# Tandem Duplication Events in the Expansion of the Small Heat Shock Protein Gene Family in *Solanum lycopersicum* (cv. Heinz 1706)

**Flavia J. Krsticevic,*,†,1 Débora P. Arce,†,‡ Joaquín Ezpeleta,*,§ and Elizabeth Tapia*,§**

*Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas, Consejo Nacional de Investigaciones Científicas y Técnicas, and §Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario, Rosario Argentina, †Grupo de Análisis, Desarrollos e Investigaciones Biomédicas, Facultad Regional San Nicolás, Universidad Tecnológica Nacional, San Nicolás, Argentina, and ‡Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Zavalla, Argentina

**ABSTRACT** In plants, fruit maturation and oxidative stress can induce small heat shock protein (sHSP) synthesis to maintain cellular homeostasis. Although the tomato reference genome was published in 2012, the actual number and functionality of sHSP genes remain unknown. Using a transcriptomic (RNA-seq) and evolutionary genomic approach, putative sHSP genes in the *Solanum lycopersicum* (cv. Heinz 1706) genome were investigated. A sHSP gene family of 33 members was established. Remarkably, roughly half of the members of this family can be explained by nine independent tandem duplication events that determined, evolutionarily, their functional fates. Within a mitochondrial class subfamily, only one duplicated member, Solyc08g078700, retained its ancestral chaperone function, while the others, Solyc08g078710 and Solyc08g078720, likely degenerated under neutrality and lack ancestral chaperone function. Functional conservation occurred within a cytosolic class I subfamily, whose four members, Solyc06g076570, Solyc06g076560, Solyc06g076540, and Solyc06g076520, support ~57% of the total sHSP RNAm in the red ripe fruit. Subfunctionalization occurred within a new subfamily, whose two members, Solyc04g082720 and Solyc04g082740, show heterogeneous differential expression profiles during fruit ripening. These findings, involving the birth/death of some genes or the preferential/plastic expression of some others during fruit ripening, highlight the importance of tandem duplication events in the expansion of the sHSP gene family in the tomato genome. Despite its evolutionary diversity, the sHSP gene family in the tomato genome seems to be endowed with a core set of four homeostasis genes: Solyc05g014280, Solyc03g082420, Solyc11g020330, and Solyc06g076560, which appear to provide a baseline protection during both fruit ripening and heat shock stress in different tomato tissues.

**KEYWORDS**
sHSP
ripening
tomato
transcriptome
RNA-seq
tandem duplication

Tomatoes are native to South America, and 13 species are currently known, including the ketchup-worthy commercial variety *Solanum lycopersicum*. The Solanaceae species are characterized by a high degree of phenotypic variation, ecological adaptability (from rainforests to deserts), and similar genomes and gene repertoires. Because of its commercial importance, *S. lycopersicum* (cv. Heinz 1706) is a centerpiece of the Solanaceae family. The complete genome of this species, comprising 950 Mb and ~35,000 protein-coding genes, was released in 2012 by the Tomato Genome Consortium. The small size of its diploid genome makes *S. lycopersicum* (cv. Heinz 1706) a good reference for the study of the Solanaceae species and explains the emerging use of this fruit as a model system for the study of fleshy fruit development (Aoki *et al.* 2013). In this regard, it is worth mentioning that the Solanum lineage has experienced two consecutive genome triplication events and that these events have led to the neo- or subfunctionalization of genes controlling important fruit characteristics such as color and fleshiness (The Tomato Genome Consortium 2012). Historically, low molecular weight (12–40 kDa) chaperone-like proteins, or small heat shock


doi: 10.1534/g3.116.032045
Manuscript received June 4, 2016; accepted for publication July 16, 2016; published Early Online August 26, 2016.

Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.032045/-/DC1.
[1]Corresponding author: Ocampo y Esmeralda, 2000 EZP Rosario, Argentina. E-mail: fkrsticevic@gmail.com

proteins (sHSPs), have been associated with stress tolerance factors by preventing the irreversible aggregation of misfolded proteins (Basha *et al.* 2015; Poulain *et al.* 2010). However, heat shock stress is not the only stimulus triggering sHSP gene expression and protein synthesis. Indeed, sHSP synthesis is also induced during fruit maturation (Low *et al.* 2000; Lawrence *et al.* 1997; Neta-Sharir *et al.* 2005), and certain development stages (Prasinos *et al.* 2005; Faurobert *et al.* 2007), in both Arabidopsis and Solanaceae plants, suggesting the existence of a complex chaperone-dependent regulating network associated with these processes to maintain cellular homeostasis. In fact, pregenomic data on tomato sHSPs provides experimental evidence for at least 14 sHSPs (Frank *et al.* 2009; Lee *et al.* 2012; Alba *et al.* 2005; Baniwal *et al.* 2004; Sanmiya *et al.* 2004). This number has almost doubled in the postgenomic era, with the finding of around 26 sHSP genes responsive to several stress situations on different tissues, including heat shock stress on leaves (Fragkostefanakis *et al.* 2015) and microspores (Frank *et al.* 2009). We note, however, that, like other gene families in tomato (Andolfo *et al.* 2014), current annotation of the sHSP gene family may not be fully defined. Multiple-copy sHSP genes may have gone unnoticed due to the intrinsic limitations of genome assembly software (Krsticevic *et al.* 2010), and multiple sHSP genes may have been misannotated into a family of functionally unrelated proteins known as ACD-like or HSP20-like (Bondino *et al.* 2012) based solely on the presence of a conserved alpha-crystallin domain IPR008978 (ACD or HSP20 domain). To uncover the actual size and organization of the sHSP gene family in the *S. lycopersicum* (cv. Heinz 1706) genome, transcriptomic data of putative sHSP genes is analyzed from an evolutionary perspective.

## MATERIALS AND METHODS

### Putative sHSP genes and transcriptome data in S. lycopersicum (cv. Heinz 1706)

A BlastP search against the Tomato protein database (ITAG2.4 Release, Sol Genomics Network) was first performed using the amino acid sequence of Solyc09g015020, characterized by a conserved ACD domain, as query. Aiming to capture all putative members of the sHSP gene family, every annotated protein containing the IPR008978 HSP20-like chaperone Interpro domain was also retrieved from the Sol Genomics Network database (ITAG release 2.40). In addition, every putative sequence related to sHSPs was retrieved from the Helmholtz-Muenchen tomato database using "small HSP" as search keyword. As a result, 58 putative sHSP sequences of *S. lycopersicum* (cv. Heinz 1706) were retrieved. Taking into account that nine β-sheets are expected in conserved ACD domains (Poulain *et al.* 2010), putative sHSP sequences were further characterized with the corresponding number of β-sheets. The ACD domain was identified with PROTEUS2, a web server supporting comprehensive protein structure prediction and structure-based annotation, http://wishart.biology.ualberta.ca/proteus2 (Montgomerie *et al.* 2008). Additionally, the number of β-sheets was estimated with Phyre2, a web server supporting the prediction of secondary structures, http://www.sbg.bio.ic.ac.uk/phyre2 (Kelley *et al.* 2015).

Fruit ripening in tomato starts at the mature green (MG) stage, where an extensive metabolic reorganization takes place, evolves to the mature breaking (MB) stage, and, 10 d later, reaches the mature red (MR) stage. Illumina RNA-seq read files of *S. lycopersicum* (cv. Heinz 1706) from two biological replicates of the MG, MB, and MR fruit ripening stages were downloaded from the DDBJ Sequence Read Archive (DRA) database (http://trace.ddbj.nig.ac.jp/dra/index_e.html). Concerning complementary analysis of putative sHSP genes during the growth stages of tomato fruit, RNA-seq read files from 1 and 2 cm immature green fruit were retrieved. Finally, concerning complementary tissue preferential analysis of putative sHSP genes, Illumina RNA-seq read files from leaf, root, flower, and flower bud were also retrieved (see Supplemental Material, Table S1). SRA files were converted to FASTQ files using the fastq-dump utility of the SRA toolkit. *S. lycopersicum* (cv. Heinz 1706). RNA-seq data from the set of MG, MB, and MR fruit maturation stages, the 1 and 2 cm immature stages, and the root, stem, and leaf plant tissues were then aligned back to the tomato genome assembly SL2.50 with annotation ITAG2.40 (The Tomato Genome Consortium 2012) using TOPHAT version 2.0.13 (Trapnell *et al.* 2009) with bowtie2 version 2.2.4 and default settings. Read counts for each gene were quantified using coverageBed from bedtools version 2.25.0. Fruit maturation entails a development process that may cause varying mRNA levels between group samples, which may discourage the use of standard normalization methods (Aanes *et al.* 2014). To shed some light on this issue, the quantroStat statistic was used (Hicks and Irizarry 2015). Differences between ripening groups were detected at the $\alpha = 0.01$ significance level. In this scenario, one must still decide if detected differences are likely to be biologically or technically driven, in which case standard normalization methods should be applied. Taking into account that only two biological replicates per group were available, which may render the quantroStat imprecise, and that relative logarithmic expressions (RLE) box plots across groups showed an almost in line appearance, the Trimmed Mean of M-values (TMM) normalization method in the edgeR Bioconductor package (Robinson *et al.* 2010) version 3.12.0 was applied. In any case, normalization factors close to unity were obtained, suggesting that the use TMM normalization to remove technical variability would have minimal impact on downstream data analysis. Fragments per kilobase of transcript per million mapped fragments (FPKM) values were calculated to determine transcript abundance of individual genes. Average FPKM values were calculated for each pair of biological replicates for each sample (see Table S2). Following the criteria of previous studies (González-Porta *et al.* 2013; Hebenstreit *et al.* 2011), genes were considered expressed when average FPKM values were $\geq 2$. Differential expression between pairs of ripening stages was assessed with the exactTest function. The MG fruit stage was taken as baseline and edgeR's exact test was applied to identify significant $\log_2$-fold changes (FC) of the MB and MR stages relative to the MG baseline at the $\alpha = 0.01$ significance level (see Figure 1). Positive $\log_2$ FC values indicated upregulation, negative values indicated downregulation, and zero indicated constant gene expression relative to the MG baseline. In these studies, the *tip41* housekeeping gene (Solyc01g107420) was used as negative control for differential gene expression while the *hsp70* gene (Solyc11g020040) was used as a positive control for upregulation during fruit ripening (Exposito-Rodriguez *et al.* 2008; Fei *et al.* 2004).

### Assessing the phylogeny of putative sHSP genes in S. lycopersicum (cv. Heinz 1706), and evolutionary times of duplication events

A phylogeny-based functional analysis was used to complete (Eisen *et al.* 1998) preliminary chaperone function prediction obtained through RNA-seq expression analysis of putative sHSPs in *S. lycopersicum* (cv. Heinz 1706). Eleven sHSP sequences of *Arabidopsis thaliana* (Waters *et al.* 2008), known to be representative and conserved between angiosperms, together with the 58 putative sHSP sequences of tomato, were used to construct an initial phylogenetic tree. Additionally, the subcellular localization of putative sHSPs was obtained from analogous clusters formed in the phylogenetic tree (see Figure S1). Molecular evolutionary analyses were conducted with MEGA version 6 (Tamura

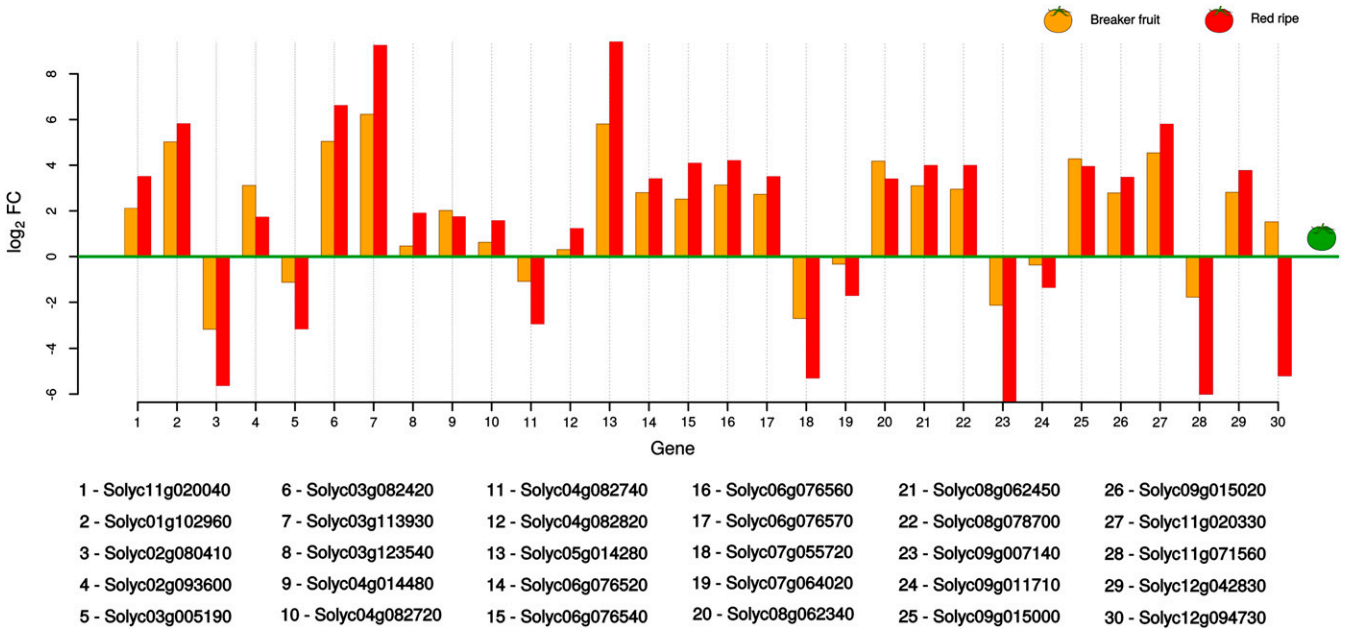


**Figure 1** Putative sHSP genes differentially expressed during tomato fruit ripening (29 out of a total of 58). Differential expression at the mature breaker and mature red fruit ripening stages relative to the reference mature green stage is quantified as $\log_2$ fold change (FC). The *hsp70* gene or Solyc11g020040 (#1) was used as a positive control of upregulation.

*et al.* 2013). Amino acid sequences were aligned using ClustalW with default settings, and a Maximum Likelihood analysis based on the Jones-Taylor-Thornton (JTT) substitution model for proteins with a gamma distribution was performed (Larkin *et al.* 2007). A condensed tree was built with a bootstrap method set to work with 100 replications and a cut-off value of 70% (Hillis and Bull 1993). Aiming to evidence evolutionary relationships in the resulting sHSP gene family, a second phylogenetic tree was built with curated sHSPs obtained after joint RNA-seq and phylogeny-based functional analysis. In this case, a Maximum Likelihood analysis based on a WAG substitution model for proteins with a gamma distribution was used.

Following Xia *et al.* (2011), the DnaSP tool (Librado and Rozas 2009) was used for the computation of occurrence time $T$ of the tandem duplicated genes. Briefly, DnaSP provides the mean number $K_S$ of synonymous substitutions per synonymous site between pairs of duplicated genes. The occurrence time of duplication events can be calculated using $T = K_S/2\lambda$, where $\lambda$ is the clockwise substitution rate, a characteristic parameter of each species. Based on the work of Yang *et al.* (2008), $\lambda = 1.5 \times 10^{-8}$ of *A. thaliana* was used to find the approximate value of $T$ in tomato.

## Data Availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

# RESULTS

## Putative small heat shock proteins (sHSPs) in the S. lycopersicum (cv. Heinz 1706) genome

To begin the reconstruction of a large-scale picture of the sHSP family organization in the *S. lycopersicum* (cv. Heinz 1706) genome, 58 putative sHSPs were mapped to their chromosomes. Putative sHSP genes were found to be distributed evenly (Pearson's Chi-squared test based on $10^7$ replicates, p-value = 0.5023) across the 12 tomato chromosomes. Despite this random distribution, an integral analysis of the expression profiles of putative sHSPs genes across multiple conditions and tissues, together with their phylogenetic relationships, may help define and characterize the sHSP gene family.

## Differential expression of putative sHSP genes during tomato fruit ripening

Differential expression analysis pointed out 29 out of the initial set of 58 putative sHSP genes as differentially expressed in the MR fruit ripening stage relative to the reference MG one (see Figure 1). A subgroup of 20 putative sHSP genes was found to be upregulated during the fruit ripening process, with expression intensity rising from MG to MR, suggesting that these sequences were indeed sHSPs. On the other hand, a subgroup of nine putative sHSP genes was found to be downregulated in the MR stage relative to the reference MG one. This pattern of downregulation may be caused by two nonmutually exclusive events: natural gene expression declination during fruit senescence, or the fact that putative downregulated sHSP genes are actually HSP20-like genes. To further support the first evidence of chaperone functionality in putative sHSP genes being upregulated in the MR stage relative to the reference MG case, and to clarify the status of those appearing as downregulated, not differentially expressed, or even not expressed, a phylogenomic analysis was performed.

## Phylogeny-based annotation of putative sHSP genes

A set of 20 clusters representing the four major monophylogenetic clades characteristic of higher plants, MCI, MCII, MCIII, and MCIV (Bondino *et al.* 2012), was identified in a phylogenetic tree built from 58 putative sHSPs sequences in *S. lycopersicum* (cv. Heinz 1706) and 11 sHSP reference sequences in *A. thaliana* (see Figure S1).

A set of 18 sHSP genes distributed in 10 clusters, containing 16 upregulated and two downregulated members during fruit ripening, was clearly identified based on functional orthology with *A. thaliana* and previous experimental evidence. Another set of 13 sHSP genes distributed in four clusters (#14, #13, #9, and #7) was identified after careful analysis of previous experimental evidence, completeness of the ACD

domain, and presence of other domains (Table S3). Cluster #14 initially had two paralogous members, Solyc08g078710 and Solyc08g078720, map together with singleton Solyc08g078700. The three sequences show around 47% sequence identity, suggesting a common origin product of a tandem duplication event (see Figure S3). Altogether, this evidence suggests Solyc08g078700 membership to Cluster #14. In addition, upregulated Solyc08g078700 is similar (46.7% identity and 63.8% similarity at the amino acid sequence level) to the mitochondrial 23.6 kDa Q96331 protein in *A. thaliana*, already classified as a sHSP (Scharf *et al.* 2001). Altogether, this evidence suggests that, although Solyc08g078710 is not expressed during fruit ripening and Solyc08g078720 is not differentially expressed, both are sHSPs. Cluster #13 contains two paralogous members, Solyc10g086680 and Solyc09g011710. Although Solyc10g086680 is not expressed during fruit ripening, and Solyc09g011710 is downregulated, they both have a complete ACD domain, and, additionally, Solyc09g011710 has been reported to be downregulated under heat shock stress in leaves (Fragkostefanakis *et al.* 2015). Altogether, this evidence indicates that both Solyc10g086680 and Solyc09g011710 are sHSPs. Cluster #9 contains four paralogous members, Solyc01g009200 and Solyc01g009220, which are not expressed during fruit ripening, but are downregulated during fruit development (see Figure S2), and Solyc11g071560 and Solyc09g007140, which are downregulated during fruit ripening. Taking into account that all members of this cluster but Solyc01g009220 have a complete ACD domain, the evidence indicates that the four members of Cluster #9 are likely sHSPs (see Table S3). Cluster #7 contains four paralogous members: Solyc04g082720, which is upregulated during fruit ripening; Solyc04g082740, which is downregulated; Solyc01g098810, which is not differentially expressed; and Solyc01g098790, which is not expressed. In addition, all members of this cluster except Solyc01g098790 have a complete ACD domain. Altogether, these sources of evidence indicate that the four members of Cluster #7 are sHSPs. Note, however, that Cluster #7 has a striking expression pattern, with both upregulated and downregulated members during tomato fruit ripening (see *Discussion*). In agreement with the joint presence of the ARID and ACD domains (Riechmann *et al.* 2000), a set of six putative sHSPs clustering together in Cluster #4 (Figure S1) turned out to be transcription factors, *i.e.*, HSP20-like genes. While two of these transcription factors have been reported previously (Bondino *et al.* 2012), four of them could be new. Similarly, another set of 14 putative sHSPs distributed in five clusters are also potential HSP20-like genes, since they all lack functional and evolutionary evidence to support sHSP family membership. Finally, two sHSP genes, Solyc02g093600 and Solyc04g072250, were identified in the set of seven putative sHSP genes that remained unclustered. Solyc02g093600 is upregulated during fruit ripening, has a complete ACD domain and previous experimental evidence (Fragkostefanakis *et al.* 2015). On the other hand, structural features of Solyc04g072250, including the presence of an ORF, absence of a premature stop codon, and a complete ACD domain, suggest its chaperone functionality. However, previous experimental evidence (Fragkostefanakis *et al.* 2015), and lack of expression during fruit ripening suggest that Solyc04g072250 is actually nonfunctional. Based on the these considerations, a sHSP gene family of around 33 members can be defined in the *S. lycopersicum* (cv. Heinz 1706) genome. Regarding the fruit ripening process, this result brings new experimental evidence about the chaperone function of four members of the sHSP gene family, Solyc04g082740, Solyc09g007140, and Solyc11g071560, which appear downregulated, and Solyc04g082720, which appears upregulated. Additionally, we extend the chaperone function of Solyc03g123540 and Solyc02g093600, already reported as sensitive to heat shock stress, to fruit ripening.

## Subcellular localization of sHSP genes

Subcellular localization of proteins can provide important evidence about their function (Emanuelsson *et al.* 2000). Based on evolutionary relationships, and in agreement with previous experimental evidence (Figure S1), 22 sHSPs are predicted to be distributed ubiquitously across cellular compartments and organelles (Table S4). Briefly, 11 of them are predicted for the cytoplasm (CI), two for chloroplasts (CP), three for the endoplasmic reticulum (ER), five for the mitochondria (MT), and one for the peroxisoma (PX). Two sHSP genes, Solyc05g014280 and Solyc03g113930, whose products are presumed for the CP and ER, show the highest levels of differential expression during fruit ripening (Table S5). Remarkably, four sHSP genes found to be highly responsive to fruit ripening stress, Solyc05g014280, Solyc03g082420, Solyc11g020330, and Solyc06g076560, whose products are presumed for the CP, ER, and CI, have been also reported to be highly responsive to heat shock stress in leaves (Fragkostefanakis *et al.* 2015) and microspores (Frank *et al.* 2009), suggesting the existence of a core set of sHSP genes important to maintain cellular homeostasis under 1, 4, 6, 8, and 9 [both stress situations (Figure 2)].

## Tandem duplication events in the sHSP gene family

Tandem and segmental duplication are the main sources of diversity for the evolution of large gene families in plants (Cannon *et al.* 2004). A phylogenetic analysis of the sHSP gene family in *S. lycopersicum* (cv. Heinz 1706) revealed 17 sHSP genes (~51%) produced by tandem duplications events (see Figure 3). These genes are organized into six subfamilies that map to chromosomes 1, 4, 6, 8, and 9. Certainly, segmental duplication has also contributed to the expansion of gene families in plants. However, its role may be less pronounced in the diversification of the sHSP family (Waters *et al.* 2008). To shed light on this issue, duplications of sHSP genes were investigated with the MCSCAN tool (Tang *et al.* 2008), and little evidence of a dominant role of segmental duplication in *S. lycopersicum* was found. Duplication analysis based on the identification of synteny blocks showed only two segmental duplications among chromosomes 6, 9, and 10 involving three genes, Solyc06g076520, Solyc09g011710, and Solyc10g086680. These segmental duplications may be attributable to the last whole-genome triplication (91–52 Myr) that occurred in the Solanum lineage (The Tomato Genome Consortium 2012).

Three sHSP subfamilies are useful to describe the alternative functional outcomes of tandem duplicated sHSP genes in *S. lycopersicum* (cv. Heinz 1706). A first subfamily involves three MT class sHSP genes mapping together to a region of ~11.4 kb in chromosome 8 (SL2.40ch08:59625875..59637274). Notably, in this subfamily, only the basal gene Solyc08g078700 appears as clearly functional, while the other two subfamily members, Solyc08g078710 and Solyc08g078720, seem to be losing their ancestral chaperone function. A second subfamily involves four functional intronless CI class sHSP genes mapping together to a ~17.9-kb region in chromosome 6 (SL2.40ch6:47.547k..47.564k). Three members of this subfamily, Solyc06g076540, Solyc06g076560, and Solyc06g076570, have been previously reported by Goyal *et al.* (2012) in *S. lycopersicum* (cv. Ohio 8245). Now, a fourth member, Solyc06g076520, is reported. Notably, the four members of this subfamily support ~57% of the sHSP transcripts in the MR fruit ripening stage (Table S2). Furthermore, subfamily members Solyc06g076540 and Solyc06g076560 are among the most differentially expressed sHSP genes during fruit ripening (see Table S5). Finally, a third subfamily involves two sHSP cytosolic/nuclear genes, Solyc04g082720 and Solyc04g082740, mapping together to a ~9.1-kb region in chromosome 4 (SL2.50ch04:66300179..66309278). Notably,

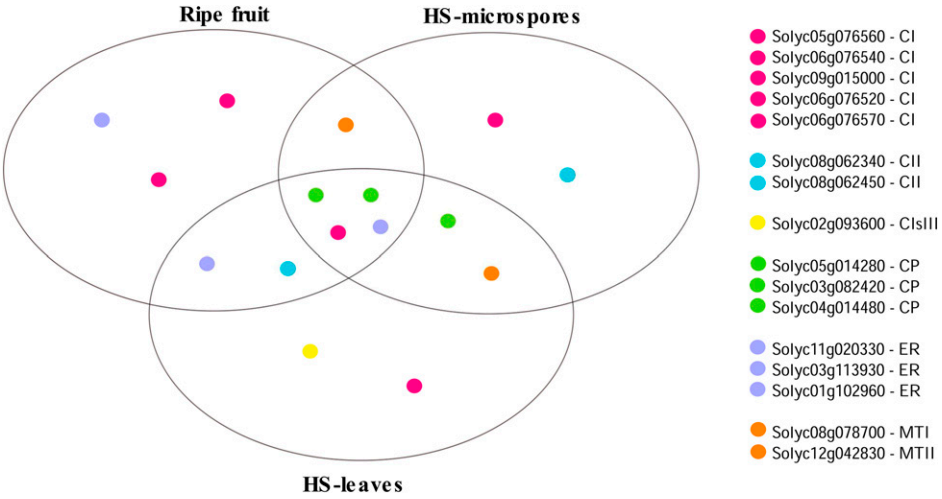


**Figure 2** Top 10 sHSP genes responsive to fruit ripening and heat-shock stress in leaves and microspores. For each condition, sHSP genes targeted to the endoplasmic reticulum (ER), the cytosolic classes I and II (CI, CIsIII and CII), perixoma (PX), chloroplast (CP) and mitochondrion (MT) are shown. Four sHSP genes, Solyc05g014280, Solyc03g082420, Solyc11g020330, and Solyc06g076560, targeted to the CP, the ER, and the CI, are responsive in all conditions.

although both members of this subfamily are functional, their temporal expression patterns over development and ripening suggest that they are undergoing a subfunctionalization process.

Identification of multiple-copy genes in tomato, like that presented here for the sHSP gene family, can contribute to reducing the uncertainty of estimations about exploitable phenotypic variation, which could be very useful in commercial tomato breeding programs.

## DISCUSSION

### Small sHSP genes in the S. lycopersicum (cv. Heinz 1706) genome

Even with the large amount of genomic data now available, the number and functionality of sHSP genes in the Solanaceae family remain largely unknown, and their functional annotation is often inconsistent across authors and databases (see Table S3). An evolutionary perspective on the transcriptome analysis of *S. lycopersicum* (cv. Heinz 1706) allowed us to define a sHSP gene family of around 33 members. Families of sHSP genes in plant species tend to be rather large and variable in size: 19 sHSP genes have been reported in *A. thaliana* (Scharf *et al.* 2001; Siddique *et al.* 2008), 39 in rice *Oryza sativa* (Ouyang *et al.* 2009) and 51 in *Glycine max* (Lopes-Caitar *et al.* 2013). Despite this variability, the proportion of sHSP genes in plant genomes is roughly constant, ranging from ~0.06 to ~0.1%. The proportion of sHSP genes in *S. lycopersicum* (cv. Heinz 1706), 0.095%, or 33 out of a total of 34,727, is in accordance with these previous studies, suggesting that the totality of members of the sHSP gene family has been uncovered in tomato. Note, however, that the actual number and location of sHSP genes in the ~7000 domesticated lines of *S. lycopersicum* collected in the EU-SOL BreeDB database (https://www.eu-sol.wur.nl) may vary according to directional selection pressures (Ercolano *et al.* 2014).

### Tandem duplication events and the expansion of the sHSP gene family in tomato

The main function of sHSPs is to maintain the homeostasis of cellular proteins. The importance of this ubiquitous function supports the presence of redundant sHSPs, so that if one of them fails, the others are ready to supply their chaperone function. Evolutionary forces have clearly affected and modeled the sHSP gene family (Ohno 1970). Roughly half (17) of the sHSP genes in the *S. lycopersicum* (cv. Heinz 1706) genome can be explained by tandem duplication events. In most of these events, redundancy tends to be eliminated, so that one of the copies retains its ancestral function while the other becomes a pseudogene (Zheng and Gerstein 2007).

Neutral evolutionary processes seem to be a valid argument to explain the behavior of two of three MT class tandem duplicated sHSP genes, Solyc08g078700, Solyc08g078710, and Solyc08g078720, mapping together to a 11.4 kb region in chromosome 8. While the basal Solyc08g078700 gene retained its ancestral chaperone function and evolved under purifying selection (see Figure S3 and associated key), its two accompanying copies, Solyc08g078710 and Solyc08g078720, degenerated. Functional redundancy also seems to a be a valid possibly under the effect of neutrality. Although Solyc08g078710 has a complete ACD domain, it is expressed neither in plant tissues (leaf, root, flower, and flower bud) nor during fruit development (1 and 2 cm), fruit ripening, or heat shock stress, probably due to variations in the promoter architecture of the 5′ UTR region. Conversely, although Solyc08g078720 is expressed in all plant tissues, it is insensitive to fruit development, fruit ripening, or heat shock stresses, probably due to the presence of an incomplete ACD domain with only seven β-sheets (see Table S3). Altogether, this evidence suggests that neither Solyc08g078710 nor Solyc08g078720 retained their ancestral chaperone function. Functional redundancy seems to be to a be a valid argument to explain the behavior of four Class I tandem duplicated intronless sHSP genes, Solyc06g076520, Solyc06g076540, Solyc06g076560, and Solyc06g076570, mapping together to a ~17.9 kb region in chromosome 6 (SL2.40ch6:47.547k..47.564k). If there is a biological reason for this sHSP gene subfamily to stay in array in a chromosome 6 region, *e.g.*, due to its important relative contribution to differential expression and transcript abundance of sHSP genes during fruit ripening, a high degree of conservation of this subfamily across close Solanum species should be expected. In effect, duplication analysis suggests that only Solyc06g076520 originated during the last whole-genome triplication in the Solanum lineage (together with Solyc09g011710 and Solyc10g086680 in Cluster #13). The remaining members of Cluster #2, Solyc06g076570, Solyc06g076540 and Solyc06g076560, seem to be the product of tandem duplication events, the first of which took place ~13 Myr ago (Figure S4). Taking this together with collinearity results between potato and tomato at the chromosome 6 region of Cluster #2, we can hypothesize that gene associations in Cluster #2 have indeed been maintained, thus reflecting their importance in the sHSP gene family. We note, however, that orthologous genes of Solyc06g076560 are absent in the *Solanum tuberosum* and *S. pennelli* genomes. Actually, Solyc06g076560 is a paralogous copy of parental Solyc06g076520

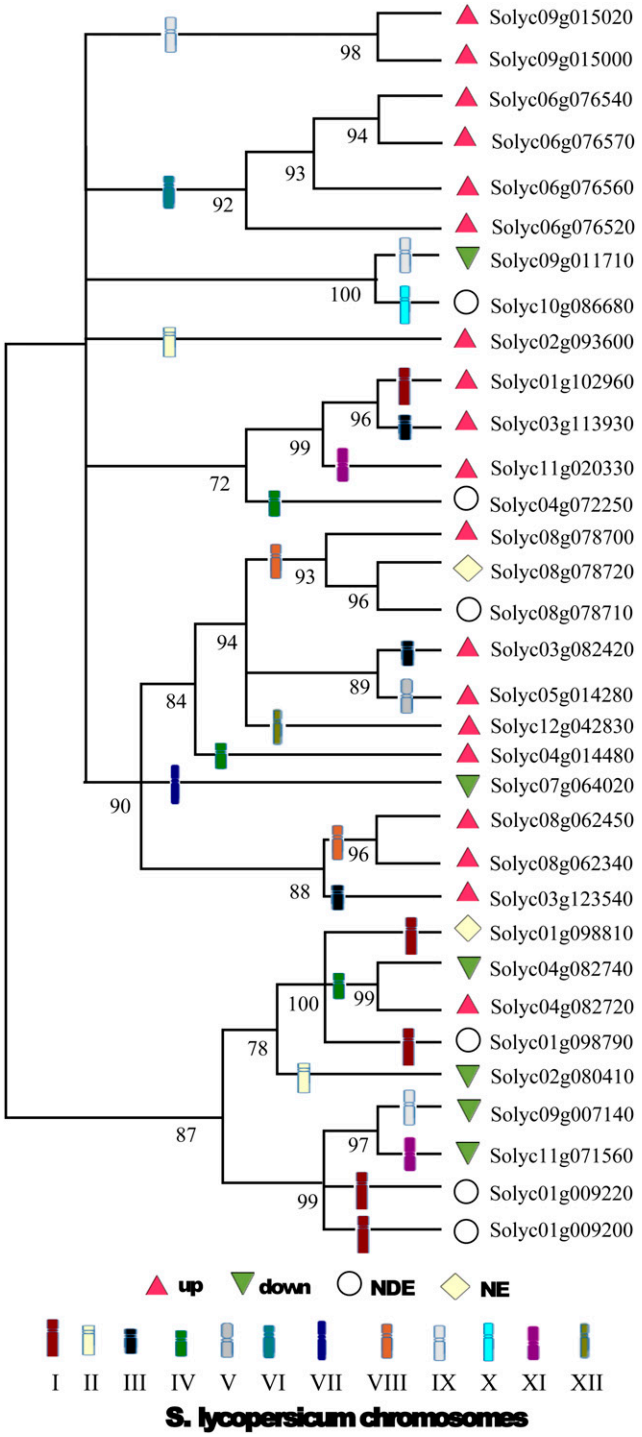


**Figure 3** Phylogenetic relationships in the sHSP family of the *S. lycopersicum* (cv. Heinz 1706) genome. Amino acid sequences deduced from the 33 members of sHSP gene family were used. Gene expression profiles during fruit ripening are shown. Differential expression is measured at the MR stage relative to the MG reference stage. Different symbols are used for depicting upregulated and downregulated, not differentially expressed (NDE), and not expressed (NE) genes. Aiming to highlight tandem duplication events that happened during the evolutionary history of this family, the chromosome localization of sHSP genes is used for branch labeling. Nine tandem duplication events are present: one in chromosome 1 involving the pair Solyc01g009200–Solyc01g009220, one in chromosome 4 involving the pair Solyc04g082720–Solyc04g082740, two in chromosome8 involving the pair Solyc08g062340–Solyc08g062350, and the trio Solyc08g078700–Solyc08g078710–Solyc08g078720, three in chromosome 6 involving the quartet Solyc06g076520–Solyc06g076540–Solyc06g076560–Solyc06g076570, and one in chromosome 9 involving the pair Solyc09g015000–Solyc09g015020.

(97.9% nucleotide identity) suggesting the occurrence of a tandem duplication event exclusive to the *S. lycopersicum* clade (Figure 3). Finally, tandem duplicated genes may diversify and undergo some degree of neo-functionalization. In this regard, either of the copies may acquire a new beneficial function, and the other retain its ancestral function, or both copies may undergo subfunctionalization, with each copy being expressed uniquely at different tissues or with a temporal expression pattern (Lynch and Force 2000; He and Zhang 2005). In effect, temporal subfunctionalization seems to be a valid argument to explain the behavior of the two tandem duplicated sHSP genes, Solyc04g082720 and Solyc04g082740, mapping together to a ~9.1 kb region in chromosome 4. These sHSP genes show a complementary temporal expression pattern (MacCarthy and Bergman 2007) during fruit development and ripening. According to the NexGenEx-Tom database (http://140.164.45.142/NexGenEx-Tom/expression/exp-search.aspx), while the peak of expression of Solyc04g082740 occurs during fruit development at 3 cm fruit size, the peak of expression of Solyc04g082720 occurs during ripening at the MR stage.

Similarly to rice, where roughly half (19) of sHSP genes in the genome have been reported to be produced from tandem duplication events (Ouyang *et al.* 2009), and differently from *A. thaliana*, where only one tandem duplicated sHSP gene has been reported (Scharf *et al.* 2001), our results suggest that tandem duplication events have contributed greatly to the expansion of sHSP gene family in *S. lycopersicum*.

## A core set of homeostasis sHSP genes in tomato

Simultaneous analysis of most differentially expressed sHSP genes during fruit ripening and heat shock stress in leaves and microspores (Fragkostefanakis *et al.* 2015; Frank *et al.* 2009) revealed the existence of four sHSP genes with a common and very sensitive response to both stress situations. Being sessile organisms, plants have evolved mechanisms to deal with and tolerate multiple stress situations. In contrast to other eukaryotic organism domains, plants are unique in expressing in a cell a multiplicity of cytosolic sHSPs and specific sHSPs targeted to the plastids, the ER, and the MT (Waters *et al.* 1996; Waters 1995; de Jong *et al.* 1998; Siddique *et al.* 2008).

Chloroplasts are responsible for photosynthesis as well as numerous other functions in a plant cell (Jarvis and López-Juez 2013). During fruit ripening, a massive transformation process of CP into chromoplasts takes place, and, thus, repair and stabilization of proteins are required (Lawrence *et al.* 1997; Zeng *et al.* 2014). A subfamily of two CP sHSP genes, Solyc03g082420 and Solyc05g014280, is present in the tomato genome. The Solyc05g014280 or *vis1* gene is highly differentially expressed during fruit ripening and seems to play a specific role in pectin depolymerization, a common event in fruit ripening, by reducing the thermal denaturation of depolymerizing enzymes in response to daytime elevated temperatures (Ramakrishna *et al.* 2003). On the other hand, the extensively characterized Solyc03g082420 or *hsp21* gene (Lawrence *et al.* 1997; Srivastava *et al.* 2010; Matas *et al.* 2011) is highly differentially expressed, not only during fruit ripening and heat shock stress in leaves and plant microspores, but also during plant development (Neta-Sharir *et al.* 2005; Lambert *et al.* 2011). The ER plays a key role in a cell endomembrane system. It is involved in folding and assembling the majority of proteins that a cell secretes, in lipid

synthesis, and in sustaining cell homeostatic balance (Walter and Ron 2011; Howell 2013). In higher plants, ER sHSPs accumulate in this compartment (Zhao *et al.* 2007), and may protect ER proteins from stress (LaFayette and Travis 1990; Sticher *et al.* 1990; Helm *et al.* 1993). A subfamily of three ER intronless genes, Solyc01g102960, Solyc11g020330, and Solyc03g113930, is present in the tomato genome. Solyc01g102960 is highly differentially expressed only during fruit ripening. On the other hand, Solyc11g020330 and Solyc03g113930 are highly differentially expressed during both fruit ripening and heat-shock stress in leaves and microspores, suggesting their role as general-purpose stress responsive sHSP genes at the ER.

Ripening involves a massive structural change comparable to that experienced by plastids when exposed to environmental stress (Giovannoni 2001). In particular, the cytosol is an important place for the activity of sHSP gene products in tomato. Like other higher plants, the subfamily of cytosolic sHSP genes in tomato is larger than those of cell organelles (Reddy *et al.* 2014). Eight sHSP genes, conforming cytosolic subfamilies CI and CII, are highly differentially expressed during fruit ripening (Table S5). In agreement with the ancient membership of cytosolic genes in the sHSP gene family (Waters and Vierling 1999), these genes account for ~77% of the total transcripts of sHSP genes in the MR ripening stage (Table S2). Remarkably, intronless Solyc06g076560, the youngest sHSP gene exclusive to the *S. lycopersicum* clade, accounting for ~17% of the total transcripts of sHSP genes in the MR ripening stage, is also highly differentially expressed during heat shock stress in leaves and microspores. In summary, Solyc03g082420, Solyc05g014280, Solyc11g020330, and Solyc06g076560, targeted to the CP, the ER, and cytosolic compartments, define a core set of sHSP genes contributing to cell homeostasis in both fruit ripening and heat shock stress, suggesting that other subsets of multi-purpose sHSP genes may coexist within the family.

The results presented here for the sHSP gene family suggest that systematic approaches built upon an evolutionary analysis of transcriptome data may be also effective to disentangle the organization and functionality of other complex gene families.

## ACKNOWLEDGMENTS

This work was funded by project PICT 2012-2513, "Multiplex systems for targeted microfluidic amplification and next-generation sequencing (NGS) sequencing," National Agency for Science and Technology Promotion, Argentina (institution: National Scientific and Technical Research Council [CONICET]) and PICT 2014-3181, "Process optimization of genetic breeding in tomato using bioinformatics tools," National Agency for Science and Technology Promotion, Argentina (institution: CONICET).

## LITERATURE CITED

Aanes, H., C. Winata, L. F. Moen, O. Østrup, S. Mathavan *et al.*, 2014 Normalization of RNA-sequencing data from samples with varying mRNA levels. PLoS One 9: e89158.

Alba, R., P. Payton, Z. Fei, R. McQuinn, P. Debbie *et al.*, 2005 Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. Plant Cell 17: 2954–2965.

Andolfo, G., F. Jupe, K. Witek, G. J. Etherington, M. R. Ercolano *et al.*, 2014 Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. BMC Plant Biol. 14: 120.

Aoki, K., Y. Ogata, K. Igarashi, K. Yano, H. Nagasaki *et al.*, 2013 Functional genomics of tomato in a post-genome-sequencing phase. Breed. Sci. 63: 14–20.

Baniwal, S. K., K. Bharti, K. Y. Chan, M. Fauth, A. Ganguli *et al.*, 2004 Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. J. Biosci. 29: 471–487.

Basha, E., H. O'Neill, and E. Vierling, 2015 Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. Trends Biochem. Sci. 37: 106–117.

Bondino, H., E. Valle, and A. ten Have, 2012 Evolution and functional diversification of the small heat shock protein alpha-crystallin family in higher plants. Planta 235: 1299–1313.

Cannon, S. B., A. Mitra, A. Baumgarten, N. D. Young, and G. May, 2004 The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol. 4: 1–21.

de Jong, W. W., G. J. Caspers, and J. A. Leunissen, 1998 Genealogy of the alpha-crystallin–small heat-shock protein superfamily. Int. J. Biol. Macromol. 22: 151–162.

Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein, 1998 Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95: 14863–14868.

Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne, 2000 Predicting subcellular localization of proteins based on their n-terminal amino acid sequence. J. Mol. Biol. 300: 1005–1016.

Ercolano, M. R., A. Sacco, F. Ferriello, R. D'Alessandro, P. Tononi *et al.*, 2014 Patchwork sequencing of tomato San Marzano and Vesuviano varieties highlights genome-wide variations. BMC Genomics 15: 138.

Exposito-Rodriguez, M., A. Borges, A. Borges-Perez, and J. Perez, 2008 Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biol. 8: 131.

Faurobert, M., C. Mihr, N. Bertin, T. Pawlowski, L. Negroni *et al.*, 2007 Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiol. 143: 1327–1346.

Fei, Z., X. Tang, R. M. Alba, J. A. White, C. M. Ronning, G. B. Martin, S. D. Tanksley, and J. J. Giovannoni, 2004 Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. Plant J. 40: 47–59.

Fragkostefanakis, S., S. Simm, P. Paul, D. Bublak, K.-D. Scharf *et al.*, 2015 Chaperone network composition in *Solanum lycopersicum* explored by transcriptome profiling and microarray meta-analysis. Plant Cell Environ. 38: 693–709.

Frank, G., E. Pressman, R. Ophir, L. Althan, R. Shaked *et al.*, 2009 Transcriptional profiling of maturing tomato (*Solanum lycopersicum* L.) microspores reveals the involvement of heat shock proteins, ROS scavengers, hormones, and sugars in the heat stress response. J. Exp. Bot. 60: 3891–3908.

Giovannoni, J., 2001 Molecular biology of fruit maturation and ripening. Annu. Rev. Plant Physiol. Plant Mol. Biol. 52: 725–749.

González-Porta, M., A. Frankish, J. Rung, J. Harrow, and A. Brazma, 2013 Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol. 14: R70.

Goyal, R. K., V. Kumar, V. Shukla, R. Mattoo, Y. Liu *et al.*, 2012 Features of a unique intronless cluster of class I small heat shock protein genes in tandem with box C/D snoRNA genes on chromosome 6 in tomato (*Solanum lycopersicum*). Planta 235: 453–471.

He, X., and J. Zhang, 2005 Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169: 1157–1164.

Hebenstreit, D., M. Fang, M. Gu, V. Charoensawan, A. van Oudenaarden *et al.*, 2011 RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol. Syst. Biol. 7: 497.

Helm, K. W., P. R. LaFayette, R. T. Nagao, J. L. Key, and E. Vierling, 1993 Localization of small heat shock proteins to the higher plant endomembrane system. Mol. Cell. Biol. 13: 238–247.

Hicks, S. C., and R. A. Irizarry, 2015 *quantro*: a data-driven approach to guide the choice of an appropriate normalization method. Genome Biol. 16: 1–8.

Hillis, D. M., and J. J. Bull, 1993 An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42: 182–192.

Howell, S. H., 2013 Endoplasmic reticulum stress responses in plants. Annu. Rev. Plant Biol. 64: 477–499.

Jarvis, P., and E. López-Juez, 2013 Biogenesis and homeostasis of chloroplasts and other plastids. Nat. Rev. Mol. Cell Biol. 14: 787–802.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, 2015 The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10: 845–858.

Krsticevic, F., H. Santos, S. Januário, C. Schrago, and A. Carvalho, 2010 Functional copies of the Mst77F gene on the Y chromosome of *Drosophila melanogaster*. Genetics 184: 295–307.

LaFayette, P., and R. Travis, 1990 Soluble and membrane-associated heat shock proteins in soybean root. Protoplasma 156: 174–182.

Lambert, W., P. J. Koeck, E. Ahrman, P. Purhonen, K. Cheng *et al.*, 2011 Subunit arrangement in the dodecameric chloroplast small heat shock protein hsp21. Protein Sci. 20: 291–301.

Larkin, M. A., G. Blackshields, N. Brown, R. Chenna, P. A. McGettigan *et al.*, 2007 Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

Lawrence, S., K. Cline, and G. Moore, 1997 Chromoplast development in ripening tomato fruit: identification of cDNAs for chromoplast-targeted proteins and characterization of a cDNA encoding a plastid-localized low-molecular-weight heat shock protein. Plant Mol. Biol. 33: 483–492.

Lee, J. M., J.-G. Joung, R. McQuinn, M.-Y. Chung, Z. Fei *et al.*, 2012 Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor slerf6 plays an important role in ripening and carotenoid accumulation. Plant J. 70: 191–204.

Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Lopes-Caitar, V. S., M. C. de Carvalho, L. M. Darben, M. K. Kuwahara, A. L. Nepomuceno *et al.*, 2013 Genome-wide analysis of the hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. BMC Genomics 14: 1–17.

Low, D., K. Brandle, L. Nover, and C. Forreiter, 2000 Cytosolic heat-stress proteins Hsp17.7 class I and Hsp17.3 class II of tomato act as molecular chaperones in vivo. Planta 211: 575–582.

Lynch, M., and A. Force, 2000 The probability of duplicate gene preservation by subfunctionalization. Genetics 154: 459–473.

MacCarthy, T., and A. Bergman, 2007 The limits of subfunctionalization. BMC Evol. Biol. 7: 1–14.

Matas, A. J., T. H. Yeats, G. J. Buda, Y. Zheng, S. Chatterjee *et al.*, 2011 Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. Plant Cell 23: 3893–3910.

Montgomerie, S., J. A. Cruz, S. Shrivastava, D. Arndt, M. Berjanskii *et al.*, 2008 Proteus2: a web server for comprehensive protein structure prediction and structure-based annotation. Nucleic Acids Res. 36: W202–W209.

Neta-Sharir, I., T. Isaacson, S. Lurie, and D. Weiss, 2005 Dual role for tomato heat shock protein 21: protecting photosystem II from oxidative stress and promoting color changes during fruit maturation. Plant Cell 17: 1829–1838.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer, Berlin.

Ouyang, Y., J. Chen, W. Xie, L. Wang, and Q. Zhang, 2009 Comprehensive sequence and expression profile analysis of hsp20 gene family in rice. Plant Mol. Biol. 70: 341–357.

Poulain, P., J.-C. Gelly, and D. Flatters, 2010 Detection and architecture of small heat shock protein monomers. PLoS One 5: e9990.

Prasinos, C., K. Krampis, D. Samakovli, and P. Hatzopoulos, 2005 Tight regulation of expression of two *Arabidopsis* cytosolic Hsp90 genes during embryo development. J. Exp. Bot. 56: 633–644.

Ramakrishna, W., Z. Deng, C. K. Ding, A. K. Handa, and R. H. Ozminkowski, 2003 A novel small heat shock protein gene, vis1, contributes to pectin depolymerization and juice viscosity in tomato fruit. Plant Physiol. 131: 725–735.

Reddy, P. S., P. B. Kavi Kishor, C. Seiler, M. Kuhlmann, L. Eschen-Lippold *et al.*, 2014 Unraveling regulation of the small heat shock proteins by the heat shock factor hvhsfb2c in barley: its implications in drought stress response and seed development. PLoS One 9: e89125.

Riechmann, J. L., J. Heard, G. Martin, L. Reuber, C.-Z. Jiang *et al.*, 2000 *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290: 2105–2110.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140.

Sanmiya, K., K. Suzuki, Y. Egawa, and M. Shono, 2004 Mitochondrial small heat-shock protein enhances thermotolerance in tobacco plants. FEBS Lett. 557: 265–268.

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

Siddique, M., S. Gernhard, P. von Koskull-Döring, E. Vierling, and K.-D. Scharf, 2008 The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. Cell Stress Chaperones 13: 183–197.

Srivastava, A., A. Gupta, T. Datsenka, A. K. Mattoo, and A. K. Handa, 2010 Maturity and ripening-stage specific modulation of tomato (*Solanum lycopersicum*) fruit transcriptome. GM Crops 1: 237–249.

Sticher, L., A. K. Biswas, D. S. Bush, and R. L. Jones, 1990 Heat shock inhibits α-amylase synthesis in barley aleurone without inhibiting the activity of endoplasmic reticulum marker enzymes. Plant Physiol. 92: 506–513.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar, 2013 Mega6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.

Tang, H., J. E. Bowers, X. Wang, R. Ming, M. Alam *et al.*, 2008 Synteny and collinearity in plant genomes. Science 320: 486–488.

The Tomato Genome Consortium, 2012 The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635–641.

Trapnell, C., L. Pachter, and S. L. Salzberg, 2009 TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111.

Walter, P., and D. Ron, 2011 The unfolded protein response: from stress pathway to homeostatic regulation. Science 334: 1081–1086.

Waters, E. R., 1995 The molecular evolution of the small heat-shock proteins in plants. Genetics 141: 785–795.

Waters, E. R., and E. Vierling, 1999 The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. Mol. Biol. Evol. 16: 127–139.

Waters, E. R., G. J. Lee, and E. Vierling, 1996 Review article: evolution, structure and function of the small heat shock proteins in plants. J. Exp. Bot. 47: 325–338.

Waters, E. R., B. D. Aevermann, and Z. Sanders-Reed, 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.

Xia, K., T. Liu, J. Ouyang, R. Wang, T. Fan *et al.*, 2011 Genome-wide identification, classification, and expression analysis of autophagy-associated gene homologues in rice (*Oryza sativa* L.). DNA Res. 18: 363–377.

Yang, Z., Y. Zhou, X. Wang, S. Gu, J. Yu *et al.*, 2008 Genomewide comparative phylogenetic and molecular evolutionary analysis of tubby-like protein family in *Arabidopsis*, rice, and poplar. Genomics 92: 246–253.

Zeng, Y., Z. Pan, L. Wang, Y. Ding, Q. Xu *et al.*, 2014 Phosphoproteomic analysis of chromoplasts from sweet orange during fruit ripening. Physiol. Plant. 150: 252–270.

Zhao, C., M. Shono, A. Sun, S. Yi, M. Li *et al.*, 2007 Constitutive expression of an endoplasmic reticulum small heat shock protein alleviates endoplasmic reticulum stress in transgenic tomato. J. Plant Physiol. 164: 835–841.

Zheng, D., and M. B. Gerstein, 2007 The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? Trends Genet. 23: 219–224.

*Communicating editor: E. Akhunov*