natl. Acad. Sci. U. S. A.* 104: 2773-2778.

de Jong WW, Caspers GJ and Leunissen JA (1998). Genealogy of the alpha-crystallin - small heat-shock protein superfamily. *Int. J. Biol. Macromol.* 22: 151-162.

Frydenberg J, Hoffmann AA and Loeschcke V (2003). DNA sequence variation and latitudinal associations in *hsp23*, *hsp26* and *hsp27* from natural populations of *Drosophila melanogaster*. *Mol. Ecol.* 12: 2025-2032.

Fu YX and Li WH (1993). Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.

Ganea E (2001). Chaperone-like activity of alpha-crystallin and other small heat shock proteins. *Curr. Protein Pept. Sci.* 2: 205-225.

Gulli M, Rampino P, Lupotto E and Marmiroli N (2005). The effect of heat stress and cadmium ions on the expression of a small hsp gene in barley and maize. *J. Cereal Sci.* 42: 25-31.

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, et al. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148: 993-1003.

Heckathorn SA, Mueller JK, Laguidice S, Zhu B, et al. (2004). Chloroplast small heat-shock proteins protect photosynthesis during heavy metal stress. *Am. J. Bot.* 91: 1312-1318.

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

Lynch M and Conery JS (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1155.

Marmiroli N, Pavesi A, Di Cola G, Hartings H, et al. (1993). Identification, characterization, and analysis of cDNA and genomic sequences encoding two different small heat shock proteins in *Hordeum vulgare*. *Genome* 36: 1111-1118.

Moore RC and Purugganan MD (2005). The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8: 122-128.

Nei M (1987). Molecular Evolutionary Genetics. Columbia University Press, New York.

Nevo E (1992). Origin, Evolution, Population Genetics and Resources for Breeding of Wild Barley, *Hordeum spontaneum*, in the Fertile Crescent. CAB International, Wallingford, 19-43.

Nevo E (2001). Evolution of genome-phenome diversity under environmental stress. *Proc. Natl. Acad. Sci. U. S. A.* 98: 6233-6240.

Nevo E (2009). Evolution in action across life at "Evolution Canyons", Israel. *Trends Evol. Biol.* 1: e-3.

Nevo E, Apelbaum-Elkaher I, Garty J and Beiles A (1997). Natural selection causes microscale allozyme diversity in wild barley and a lichen at 'Evolution Canyon' Mt Carmel, Israel. *Heredity* 78: 373-382.

Nevo E, Fragman O, Dafni A and Beiles A (1999). Biodiversity and interslope divergence of vascular plants caused by microclimatic differences at 'Evolution Canyon', Lower Nahal Oren, Mount Carmel, Israel. *Isr. J. Plant Sci.* 47: 49-59.

Nevo E, Beharav A, Meyer RC and Hackett CA (2005). Genomic microsatellite adaptive divergence of wild barley by microclimatic stress in 'Evolution Canyon', Israel. *Biol. J. Linn. Soc.* 84: 205-224.

Ohnu S (1970). Evolution by Gene Duplication. Springer Verlag, Berlin.

Pavlicek T, Sharon D, Kravchenko V and Saaroni H (2003). Microclimatic interslope differences underlying biodiversity contrasts in 'Evolution Canyon', Mt Carmel, Israel. *Isr. J. Earth Sci.* 52: 1-9.

Poukheirandish M and Komatsuda T (2007). The importance of barley genetics and domestication in a global perspective. *Ann. Bot.* 100: 999-1008.

Rozas J, Sánchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.

Sato K, Shin I, Seki M, Shinozaki K, et al. (2009). Development of 5006 full-length CDNAs in barley: a tool for accessing cereal genomics resources. *DNA Res.* 16: 81-89.

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

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Taylor JS and Raes J (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38: 615-643.

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

Waters ER (1995). The molecular evolution of the small heat-shock proteins in plants. *Genetics* 141: 785-795.

Waters ER and Vierling E (1999). Chloroplast small heat shock proteins: evidence for atypical evolution of an organelle-localized protein. *Proc. Natl. Acad. Sci. U. S. A.* 96: 14394-14399.

Waters ER, Aevermann BD and Sanders-Reed Z (2008). Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress Chaperones* 13: 127-142.

Wu ML, Lin TP, Lin MY, Cheng YP, et al. (2007). Divergent evolution of the chloroplast small heat shock protein gene in the genera *Rhododendron* (Ericaceae) and *Machilus* (Lauraceae). *Ann. Bot.* 99: 461-475.

Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586-1591.

Yang Z and Nielsen R (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19: 908-917.

Yang Z, Zhang T, Bolshoy A, Beharav A, et al. (2009). Adaptive microclimatic structural and expressional dehydrin 1 evolution in wild barley, *Hordeum spontaneum*, at 'Evolution Canyon', Mount Carmel, Israel. *Mol. Ecol.* 18: 2063-2075.

Yang Z, Zhang T, Li G and Nevo E (2011). Adaptive microclimatic evolution of the dehydrin 6 gene in wild barley at "Evolution Canyon", Israel. *Genetica* 139: 1429-1438.

Zou C, Lehti-Shiu MD, Thomashow M and Shiu SH (2009). Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*. *PLoS Genet.* 5: e1000581.