New Phytologist

Research

Natural variation in HsfA2 pre-mRNA splicing is associated with changes in thermotolerance during tomato domestication

Yangjie Hu[1], Anida Mesihovic[1], José M. Jiménez-Gómez[2,3], Sascha Röth[1], Philipp Gebhardt[1], Daniela Bublak[1], Arnaud Bovy[4], Klaus-Dieter Scharf[1], Enrico Schleiff[1,5,6,7] and Sotirios Fragkostefanakis[1]

[1]Department of Biosciences, Molecular Cell Biology of Plants, Goethe University, D-60438, Frankfurt am Main, Germany; [2]Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, RD10, 78026, Versailles Cedex, France; [3]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829, Cologne, Germany; [4]Plant Breeding, Wageningen University, Wageningen 6708PB, the Netherlands; [5]Cluster of Excellence Frankfurt, Goethe University, D-60438, Frankfurt am Main, Germany; [6]Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University, D-60438, Frankfurt am Main, Germany; [7]Frankfurt Institute of Advanced Studies (FIAS), D-60438, Frankfurt am Main, Germany

Author for correspondence:
Enrico Schleiff
Tel: +49 69 798 29287
Email: schleiff@bio.uni-frankfurt.de

Received: *5 June 2019*
Accepted: *18 September 2019*

New Phytologist (2020) **225:** 1297–1310
doi: 10.1111/nph.16221

Key words: acclimation, high temperature, polymorphism, pre-mRNA splicing, *Solanum*, stress response.

Summary

- Wild relatives of crops thrive in habitats where environmental conditions can be restrictive for productivity and survival of cultivated species. The genetic basis of this variability, particularly for tolerance to high temperatures, is not well understood. We examined the capacity of wild and cultivated accessions to acclimate to rapid temperature elevations that cause heat stress (HS).
- We investigated genotypic variation in thermotolerance of seedlings of wild and cultivated accessions. The contribution of polymorphisms associated with thermotolerance variation was examined regarding alterations in function of the identified gene.
- We show that tomato germplasm underwent a progressive loss of acclimation to strong temperature elevations. Sensitivity is associated with intronic polymorphisms in the HS transcription factor HsfA2 which affect the splicing efficiency of its pre-mRNA. Intron splicing in wild species results in increased synthesis of isoform HsfA2-II, implicated in the early stress response, at the expense of HsfA2-I which is involved in establishing short-term acclimation and thermotolerance.
- We propose that the selection for modern HsfA2 haplotypes reduced the ability of cultivated tomatoes to rapidly acclimate to temperature elevations, but enhanced their short-term acclimation capacity. Hence, we provide evidence that alternative splicing has a central role in the definition of plant fitness plasticity to stressful conditions.

Introduction

Plants have developed mechanisms to sense and respond to suboptimal temperature conditions (Neumann *et al.*, 1989). In particular, temperatures 10–15°C above the optimum for growth and development cause heat stress (HS) (Wahid *et al.*, 2007). HS can pose a threat for survival and productivity of species or even whole ecosystems. Even high-bred strains show enhanced sensitivity to stressful conditions that may eventually amount to 50–80% yield losses (Neumann *et al.*, 1989; Zhao *et al.*, 2017). Understanding the molecular and genetic basis of thermotolerance is essential for the generation of HS-resilient crops that are currently under threat as a result of global warming.

Acclimation and recovery from HS depend on the activation of cellular stress response mechanisms, collectively called HS response (HSR). HSR finely tunes the activity of various gene and protein networks according to the cellular demands. Most characteristic of HSR is the accumulation of heat shock proteins (Hsps), which act as molecular chaperones to protect proteome integrity (Parsell & Lindquist, 1993). The core regulatory factors for HSR and thermotolerance are the HS transcription factors (Hsfs), required for the induction of the majority of HS-induced genes (Scharf *et al.*, 2012). Today, > 20 Hsfs have been identified in various plant genomes (Scharf *et al.*, 2012; Berz *et al.*, 2019).

The genetically predisposed limits for thermotolerance against a direct and often lethal HS can be exceeded by pre-exposure to moderately high temperature or gradual increase to an otherwise lethal temperature (Hong & Vierling, 2000). Key factors involved in HS acclimation have been discovered based on forward and reverse genetic approaches that have mainly uncovered the crucial role of specific Hsfs and Hsps (Hong & Vierling, 2000; Queitsch *et al.*, 2000; Mishra *et al.*, 2002; Larkindale *et al.*, 2005; Charng *et al.*, 2006; Wu *et al.*, 2013; Fragkostefanakis *et al.*, 2016, 2018; Lämke *et al.*, 2016). As acclimation mechanisms are believed to be ancient, the highly conserved functions of such key thermotolerance players in different species is not a surprise (Hong & Vierling, 2000; Queitsch *et al.*, 2000; Charng

et al., 2006; Schramm *et al.*, 2006; Wu *et al.*, 2013; Fragkostefanakis *et al.*, 2016).

However, natural variation in the thermotolerance of plants has not been well explored and consequently knowledge on alleles related to enhanced acclimation capacity is scarce. In particular, the question of whether HS acclimation capacity has been altered during domestication, improvement and spreading of crops to a broader geographical range has not been widely addressed. It was recently shown that selection for specific haplotypes of the quantitative locus TT1 encoding for an α2 subunit of the 26S proteasome has played a major role in thermotolerance during rice evolution (Li *et al.*, 2015). The discovery of such thermotolerance loci in important crops will boost efforts for the generation of stress-resilient elite germplasm. Wild crop relatives in many cases thrive under less favourable conditions than domesticated crops and are considered a valuable source of exotic loci for crop improvement (Hajjar & Hodgkin, 2007; Bolger *et al.*, 2014).

Tomato (*Solanum lycopersicum* var. *lycoperiscum*) is an important crop plant worldwide and serves as a model for elaboration of the fleshy fruit development and molecular and physiological HSR mechanisms (Mishra *et al.*, 2002; Frank *et al.*, 2009). The domesticated tomato originated in South America and nowadays is mainly cultivated in areas affected by global warming (Lobell & Field, 2007; Blanca *et al.*, 2015). As one aspect of global warming is the increase in frequency, intensity and duration of heatwaves (Easterling *et al.*, 2000; Rahmstorf & Coumou, 2011), we looked for mechanisms involved in tomato acclimation to HS in wild and domesticated genotypes. We provide evidence that natural variation in acclimation capacity of tomato has been altered in the course of domestication. We show that this phenomenon is associated with variation in alternative splicing of HsfA2 pre-mRNA, linking an important regulatory mechanism to plant fitness plasticity under stress conditions.

Materials and Methods

Plant material

Tomato accessions were obtained from the Charles M. Rick Tomato Genetics Resource Center at the University of California, Davis (Supporting Information Table S1; www.tgrc.ucdavis.edu). The *Solanum pennellii* introgression lines (ILs) were provided by Wageningen University (the Netherlands). ILs of chromosome 8 used in this study (IL8-1, LA3509; IL8-2, LA3510; IL8-3, LA3511) were selected from a collection of ILs generated by crossing the tomato cv M82 (LA3475) with *S. pennellii* (LA716) (Eshed & Zamir, 1995). By sequence analysis we confirmed that the introgression carrying *S. pennellii* HsfA2 is IL8-1. HsfA2 RNA interference (RNAi) lines -3 and -7 have been previously characterized (Fragkostefanakis *et al.*, 2016).

Generation of CRISPR/Cas9 mutants

The single guiding RNAs (sgRNAs) for HsfA2 (UCCGACGGCCGUGCUGCCUA) for CRISPR/Cas9-mediated plant genome editing was selected using CRISPR-PLANT (Xie *et al.*, 2014). Plasmid constructs were assembled via Golden Gate cloning technology (Weber *et al.*, 2011) according to SynBio (www.synbio.tsl.ac.uk) instructions. Vectors included pICSL01009::AtU6p (Addgene #46968) plasmid and plasmids from the MoClo Toolkit (Addgene #1000000044) (Engler *et al.*, 2014). The binary vector (pICSL002208) for plant transformation was kindly provided by Dr Nicola Patron (Earlham Institute, UK), contained between the left and right borders as follows: a sgRNA under the control of the *Arabidopsis thaliana* U6 small nuclear RNA (snRNA) promoter, *Streptococcus pyogenes* Cas9 controlled by the CaMV 35S promoter and kanamycin resistance gene neomycin phosphotransferase II (nptII) controlled by the CaMV 35S for selection based on kanamycin resistance. Final vectors were transformed into *Agrobacterium tumefaciens* strain GV3101. *Agrobacterium*-mediated transformation of *S. lycopersicum* cv Moneymaker cotyledons was performed as described previously (McCormick *et al.*, 1986). Genotyping was performed initially by genomic DNA sequencing by amplification of the CRISPR/Cas9-predicted editing region with specific primers (Table S2). Experiments were performed on homozygous, T-DNA-free plants on T3 generation.

Seedling thermotolerance

Seedling thermotolerance was determined by the relative hypocotyl elongation assay following heat stress treatment as previously described (Fragkostefanakis *et al.*, 2016). Four-day-old etiolated seedlings germinated and grown on wet paper towels, under a constant temperature of 25°C, were subjected to 37.5°C stress treatment for 90 min, to allow induction and activation of HsfA2, and directly exposed for another 90 min to 50°C stress. Treatments were performed in sealed Petri dishes in water baths. A 37.5°C treatment did not yield any significant effect for any of the accessions examined (data not shown) while a direct 90 min, 50°C treatment was lethal for all accessions. Seedlings kept for the same time at 25°C served as controls. Following the stress treatment, the seedlings were allowed to recover in the dark at 25°C for 3 d. Seedlings were photographed every day and hypocotyl length for stressed and control seedlings was determined by ImageJ. Thermotolerance is presented as relative hypocotyl elongation of the stressed seedlings compared with the control for each accession to avoid minor growth variations as a result of genetic background (Fig. S1).

Identification of splice variants and expression analysis

The presence of HsfA2 transcript variants was examined by PCR using specific primers (Table S2). Total RNA was extracted using the E.Z.N.A. Plant RNA Kit (Omega Biotek, Norcross, GA, USA) following the manufacturer's instructions. The presence of genomic DNA was inspected by control reactions without reverse transcriptase (–RT). cDNA was synthesized with 1 μg of total RNA using the Revert Aid reverse transcriptase (Thermo Scientific, Waltham, MA, USA). The PCR fragments corresponding to putative splice variants were subsequently sequenced using specific primers (GATC Biotech AG, Konstanz, Germany). Expression analysis of selected genes was done by quantitative real-time polymerase chain

reaction (qRT-PCR) on a Stratagene Mx3000P cycler (Agilent Technologies, Palo Alto, CA, USA). The qRT-PCR reaction consisted of gene-specific primers (Table S2), PerfeCTa SYBR Green FastMix Low ROX (Quanta Biosciences, Beverly, MA, USA), and the cDNA template. Thermal cycling conditions were 95°C for 3 min followed by 95°C for 15 s, 60°C for 30 s, and 72°C for 30 s for 40 cycles. Elongation factor 1-alpha (Solyc06g005060) and ubiquitin (Solyc07g064130) were used as housekeeping genes (Fragkostefanakis *et al.*, 2014, 2016). Data were analyzed using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001).

Immunodetection analysis was performed using an antibody generated against a recombinant full-length HsfA2 (Scharf *et al.*, 1998; Mishra *et al.*, 2002). The green fluorescent protein (GFP)-HsfA2 isoforms in the minigene system were detected by a monoclonal GFP antibody (Roche) and a mouse horseradish peroxidase (HRP) conjugated secondary antibody (Thermo Fischer Scientific, Waltham, MA, USA). HsfA2 isoforms were detected using a polyclonal anti-SpHsfA2 and a rabbit HRP conjugated secondary antibody (Thermo Fischer Scientific). In brief, *c.* 100 mg of homogenized seedlings were used for protein extraction (Mishra *et al.*, 2002). Extracts containing *c.* 15–20 μg of protein were separated on a 10% sodium dodecyl sulfate-polyacrylamide gel. Proteins were transferred to nitrocellulose membrane (Protran nitrocellulose transfer membrane; Whatman Ltd, Little Chalfont, UK) and further processed for chemiluminescence detection following the manufacturer's protocol (Perkin Elmer, Waltham, MA, USA). Anti-actin (Sigma) was used for verification of equal loading.

Transcriptional activation activity assay

The coding sequences of HsfA2-Iα and HsfA2-II isoforms were amplified from cDNA of the indicated genotype and cloned into the *Sal*I–*Xba*I sites of pRT-HA or *Acc*65I–*Xba*I sites of the pRT-GFP vector (Röth *et al.*, 2016). The transcriptional activation activity of the isoforms was examined in tomato mesophyll protoplasts using the reporter Hsf-dependent promoter *PGmhsp17.3BCI*::GUS (Treuter *et al.*, 1993). Fifty thousand protoplasts were transformed with 0.5 μg plasmid DNA carrying the appropriate Hsf expression construct and 1 μg of β-glucuronidase (GUS) reporter plasmid DNA. The total DNA amount was adjusted to 10 μg by adding pRT-Neo plasmid encoding for the Neomycin resistance gene as mock plasmid DNA (Scharf *et al.*, 1998). Measurements were taken for several time points. 4-methylumbelliferone was measured in the 'Fluostar' fluorometer (BMG LabTechnologies GmbH, Offenburg, Germany). Background fluorescence was subtracted and values were calculated relative to samples transformed only with GUS reporter and Neo mock control. Nuclear export was inhibited by leptomycin B (LMB; final concentration 22 ng ml^{-1}), as previously described (Kudo *et al.*, 1998; Port *et al.*, 2004). LMB was added in protoplast samples 3 h before harvesting.

Protein stability

The stability of the two HsfA2 isoforms was determined in a chase experiment in tomato protoplasts as previously described (Röth *et al.*, 2016). Transformed protoplasts with plasmids carrying the GFP-HsfA2-I and GFP-HsfA2-II expression cassettes were treated with 20 μg ml^{-1} of cycloheximide, 4 h after transformation. Each time point was an individual transformation event. Samples were harvested at the indicated time points and quantification was performed on the basis of an immunoblot analysis. Band intensities were quantified by ImageJ after background subtraction (Abràmoff *et al.*, 2004). Protein abundances were calculated as relative to time point zero. Values are the average of three independent experiments.

Subcellular localization of HsfA2 isoforms

The subcellular distribution of GFP-HsfA2 fusions expressed in tomato protoplasts was examined under a Leica SP5 confocal laser-scanning microscope (Leica Microsystems, Heidelberg, Germany). An ENP1-mCherry fusion protein was used as a nuclear marker. An Hsp17.4-CII-mCherry construct was used to study the colocalization of the Hsp17.4-CII with HsfA2 isoforms. Excitation of GFP was at 488 nm and fluorescence emission at 490–548 nm, while the corresponding wavelengths for mCherry were 561 and 570–656 nm. Images are representative of several individual transformation events yielding the same results.

Minigene splicing reporter assay

Minigene reporters contained a genomic segment of the HsfA2 gene from either *S. lycopersicum* cv Moneymaker or *S. peruvianum*, including the intron of interest, flanked by the 184 nt 3′-end of exon 2 and the 135 nt 5′-end of exon 3. The HsfA2 minigene was fused in frame to the carboxyl-terminal of GFP. The genomic segment was generated by PCR amplification with high-fidelity Pfu polymerase, using genomic DNA of *S. lycopersicum* cv Moneymaker or *S. peruvianum* as a template. The two amplified DNA fragments were cloned between the *Acc*65I and *Xba*I sites of the pRT-dS vector.

Site-directed mutagenesis to investigate the contribution of each polymorphism in splicing efficiency of HsfA2 was performed using QuickChange mutagenesis PCR (Liu & Naismith, 2008).

Protoplasts transformed with the appropriate plasmids carrying the expression cassettes of the minigenes were allowed to generate GFP-HsfA2 mRNA and then exposed to 37.5°C for 1 h. The splicing efficiency of the minigene constructs was examined by PCR, using a GFP-specific forward and an exon 3 reverse primer pair. The intensity of each band corresponding to a different transcript variant was quantified by ImageJ using DNA agarose gels stained with ethidium bromide.

Genomic databases

HsfA2 co-orthologs in different plant species have been reported by Scharf *et al.* (2012) and can be found in the HEATSTER database (Berz *et al.*, 2019). The presence or absence of a nuclear export signal (NES) is described therein.

Genome-wide association study

Forty-two accessions for which resequencing data were available were measured for thermotolerance. We used GATK UnifiedGenotyper v.3.8 (Depristo *et al.*, 2011) to obtain a set of genome-wide variants in these accessions. From this set, we ended up with 1469 000 variants that were single nucleotide polymorphisms (SNPs), biallelic, genotyped in at least 41 of the 42 accessions, with a minimum allele frequency of 5% in our dataset. We then used Emmax (Kang *et al.*, 2010) with an identity by state kinship matrix and default parameters to calculate associations between each variant and the thermotolerance phenotype. Associations with $P < 10^{-4}$ were considered to be significant. Linkage disequilibrium (LD) for all polymorphisms of HsfA2 was calculated using Plink v.1.07 (Chang *et al.*, 2015).

Classification of resequenced tomato accessions in phylogenetic groups

Whole-genome sequencing data are available for 601 accessions (Zhu *et al.*, 2018). Short reads from all accessions were downloaded from NCBI-SRA and aligned to the *S. lycopersicum* reference genome (Tomato Genome Consortium, 2012) using Bowtie2 v.2.50 (Langmead & Salzberg, 2012) with default parameters. Reads mapping to multiple locations were removed using Samtools v.1.7 (parameter -q 5), duplicated reads were removed using Picard Tools' MarkDuplicates (http://broadinstitute.github.io/picard) and indels were realigned using GATK RealignerTargetCreator and IndelRealigner v.3.8 (Depristo *et al.*, 2011).

These 601 resequenced accessions where assigned to a phylogenetic group by comparison with the accessions classified in Blanca *et al.* (2015). In Blanca *et al.*'s work, 1008 tomato accessions were classified into phylogenetic groups using 8700 genome-wide SNPs genotyped with the SolCAP Infinium Chip (Blanca *et al.*, 2015). We genotyped the 601 resequenced accessions at the SolCAP Infinium array positions (indicated in the ITAG2.4_solCAP.gff3 file available at www.ftp.solgenomics.net) using GATK UnifiedGenotyper with default parameters. This list of variants was filtered to leave only biallelic SNPs, resulting in a total of 7486 variants, which were merged with the published matrix containing the genotypes for 1008 accessions (Blanca *et al.*, 2015). In all, 1630 variants that did not agree between the two datasets were removed, and the resulting list was filtered for loci in LD using the 'LD' pruning option of Plink with parameters --mind 0.1 --geno 0.1 --indep 50 5 2.8 (Chang *et al.*, 2015). A phylogenetic tree was estimated from the final matrix (1536 variants in 1609 accessions) using the ape package in R and the neighbor-joining method including *S. pennellii* LA0716 as a root (Paradis *et al.*, 2004). The resulting tree was plotted using the ggtree package in R (Yu *et al.*, 2017). Tomato accessions in the tree were classified manually taking into account the previously described classification (Blanca *et al.*, 2015) and their position in the tree (Table S3).

All 601 accessions were genotyped for the three intronic variants of interest (SL2.50ch08 52380445, 52380439 and 52380421) by calling SNPs simultaneously in all alignment files using GATK UnifiedGenotyper v.3.8.

Detection of introgressions from wild tomatoes in the region of HsfA2 in cultivated tomato

Accessions TS–175, TS–258 and TS–64 belong to the cultivated tomato Vintage group but presented the HsfA2^{GAG} haplotype characteristic of noncultivated accessions, which could result from an introgression from a wild species in the chromosomal region of HsfA2. We investigated this possibility by calculating the frequency of nonreference SNPs in all accessions from the same phylogroup (SLL Vintage in Table S3) in windows of 1000 SNPs and steps of 500. Analysis of these frequencies along chromosome 8 showed that these accessions did indeed present an introgression from a wild species that spans the centromere, includes the gene HsfA2 and is not present in any other vintage tomato accession.

Calculation of nucleotide diversity

We estimated nucleotide diversity (π) in the accessions sequenced in Zhu *et al.* (2018) belonging to the phylogenetic groups established in the previous section (Table S3). For this, we extracted a variant matrix using GATK with default parameters from a region spanning 10 Mb around the HsfA2 gene (from 47 to 57 Mb; HsfA2 is located at 52.4 Mb). We then used VCFtools (Danecek *et al.*, 2011) to calculate π in 10 kb windows with 1 kb steps on biallelic SNPs from each group of accessions (cultivated tomato, *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme*) as specified in Table S3. Regions showing signals of positive selection were defined as those whose π ratio was above a 95% threshold calculated for the whole chromosome.

Results

Regulation of rapid acclimation against heat stress by HsfA2

Heat stress acclimation capacity, also known as acquired thermotolerance, is examined by exposure of plants to a mild HS (MHS) temperature, followed by a second challenging HS (CHS) that, in the absence of MHS, is lethal (Larkindale & Vierling, 2007; Yeh *et al.*, 2012). A recovery of 1–3 h between MHS and CHS is referred to as short-term acquired thermotolerance treatment, while an interval longer than 24 h is referred to as long-term acquired thermotolerance treatment (Yeh *et al.*, 2012). To inspect the capacity of tomato wild and domesticated accessions to rapidly acclimate to HS conditions, 52 genotypes were exposed to MHS (37.5°C for 1.5 h) followed directly by CHS (50°C for 1.5 h) treatment (Fig. 1a, pictogram). The time and temperature of MHS allowed survival and recovery of the seedlings of all accessions upon direct CHS. The analysis included eight *S. lycopersicum* var. *lycopersicum* accessions (SLL), representing tomato modern cultivars, 15 *S. lycopersicum* var. *cerasiforme* (SLC) as semi-domesticated genotypes, 19

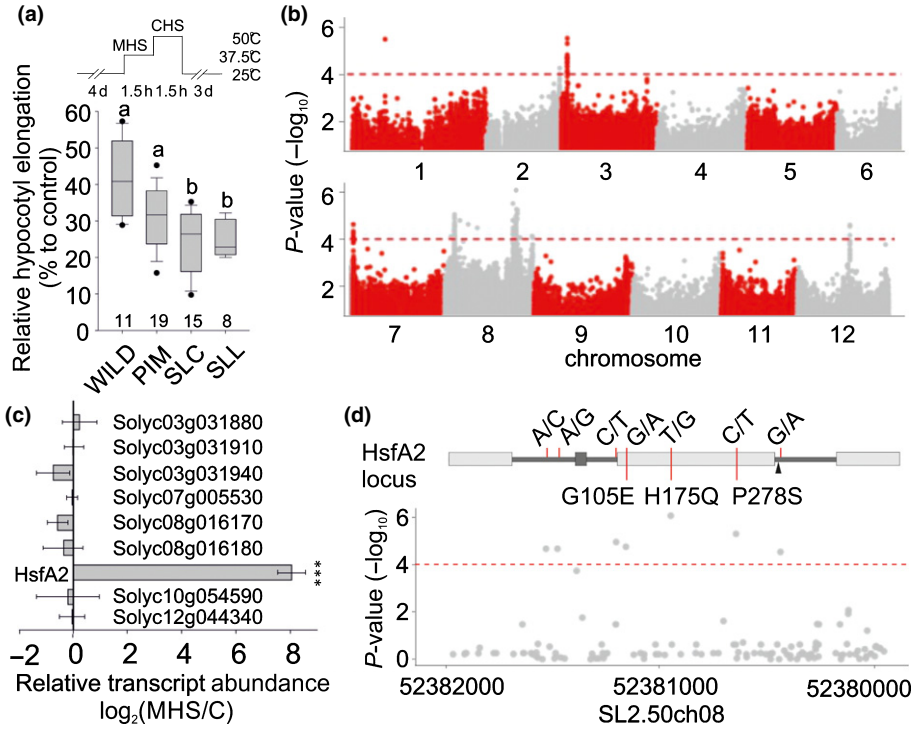

Fig. 1 Genome wide association study (GWAS) for thermotolerance of tomato seedlings. (a) Upper panel: pictogram illustrating the heat stress (HS) regime used for seedling thermotolerance. MHS, mild HS; CHS, challenging HS. Lower panel: hypocotyl elongation of seedlings from distant wild species (WILD; $n = 11$), *Solanum pimpinellifolium* (PIM; $n = 19$), *Solanum lycopersicum* var. *cerasiforme* (SLC; $n = 15$) and *Solanum lycopersicum* var. *lycopersicum* (SLL; $n = 8$). Box plots show the median and the upper quartiles. Significant differences are depicted by different letters based on one-way ANOVA ($P < 0.05$). (b) Manhattan plot of the GWAS for seedling thermotolerance. The dashed horizontal line shows the Bonferroni-adjusted significance threshold ($P = 10^{-4}$). (c) Quantitative real-time polymerase chain reaction (qRT-PCR) analysis showing the transcript abundances of genes with significant variants identified by GWAS in response to a MHS treatment in SLL cv Moneymaker. Differential expression is shown as $-\Delta\Delta$Ct of MHS vs control (C, 25°C) samples normalized to Elongation factor 1-alpha (EF1a) housekeeping gene. Error bars are the average of three independent replicates ± SD. Solyc03g031870 and Solyc03g031900 are not expressed. (d) Exon/intron structure and regional Manhattan plot of HsfA2 genomic region on chromosome 8. Light gray boxes indicate exons and dark gray line indicate introns. The dark gray box within intron 1 is a mini-exon. Significant single nucleotide polymorphisms are shown on the gene model.

S. pimpinellifolium (PIM) and 11 more distant wild accessions (WILD). Thermotolerance was determined as the relative hypocotyl elongation of seedlings following HS relative to non-stressed samples for each genotype. WILD and PIM accessions are significantly more thermotolerant than are SLC and SLL accessions (Figs 1a, S1; Table S1). This indicates a significant reduction in the capacity of domesticated and modern tomato to rapidly acclimate against HS as compared with wild species, at least for the experimental conditions applied here.

The allelic variation associated with thermotolerance was explored by a genome-wide association study (GWAS) on 42 resequenced accessions (Aflitos *et al.*, 2014; T. Lin *et al.*, 2014). Genome-wide biallelic SNPs genotyped in at least 41 out of the 42 accessions with a minimum allele frequency of 5% were selected for GWAS, resulting in a total of 1469 000 SNPs. Variants significantly associated with thermotolerance were resolved in 11 genes located in six genomic regions spanning chromosomes 3, 7, 8 and 12 (Fig. 1b; Table S4). The SNP that showed the highest association with thermotolerance was located in the gene Solyc08g062960 (Table S4) encoding HsfA2 (Scharf *et al.*, 1998; Fragkostefanakis *et al.*, 2016). Moreover, HsfA2 is the only HS-induced gene among this set of 11 genes in seedlings of *S. lycopersicum* cv Moneymaker exposed to MHS treatment (Fig. 1c). Detailed association analysis of the SNPs in the region encoding the HsfA2 gene resulted in seven significant SNPs, three in intron 1, three nonsynonymous in exon 2 and one in intron 2 (Fig. 1d).

Tomato HsfA2 is involved in acquired, but not basal, thermotolerance or response to acute HS, mainly acting as the major coactivator of the master regulator of HSR HsfA1a (Scharf *et al.*, 1998; Fragkostefanakis *et al.*, 2016). We first examined the involvement of HsfA2 in rapid acclimation capacity to HS that causes HS by generating CRISPR/Cas9 and RNAi suppression transgenic plants in SLL cv Moneymaker background. Two independent CRISPR mutants with homozygous indels in exon 1 causing a frame shift failed to produce HsfA2 protein upon MHS, while two independent RNAi lines accumulated reduced HsfA2 protein abundances (Figs 2a, b, S1c). Mutant and transgenic plants are more sensitive to the HS treatment than are wild-type plants (Fig. 2c), and accumulate lower levels of the HS-induced *Hsp17.7A-CI* and *Hsp70-5*, confirming the involvement of HsfA2 in response to MHS pretreatment (Fig. 2d).

We confirmed the existence of natural variation for HsfA2 by using an existing population of ILs containing individual

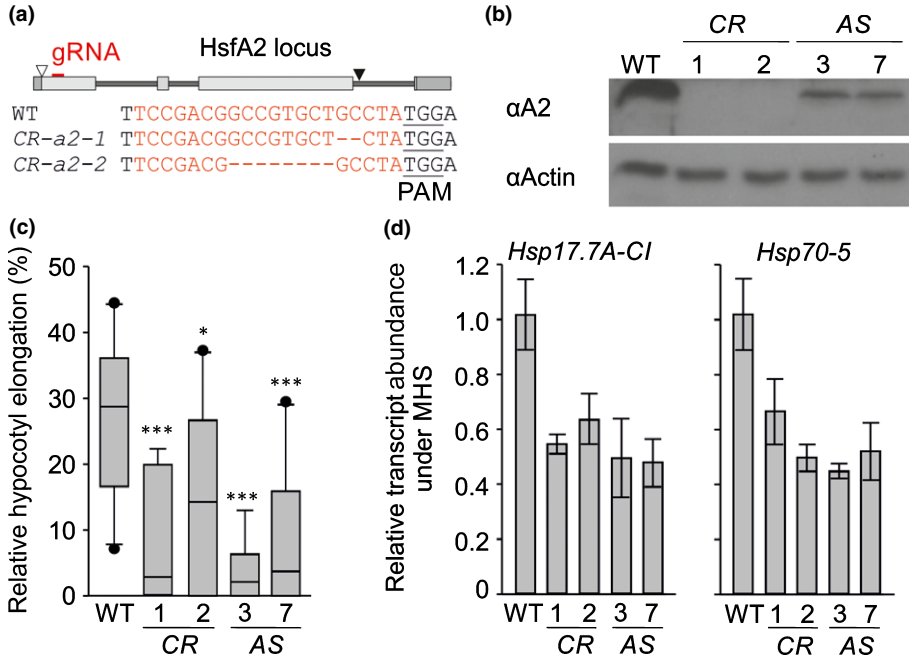

Fig. 2 HsfA2 is involved in response of seedlings to rapid heat stress (HS) acclimation. (a) HsfA2 gene structure (exons, gray boxes; introns, black lines). White and black triangles indicate position of initiation and termination codons, and the red line indicates the gRNA targeting position. (b) Immunodetection of HsfA2 in wild-type (WT), CRISPR mutants (*CR-hsfa2-1*, *CR-hsfa2-2*) and transgenic antisense (A2AS-3 and A2AS-7) seedlings exposed to 37.5°C for 1.5 h. β-Actin is shown as loading control. Note that the protein detected is HsfA2-I as discussed below. (c) Thermotolerance of WT, HsfA2 mutants and transgenic antisense plants as described for Fig. 1(a). Box plots represent data obtained from 30 seedlings from three independent experiments. Box plots show the median and upper quartiles. Significant differences between the transgenic or mutant lines with WT are depicted by asterisks (*, $P < 0.05$; ***, $P < 0.001$) based on *t*-test. (d) Relative transcript abundances of *Hsp17.7A-CI* and *Hsp70-5* in seedlings exposed to mild HS (MHS), determined by quantitative real-time polymerase chain reaction (qRT-PCR). Values are normalized against the WT ($n = 3$, ± SE).

genomic fragments from *S. pennellii* (LA716) in a cultivated tomato background (cv M82) (Eshed & Zamir, 1995). Line IL8-1 from this population carries the *S. pennellii* allele of HsfA2, whereas lines IL8-2 and IL8-3 carry the HsfA2 allele of M82 (Fig. S2a). Consistent with differential effects between the wild and the cultivated allele of HsfA2, line IL8-1 and *S. pennellii* showed enhanced thermotolerance compared with *S. lycopersicum* cv M82 and lines IL8-2 or IL8-3 (Fig. 3a). In agreement with this, we observed slightly but significantly higher *Hsp17.7A-CI* levels in IL8-1 compared with M82, which are similar to the level observed in *S. pennelii* (Fig. 3b). We assume that such a weak but significant increase in several HS-induced genes, including other members of the sHsp family, has a cumulative effect on the enhanced thermotolerance.

The effect of nonsynonymous polymorphisms of HsfA2 was examined in a GUS reporter assay in tomato mesophyll

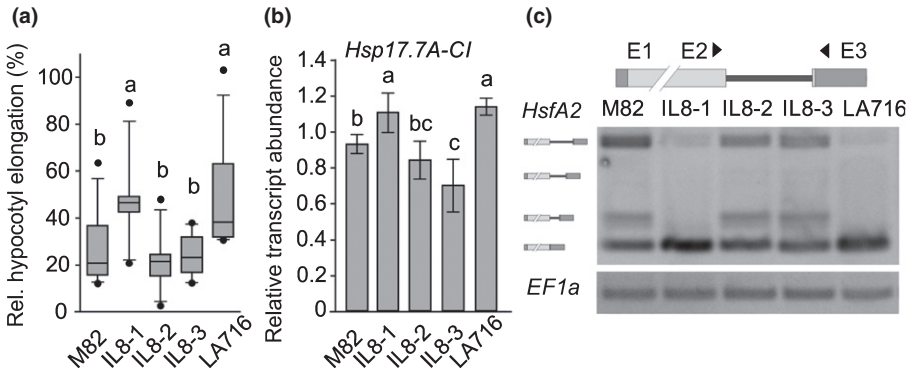

Fig. 3 Introgression lines with wild HsfA2 allele are more thermotolerant. (a) Thermotolerance of young seedlings from chromosome 8 introgression lines (ILs) and the parental lines M82 and LA716. Box plots show the median and the upper quartiles. (b) Quantitative real-time polymerase chain reaction (qRT-PCR) analysis of *Hsp17.7A-CI* in *Solanum pennellii* ILs and M82. Values are relative to the M82 mild heat stress (MHS) sample. Significant differences were calculated by ANOVA with Duncan test for $P < 0.05$ and are depicted by different letters ($n = 3$, ± SE). (c) RT-PCR analysis showing the splicing pattern of HsfA2 in seedlings of ILs and M82 exposed to MHS. Forward and reverse primers annealed to exon 2 and exon 3, respectively. Models for splice variants are shown on the left. Elongation factor 1-alpha (EF1a) was used as housekeeping gene.

protoplasts. HsfA2 from SLL (cv Moneymaker) was compared with WILD *S. peruvianum* (LA1955) containing the same non-synonymous polymorphisms G105E and H175Q (Fig. 1d) as *S. pennellii*. HsfA2 open reading frames from the two species were cloned under the control of CaMV 35S. We did not observe significant differences in the transactivation capacity of the SLL and WILD HsfA2 proteins when expressed alone or in combination with HsfA1a (Fig. S3). This indicates that the two nonsynonymous SNPs have no effect on HsfA2 function.

Consequently, we examined whether the intronic polymorphisms of the HsfA2 gene affect the pre-mRNA splicing of HsfA2. We performed RT-PCR on mRNA from seedlings exposed to MHS using pairs of primers amplifying exon-intron junctions specifically for intron 1 or intron 2. We did not observe any alternative splicing event for intron 1 (data not shown). However, we detected four DNA fragments corresponding to different HsfA2 splice variants in intron 2 (Fig. 3c). In cultivated tomato (var. M82), the shortest transcript is generated by full splicing of intron 2, and larger isoforms result from full or partial intron retention (Fig. 3c). As expected, the *S. pennellii* HsfA2 allele present in IL8-1 shows a different splicing pattern than the HsfA2 allele present in M82, IL8-2 and IL8-3 (Fig. 3c), and it is similar to the pattern observed in *S. pennellii* (Fig. S2b).

HsfA2 pre-mRNA is alternatively spliced in a temperature-dependent manner

The currently annotated tomato HsfA2 protein contains a C-terminal nuclear export signal that is encoded in the second intron (Scharf *et al.*, 1998), similar to *A. thaliana* HsfA2 (Kotak *et al.*, 2004). Therefore, this model is based on a transcript generated by intron retention. We performed a PCR using primers annealing in exons 1 and 3 to identify splicing patterns of HsfA2 pre-mRNA in seedlings of *S. lycopersicum* cv Moneymaker exposed to temperatures ranging from 25 to 50°C for 1 h (Fig. 4a). Partial or full retention of one or both introns results in the generation of six isoforms (Fig. 4a). An additional isoform appears during recovery from 42.5°C (Fig. S4a). The seven transcripts are classified into three groups (I–III) (Fig. 4b).

HsfA2-Iα, *-β* and *−γ* are generated by alternative splicing in intron 2 at 30–45°C (Fig. 4a). They encode for the previously annotated HsfA2 protein isoform with 351 amino acid residues (Fig. 4b). HsfA2-I is activated during repeated HS (Scharf *et al.*, 1998) cycles and is important for short- and long term acquired thermotolerance (Fragkostefanakis *et al.*, 2016). *HsfA2-II* observed in samples at 30–40°C is generated by splicing of both introns (Fig. 4a). HsfA2-II has a truncated C-terminal activation motif and lacks the nuclear export signal (NES; Fig. 4b,c). We confirmed the presence of HsfA2-II protein by immunodetection in *S. lycopersicum* cv Moneymaker (Fig. 4d). HsfA2-II is less abundant than HsfA2-I (Fig. 4d).

Group III transcripts (α, β, γ) are generated from the rarely observed alternative splicing in intron 1 (Fig. 4a,b). Similar alternative splicing has been shown for *A. thaliana* HsfA2 (Liu *et al.*, 2013). Alternative splicing of intron 1 leads to the formation of a premature termination codon (Fig. 4b). *HsfA2-III* transcripts are targeted for degradation, as suppression of nonsense mRNA decay (NMD) by the translation inhibitor cycloheximide (Carter *et al.*, 1995) enhances their accumulation (Fig. S4b). Therefore, in contrast to HsfA2-III in *A. thaliana* which becomes detectable only upon severe HS and encodes for a small protein with a truncated DNA-binding domain (DBD) (Liu *et al.*, 2013), in tomato such a protein probably does not exist.

Alternative splicing of HsfA2 pre-mRNA results in differential protein localization and stability

The coexistence of the nuclear localization signal and NES in HsfA2-I leads to nucleocytoplasmic shuttling of the protein with dominant cytosolic localization (Fig. 4b,e). For effective nuclear retention, HsfA2-I requires interaction with HsfA1a (Scharf *et al.*, 1998). The absence of an NES in HsfA2-II leads to HsfA1a-independent nuclear retention, as documented for GFP-HsfA2-II expressed in tomato protoplasts (Figs 4e, S5).

Coexpression of HsfA1a and HsfA2-I stimulates transcription of target genes in a synergistic manner (Fig. 4f). The activities of HsfA1a–HsfA2-II and HsfA1a–HsfA2-I complexes on the reporter construct *PGmHsp17-CI*::GUS are comparable (Figs 4f, S6). Thus, both isoforms can induce HS-responsive genes in cooperation with HsfA1a at the same level. HsfA2-II alone has higher activity than HsfA2-I owing to its strong nuclear retention, as inhibition of HsfA2-I nuclear export by leptomycin B leads to similar activity with HsfA2-II (Fig. 4f). Further, HsfA2-I exhibits higher protein stability than HsfA2-II, as indicated by the relative protein abundance in protoplasts treated with cycloheximide (Fig. 4g). Thus, nuclear retention of HsfA2-II leads to an increased turnover compared with HsfA2-I.

HsfA2-I interacts with Hsp17-CII proteins during HSR (Port *et al.*, 2004). This interaction represses the transcription factor function by recruitment of HsfA2-I in HS granules (HSGs) that serve as temporary deposition and storage sites for HsfA2-I (Scharf *et al.*, 1998; Port *et al.*, 2004). Accordingly, increasing amounts of Hsp17.4-CII in protoplasts lead to the accumulation of HsfA2-I in cytosolic foci resembling HSGs and to the lower transactivation activity of HsfA2-I (Fig. 4h,i). By contrast, HsfA2-II largely escapes the repressor effect of Hsp17.4-CII and only a minor fraction of HsfA2-II colocalizes with Hsp17.4-CII in perinuclear aggregates (Fig. 4h,i).

In conclusion, both HsfA2-I and HsfA2-II are active and stimulate the transcription of downstream genes in cooperation with HsfA1a. However, HsfA2-II is rapidly degraded while HsfA2-I is stored in HSGs and utilized to enhance acquired thermotolerance and acclimation in case of repeated HS cycles.

HsfA2 pre-mRNA splicing is associated with natural variation in thermotolerance

The difference in splicing efficiency in intron 2 of HsfA2 between the *S. pennellii* allele and the *S. lycopersicum* cv M82 allele prompted us to study the polymorphisms present in this intron (Fig. 1a). Our GWAS results pointed to a SNP in intron 2

New Phytologist

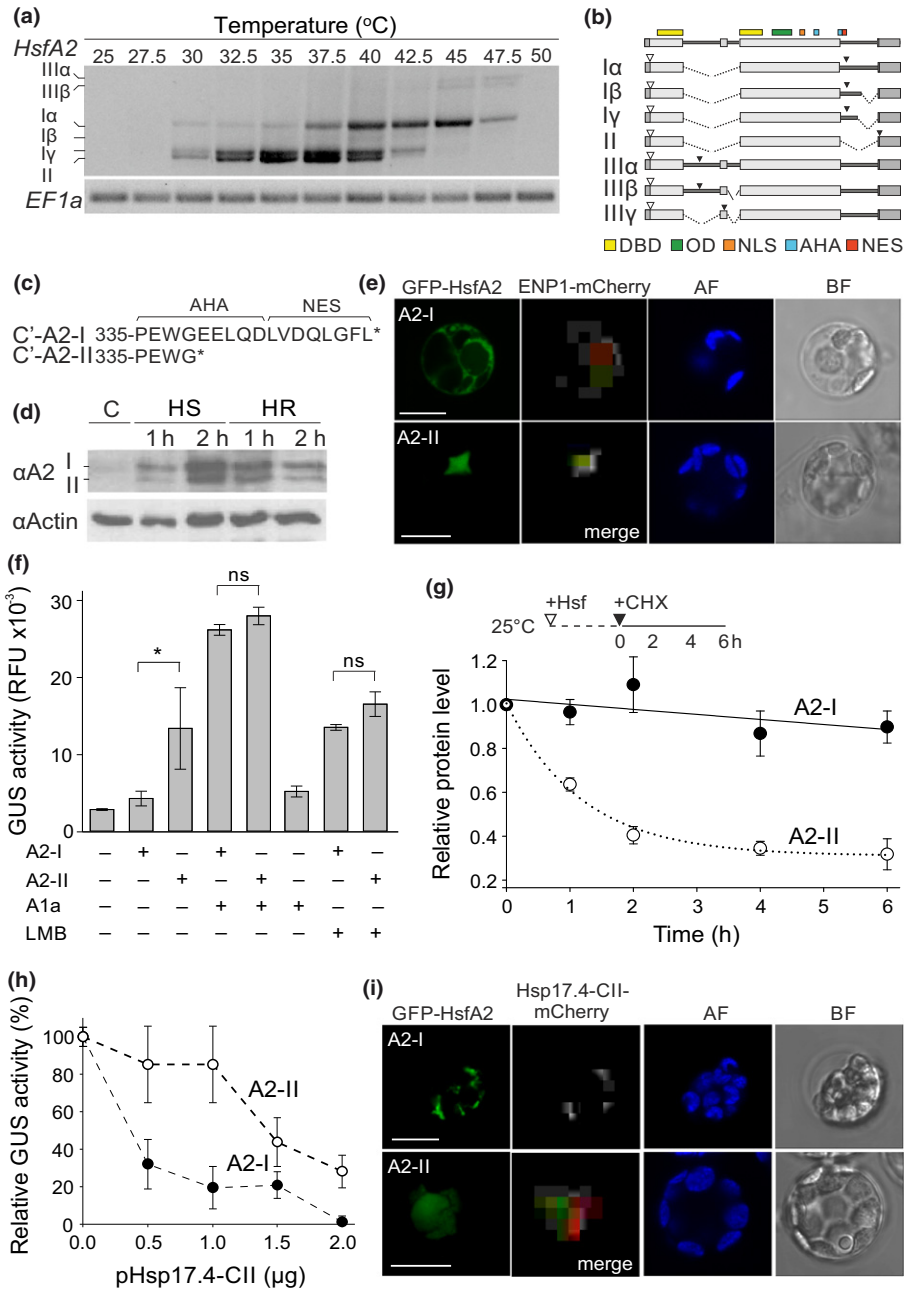

Fig. 4 HsfA2 isoforms I and II differ in their properties. (a) Real-time polymerase chain reaction (RT-PCR) of HsfA2 on cDNA from *Solanum lycopersicum* var. *lycopersicum* cv Moneymaker seedlings exposed to different temperatures for 1 h. Elongation factor 1-alpha (EF1a) serves as control. (b) Exon (thick bars)/intron (thin bars) structure of HsfA2 locus and transcript variants. White and black triangles depict initiation and termination codons, respectively. Dark gray bars are 5′- and 3′-untranslated region. DBD, DNA-binding domain; OD, oligomerization domain; NLS, nuclear localization signal; AHA, activation motif rich in aromatic, large hydrophobic, acidic amino acids, NES, nuclear export signal. Dashed lines indicate splicing. (c) Amino acid sequence of the C-terminus of HsfA2 isoforms. (d) Immunodetection of HsfA2 isoforms in seedlings exposed to 37.5°C for 1 or 2 h (heat stress (HS)), or after 1 h exposure to 37.5°C followed by 1 or 2 h recovery at 25°C (heat recovery (HR)). Actin is shown as control. (e) Localization of green fluorescent protein (GFP)-HsfA2-I and GFP-HsfA2-II alone and as overlay with ENP1-mCherry (nuclear marker) in tomato protoplasts. AF, autofluorescence; BF, bright field. Bars, 10 μm. (f) β-Glucuronidase (GUS) reporter transactivation assay of HsfA2-I or HsfA2-II alone or in combination with HsfA1a ($n = 3$, ± SD). Asterisk indicates statistically significant difference ($P < 0.05$) based on Student's *t*-test. LMB, leptomycin B; ns, not significant. (g) Stability of HsfA2 isoforms ($n = 3$; ± SE). T_0, cycloheximide was added to protoplasts expressing HsfA2-I or HsfA2-II. Lines represent the least-squares fit analysis with an exponential equation. Values are the average of three replicates ($n = 3$) and error bars are SE. (h) Relative activity of HsfA2-I and HsfA2-II upon coexpression with increasing amounts of plasmid encoding for Hsp17.4-CII in protoplasts. (i) Colocalization of GFP-HsfA2-I and GFP-HsfA2-II fusions with Hsp17.4-CII-mCherry in protoplasts. Bars, 10 μm.

associated with the thermotolerance phenotype (position SL2.50ch08:52380439; $P = 2.791 \times 10^{-5}$; Fig. 1d). This nucleotide is A in WILD, G or A in PIM and SLC, and G in SLL genotypes (Fig. 5a). Interestingly, the GT dinucleotide that serves as a donor splicing site for *HsfA2-Iγ* in *SLL* (SL2.50ch08:52380421; Fig. 5a) is changed to AT in WILD, which explains the absence of *HsfA2-Iγ* in IL8-1 and *S. pennellii* (Fig. 3c). A third SNP at position 52380445 shows complete LD with the SNP at position 52380421 (Fig. 5a). Consequently, in our dataset we find three haplotypes for intron 2: HsfA2^{AAA} present only in WILD; HsfA2^{GAG} in PIM and SLC; and HsfA2^{GGG} in PIM, SLC or SLL (Fig. 5a).

The splicing profile of HsfA2 intron 2 was examined in 33 tomato genotypes carrying different haplotypes at this intron. All distant wild tomatoes having the HsfA2^{AAA} haplotype lack *HsfA2-Iγ* and show significantly enhanced splicing of intron 2, with HsfA2-II transcripts comprising *c.* 90% of the total HsfA2 transcripts (Fig. 5b,c). Compared with this, retention of intron 2

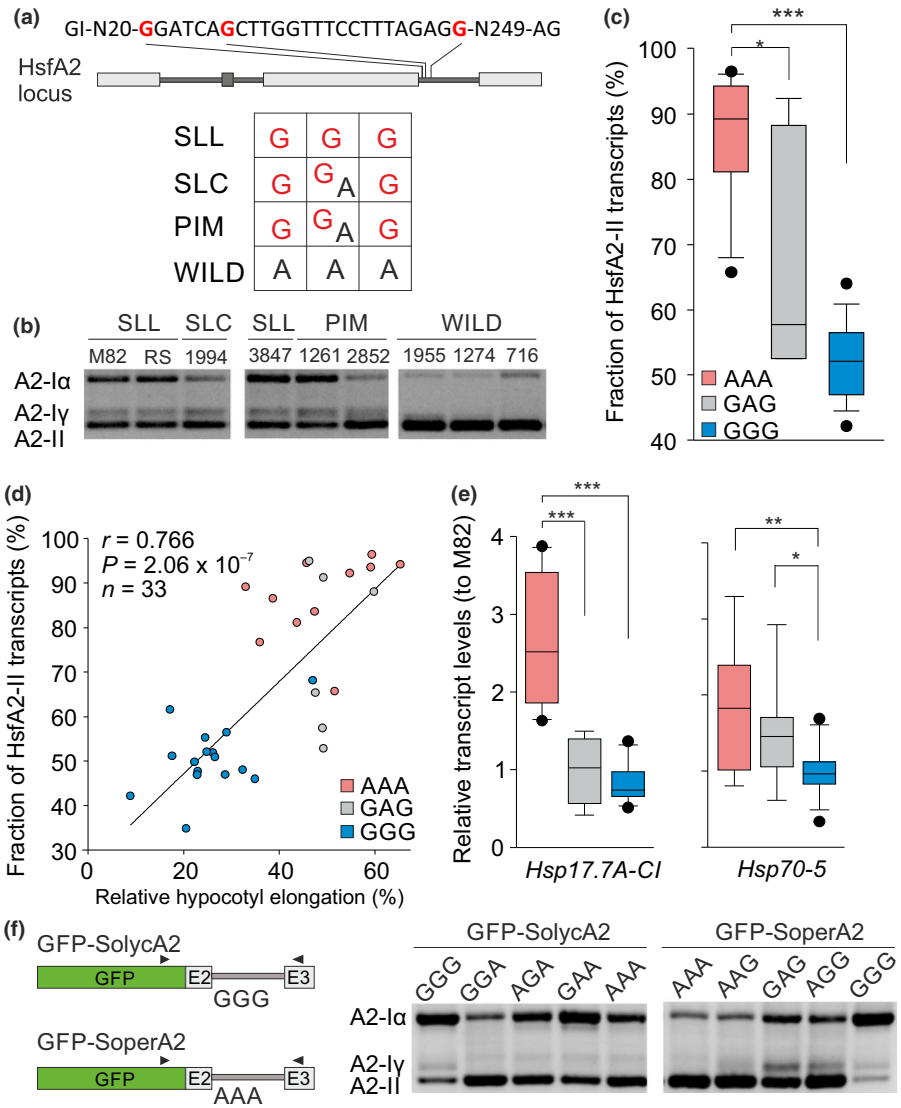

Fig. 5 HsfA2 intron 2 polymorphisms are associated with thermotolerance. (a) Positions of G/A polymorphisms in tomato species grouped as in Fig. 1(a). On top the positions of the SNPs in HsfA2 gene is shown. (b) Real-time polymerase chain reaction (RT-PCR) on cDNA from seedlings exposed to 37.5°C for 1 h. Numbers refer to LA accession numbers from the Charles M. Rick Tomato Genetics Resource Center. RS, red setter. (c) Splicing efficiency of HsfA2^{AAA} ($n = 11$), HsfA2^{GAG} ($n = 7$) and HsfA2^{GGG} ($n = 15$) haplotypes calculated as a fraction of *HsfA2-II* to total *HsfA2* transcripts based on RT-PCR analysis. (d) Linear regression of relative hypocotyl elongation in response to mild/challenging heat stress (MHS/CHS) treatment and splicing efficiency of HsfA2 intron 2 in tomato accessions upon MHS treatment, based on Pearson correlation indicated by the *r*-value and significance by *P*-value. (e) Relative transcript abundances of *Hsp17.7A-CI* and *Hsp70-5* determined by quantitative RT-PCR in seedlings exposed to MHS. M82 is used as a control sample for normalization (one-fold). Significant differences in (c) and (e) are depicted by asterisks (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$) based on *t*-test. Box plots show the median and the upper quartiles. (f) Influence of polymorphisms on intron splicing by minigene assay using *Solanum lycopersicum* cv Moneymaker (SolycA2) or *Solanum peruvianum* (SoperA2). Intron splicing was determined by RT-PCR. Arrowheads show the annealing position of primers on the minigene. The letters denote the wild-type and mutated minigene versions. SLL, *Solanum lycopersicum* var. *lycopersicum*; SLC, *Solanum lycopersicum* var. *cerasiforme*; PIM, *Solanum pimpinellifolium*; WILD, wild accession.

in accessions with a HsfA2^{GGG} haplotype was observed for 50% of the transcripts (Fig. 5c). Instead PIM and SLC accessions with the HsfA2^{GAG} allele show an intermediate and variable splicing profile (Fig. 5c). We observed a correlation between splicing efficiency and thermotolerance ($r = 0.766$) among the 33 accessions, linking HsfA2 splicing to the protection capacity generated during the MHS treatment (Fig. 5d). The transcript abundances of *Hsp17.7A-CI* and *Hsp70-5* are significantly greater in seedlings of accessions carrying HsfA2^{AAA} vs HsfA2^{GGG} alleles in response to MHS, while *Hsp70-5* is significantly more abundant in HsfA2^{GAG} than in HsfA2^{GGG} haplotypes (Fig. 5e). We confirmed the presence of HsfA2-I and HsfA2-II in representative accessions having GGG or AAA alleles, with the latter showing higher levels of HsfA2-II, but also Hsp17-CI (Fig. S7a).

The contribution of each of the three nucleotides to HsfA2 alternative splicing was evaluated in a minigene splicing assay using intron 2 flanked by parts of exon 2 and 3 of *S. lycopersicum* cv Moneymaker (GFP-SolycA2^{GGG}) or *S. peruvianum* HsfA2 (GFP-SoperA2^{AAA}). Exchange of a single or multiple G to A enhances splicing, while A to G mutations promote intron retention in protoplasts transformed with these minigenes after exposure to 37.5°C (Figs 5f, S7b). In agreement with the splicing profile of HsfA2 in HsfA2^{GAG} haplotypes, GFP-SoperA2^{GAG} minigene resulted in an intermediate splicing efficiency between GFP-SolycA2^{GGG} and GFP-SoperA2^{AAA} (Figs 5f, S7b). LD analysis for the three SNPs with all polymorphisms in the HsfA2 gene reveal that only the first and third SNPs are in LD, but also that both are in LD with two downstream intron SNPs (Fig. S8). In addition the third SNP is in LD with a polymorphism in the mini-exon. However, the results of the minigene splicing assay show that these additional SNPs do not affect the splicing profile of HsfA2, as splicing occurs independently of the mini-exon but also irrespective of the additional SNPs in intron 2 (Fig. 5f). Thus, the three intronic nucleotides are important for the regulation of intron 2 splicing efficiency.

The evolution of the three HsfA2 intron 2 haplotypes was examined in a set of 601 sequenced tomato accessions (Zhu *et al.*, 2018) (Fig. S9a; Table S3). Whereas all WILD accessions possess the HsfA2^{AAA} haplotype, all but three accessions of modern and vintage tomatoes (SLL) carry the HsfA2^{GGG} haplotype (333 accessions). Further analysis showed that the three SLL accessions that did not carry HsfA2^{GGG} alleles have an introgression from a wild species in the HsfA2 region (Fig. S9b). The frequency of the HsfA2^{GGG} haplotype decreases with phylogenetic distance with respect to cultivated tomato, being less abundant in more ancestral SLC and very rare in PIM (Fig. 6a). The increase in frequency of the HsfA2^{GGG} in SLC in comparison to PIM could be the result of a gradual selection starting during the migration of tomato out of the Andean region and a strong fixation during domestication. Supporting this notion, we found evidence of positive selection on the region of HsfA2 when comparing SLL and SLC, but not when comparing SLC and PIM (Fig. 6b; Table S5).

The HsfA2^{GGG} haplotype seems to be specific for SLL and *S. galapagense* among Solanaceae species, as other nontomato species, including *Solanum tuberosum* (potato) and *Capsicum anuum* (pepper), possess A in the first two SNP positions and show enhanced splicing after 37.5°C treatment (Fig. S10a,b). Similar to SLL cv Moneymaker, splicing is inhibited after 42.5°C treatment in all tested Solanaceae species, suggesting that the polymorphisms are not relevant for splicing efficiency under severe HS where a general inhibition of splicing is assumed (Fig. S10a). A similar regulatory mechanism by alternative splicing has not been reported for any plant Hsf. However, we observed the presence of multiple HsfA2 orthologs in other plant species and particularly in several monocots from which at least one lacks an NES (Table S6).

Tomato HsfA2-I is involved in thermotolerance acquisition upon repeated cycles of HS as the protein accumulates during the recovery phase from a mild HS treatment and is maintained in the cell until utilization in case of an upcoming challenging stress (Scharf *et al.*, 1998; Hahn *et al.*, 2011). We assumed that selection for HsfA2^{GGG} allele during domestication ensured the accumulation of HsfA2-I to confer short-term acclimation and

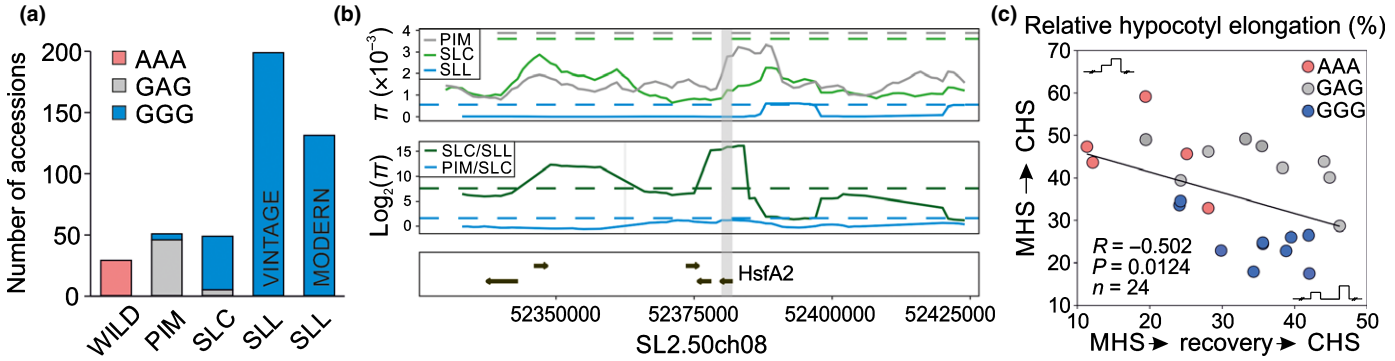

Fig. 6 Selection of HsfA2^{GGG} haplotype during tomato domestication. (a) Number of accessions possessing different haplotypes classified by HsfA2 allele and phylogenetic group. WILD, wild accession; PIM, *Solanum pimpinellifolium*; SLL, *Solanum lycopersicum* var. *lycopersicum*; SLC, *Solanum lycopersicum* var. *cerasiforme*. (b) Ratio of nucleotide diversity (π) values ($\log_2$) between different groups and SLL and the corresponding position of HsfA2 in the 13 Mb region (upper panel). π was calculated for each group of accessions in 10 kb windows and 1 kb steps. Thresholds shown as dotted lines represent the 95% quantile for a 13 Mb region centered in HsfA2. Gene positions are depicted by arrows in the lower panel. (c) Pearson correlation between thermotolerance of seedlings exposed to 37.5°C (mild heat stress (MHS)) and then shifted directly to 50°C (challenging HS (CHS)), or with an interval of 3 h recovery at 25°C between the two stress regimes.

probably stress memory to repeated cycles of HS (Lämke *et al.*, 2016). The acclimation capacity of 24 accessions was examined by introducing a 3 h recovery interval at 25°C between the pretreatment and the challenging stress (Fig. 6c). The release of HsfA2-I from HSGs and its accumulation in soluble but inactive complexes with small heat shock proteins (sHsps) during recovery from the pretreatment allow a faster activation in case of a subsequent HS (Scharf *et al.*, 1998). Indeed, there is a significant negative correlation ($R = -0.502$, $P = 0.0124$) between the rapid and gradual acclimation capacity of seedlings, which supports the notion that cellular fitness under these two regimes is dependent on the relative abundance of the two protein isoforms (Fig. 6c).

Discussion

Acclimation is an evolutionarily important feature of all biological systems that allows adaptation and survival of organisms under a variety of stressful conditions (Hong & Vierling, 2000). Cultivated tomatoes exhibit a weaker HSR in response to MHS treatment, and consequently show reduced thermotolerance in the case of a subsequent direct CHS as compared with wild species (Figs 1a, 5e). Polymorphisms in intron 2 that affect HsfA2 pre-mRNA splicing during MHS are associated with the variation in thermotolerance and, consequently, the ratio of two protein isoforms with distinct functions in acclimation of tomato to HS (Figs 5, 6). Intron 2 is present in several Solanaceae but is rare in other species. So far only in *Populus* has a second intron in HsfA2 been reported, suggesting that similar regulation by alternative splicing might exist in other species as well (Zhang *et al.*, 2015).

HsfA2-I comprises two AHA motifs and a C-terminal NES (Fig. 4) and contributes to acquired thermotolerance (Fig. 6). This isoform requires interaction with HsfA1a for efficient nuclear retention, and gains high transactivation activity in complex with HsfA1a (Chan-Schaminet *et al.*, 2009). Under HS the majority of HsfA2-I protein is recruited into HSGs, while during recovery periods the factor is released and stored in soluble complexes with cytosolic sHsps and kept inactive until repeated HS periods occur (Port *et al.*, 2004; Fig. 4). Both interaction with HsfA1a and sHsps enforce the long-term stability of HsfA2-I as a prerequisite for its contribution in acquired thermotolerance (Scharf *et al.*, 1998; Chan-Schaminet *et al.*, 2009) (Fig. 6).

The second isoform HsfA2-II is central for the direct response to HS conditions (Figs 4, 5). It lacks an NES and has a truncated C-terminal AHA motif (Fig. 4). The C-terminal region of HsfA2 generally has a tripartite role: control of the nucleocytosolic equilibrium; interaction with cytosolic sHsps of class CI and CII; and interaction with general transcription factors via the AHA motifs (Scharf *et al.*, 1998; Port *et al.*, 2004; Chan-Schaminet *et al.*, 2009). Owing to the C-terminal truncation, HsfA2-II escapes from the repressor effect of sHsps and is retained in the nucleus (Fig. 4). The latter is accompanied by an increased turnover of HsfA2-II which, together with its accumulation in a more narrow temperature range than HsfA2-I, points to a transient activity of this isoform (Fig. 4). Current models on regulation of Hsfs have focused primarily on transcriptional control and protein interaction with chaperones and co-chaperones (Scharf *et al.*, 1998; Meiri & Breiman, 2009; Hahn *et al.*, 2011; Ohama *et al.*, 2015). Furthermore, alternative splicing of the conserved intron spanning the DBD encoding region of all plant Hsfs is related to NMD and leads to reduced levels of mature protein-coding Hsf transcripts (34; Figs 4, S4). Our results underline the importance of functional diversification of plant Hsfs by alternative splicing, adding another level of Hsf regulation.

Remarkably, the observed differences in thermotolerance between tomato genotypes correlate with variation in relative abundance of the two HsfA2 isoforms. Wild tomato accessions that produce HsfA2-II at the expense of HsfA2-I show increased capacity to rapidly acclimate against severe HS as a result of their ability to stimulate a stronger HSR as compared with modern cultivars (Figs 1, 2). Instead, a shift towards intron retention in modern SLL accessions as a result of the selection of $HsfA2^{GGG}$ haplotypes during domestication favoured the synthesis of HsfA2-I under MHS conditions. The latter probably led to the loss of rapid acclimation capacity (Figs 1–3, 5, 6). In turn, these plants show a better response and enhanced short-term acclimation to repeated HS cycles (Fig. 6). A similar tradeoff has been proposed for *japonica* and *indica*, two rice cultivars with opposite performance in response to basal and acquired thermotolerance (M-Y. Lin *et al.*, 2014). Differences in thermotolerance against different stress scenarios might be related to local adaptation, but further studies are required for a detailed understanding of this.

The molecular mechanism established by the two HsfA2 isoforms in Solanaceae probably parallels the one in monocots, with the distinction that monocots carry multiple HsfA2 orthologs, of which at least one lacks an NES. This suggests that functional diversification conferred by alternative splicing in Solanaceae HsfA2 is substituted by gene duplication in monocots. This notion is consistent with the loss of introns in Hsf coding genes from aquatic to lower land plants and eventually to higher plants (Wang *et al.*, 2013). In consequence, fine regulation of multiple HsfA2 orthologs is expected to be achieved by control of gene expression. Interestingly and in support of our observation for tomato HsfA2, in rice the accumulation of the NES lacking *OsHsfA2e* transcripts is only transiently induced by HS, while transcripts of the NES containing *OsHsfA2a* and *OsHsfA2b* sustain high levels in response to long-term HS treatments (Yokotani *et al.*, 2008; Wang *et al.*, 2009).

At this stage, we cannot exclude the possibility that the selection of the $HsfA2^{GGG}$ allele is a result of LD with another quantitative trait locus. For example, the wild malodorous locus related to the undesired aroma as a result of phenylacetaldehyde synthesis is also located in chromosome 8 and was selected against during tomato domestication (Tadmor *et al.*, 2002). However, no major quantitative trait loci have been reported in chromosome 8 within the proximity of the HsfA2 region, to the best of our knowledge. In conclusion, we report alternative splicing of HsfA2 as a new regulatory mechanism that contributes to plant acclimation capacity under different HS conditions. Future attempts will focus on agronomically relevant traits to decipher the contribution of HsfA2 to important attributes of tomato farming under various climate scenarios, in order to unravel the advantage conferred by each haplotype.

Acknowledgements

We acknowledge Lutz Nover, Elizabeth Vierling, Christos Bazakos and Miltos Tsiantis for critical reading and suggestions. This work was supported by grants from DFG Germany to SF (413737959) and to ES (SFB902), a SPOT-ITN Marie Curie grant (289220) to ES, K-DS and AB, and a China Scholarship Council (CSC) to YH.

Author contributions

YH carried out the majority of the experiments, AM generated the mutants, PG performed thermotolerance and SR and DB conducted the activity assays. JMJG performed the GWAS and population genetics analyses. AB carried out the selection of ILs. SF conceptualized the project and, with K-DS and ES, headed it up. SF, JMJG, K-DS and ES wrote the manuscript. All authors have read and approved the manuscript.

ORCID

Sotirios Fragkostefanakis https://orcid.org/0000-0001-5311-7868
Philipp Gebhardt https://orcid.org/0000-0002-8539-3469
Yangjie Hu https://orcid.org/0000-0002-0910-6384
José M. Jiménez-Gómez https://orcid.org/0000-0002-5033-7192
Anida Mesihovic https://orcid.org/0000-0001-6277-5141
Sascha Röth https://orcid.org/0000-0001-7273-6701
Enrico Schleiff https://orcid.org/0000-0002-0518-3489

References

Abràmoff MD, Magalhães PJ, Ram SJ. 2004. Image processing with imageJ. *Biophotonics International* **11**: 36–42.

Aflitos S, Schijlen E, De Jong H, De Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L *et al.* 2014. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *The Plant Journal* **80**: 136–148.

Berz J, Simm S, Schuster S, Scharf K, Schleiff E, Ebersberger I. 2019. HEATSTER: a database and web server for identification and classification of heat stress transcription factors in plants. *Bioinformatics and Biology Insights* **13**: doi: 10.1177/1177932218821365.

Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, van der Knaap E, Cañizares J. 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics* **16**: 257.

Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sørensen I, Lichtenstein G *et al.* 2014. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature Genetics* **46**: 1034–1038.

Carter MS, Doskow J, Morris P, Li SL, Nhim RP, Sandstedt S, Wilkinson MF. 1995. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts *in-vivo* is reversed by protein-synthesis inhibitors *in-vitro*. *Journal of Biological Chemistry* **270**: 28995–29003.

Chan-Schaminet KY, Baniwal SK, Bublak D, Nover L, Scharf K-D. 2009. Specific interaction between tomato HsfA1 and HsfA2 creates hetero-oligomeric superactivator complexes for synergistic activation of heat stress gene expression. *The Journal of Biological Chemistry* **284**: 20848–20857.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**: 7.

Charng Y-Y, Liu H-C, Liu N-Y, Chi W-T, Wang C-N, Chang S-H, Wang T-T. 2006. A heat-inducible transcription factor, Hsf A2, is required for extension of acquired thermotolerance in Arabidopsis. *Plant Physiology* **143**: 251–262.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491.

Easterling DR, Evans JL, Groisman PY, Karl TR, Kunkel KE, Ambenje P. 2000. Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society* **81**: 417–426.

Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JDG, Patron NJ, Marillonnet S. 2014. A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology* **3**: 839–843.

Eshed Y, Zamir D. 1995. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**: 1147–1162.

Fragkostefanakis S, Mesihovic A, Simm S, Paupière MJ, Hu Y, Paul P, Mishra SK, Tschiersch B, Theres K, Bovy A *et al.* 2016. HsfA2 controls the activity of developmentally and stress-regulated heat stress protection mechanisms in tomato male reproductive tissues. *Plant Physiology* **170**: 2461–2477.

Fragkostefanakis S, Simm S, El-Shershaby A, Hu Y, Bublak D, Mesihovic A, Darm K, Mishra SK, Tschiersch B, Theres K *et al.* 2018. The repressor and co-activator HsfB1 regulates the major heat stress transcription factors in tomato. *Plant, Cell & Environment* **42**: 874–890.

Fragkostefanakis S, Simm S, Paul P, Bublak D, Scharf K-D, Schleiff E. 2014. Chaperone network composition in *Solanum lycopersicum* explored by transcriptome profiling and microarray meta-analysis. *Plant, Cell & Environment* **38**: 693–709.

Frank G, Pressman E, Ophir R, Althan L, Shaked R, Freedman M, Shen S, Firon N. 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. *Journal of Experimental Botany* **60**: 3891–3908.

Hahn A, Bublak D, Schleiff E, Scharf K-D. 2011. Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. *Plant Cell* **23**: 741–755.

Hajjar R, Hodgkin T. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* **156**: 1–13.

Hong SW, Vierling E. 2000. Mutants of *Arabidopsis thaliana* defective in the acquisition of tolerance to high temperature stress. *Proceedings of the National Academy of Sciences, USA* **97**: 4392–4397.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**: 348.

Kotak S, Port M, Ganguli A, Bicker F, Von Koskull-Döring P. 2004. Characterization of C-terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class a Hsfs with AHA and NES motifs essential for activator function and intracellular localization. *The Plant Journal* **39**: 98–112.

Kudo N, Wolff B, Sekimoto T, Schreiner EP, Yoneda Y, Yanagida M, Horinouchi S, Yoshida M. 1998. Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1. *Experimental Cell Research* **242**: 540–547.

Lämke J, Brzezinka K, Altmann S, Bäurle I. 2016. A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO Journal* **35**: 162–175.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357.

Larkindale J, Hall JD, Knight MR, Vierling E. 2005. Heat stress phenotypes of Arabidopsis mutants implicate multiple signaling pathways in the acquisition of thermotolerance. *Plant Physiology* **138**: 882–897.

Larkindale J, Vierling E. 2007. Core genome responses involved in acclimation to high temperature. *Plant Physiology* **146**: 748–761.

Li XM, Chao DY, Wu Y, Huang X, Chen K, Cui LG, Su L, Ye WW, Chen H, Chen HC *et al.* 2015. Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of African rice. *Nature Genetics* **47**: 827–833.

Lin M-Y, Chai K-H, Ko S-S, Kuang L-Y, Lur H-S, Charng Y-Y. 2014. A Positive Feedback loop between HEAT SHOCK PROTEIN101 and HEAT STRESS-ASSOCIATED 32-KD PROTEIN modulates long-term acquired thermotolerance illustrating diverse heat stress responses in rice varieties. *Plant Physiology* **164**: 2045–2053.

Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X *et al.* 2014. Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics* **46**: 1220–1226.

Liu H, Naismith JH. 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology* **8**: 91.

Liu J, Sun N, Liu M, Liu J, Du B, Wang X, Qi X. 2013. An autoregulatory loop controlling Arabidopsis HsfA2 expression: role of heat shock-induced alternative splicing. *Plant Physiology* **162**: 512–521.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* **25**: 402–408.

Lobell DB, Field CB. 2007. Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters* **2**: 014002.

McCormick S, Niedermeyer J, Fry J, Barnason A, Horsch R, Fraley R. 1986. Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*. *Plant Cell Reports* **5**: 81–84.

Meiri D, Breiman A. 2009. Arabidopsis ROF1 (FKBP62) modulates thermotolerance by interacting with HSP90.1 and affecting the accumulation of HsfA2-regulated sHSPs. *The Plant Journal* **59**: 387–399.

Mishra SK, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, Scharf K-D. 2002. In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. *Genes & Development* **16**: 1555–1567.

Neumann D, Nover L, Parthier B, Rieger R, Scharf K, Wollgiehn R, Nieden U.1989. *Heat shock and other stress response systems of plants* (Nover L, Neumann D, Scharf K-D, eds.). Berlin, Germany: Springer-Verlag.

Ohama N, Kusakabe K, Mizoi J, Zhao H, Kidokoro S, Koizumi S, Takahashi F, Ishida T, Yanagisawa S, Shinozaki K *et al.* 2015. The transcriptional cascade in the heat stress response of *Arabidopsis* is strictly regulated at the expression levels of transcription factors. *Plant Cell* **28**: 181–201.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

Parsell DA, Lindquist S. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual Review of Genetics* **27**: 437–496.

Port M, Tripp J, Zielinski D, Weber C, Heerklotz D, Winkelhaus S, Bublak D, Scharf K-D. 2004. Role of Hsp17.4-CII as coregulator and cytoplasmic retention factor of tomato heat stress transcription factor HsfA2. *Plant Physiology* **135**: 1457–1470.

Queitsch C, Hong SW, Vierling E, Lindquist S. 2000. Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis. *Plant Cell* **12**: 479–492.

Rahmstorf S, Coumou D. 2011. Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences, USA* **108**: 17905–17909.

Röth S, Mirus O, Bublak D, Klaus-Dieter S, Schleiff E. 2016. DNA-binding and repressor function are prerequisite for the turnover of the tomato heat stress transcription factor HsfB1. *The Plant Journal* **42**: 1–14.

Scharf KD, Berberich T, Ebersberger I, Nover L. 2012. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochim Biophys Acta* **1819**: 104–119.

Scharf KD, Heider H, Höhfeld I, Lyck R, Schmidt E, Nover L, Hohfeld I, Lyck R, Schmidt E, Nover L. 1998. The tomato Hsf system: HsfA2 needs interaction with HsfA1 for efficient nuclear import and may be localized in cytoplasmic heat stress granules. *Molecular and Cellular Biology* **18**: 2240–2251.

Schramm F, Ganguli A, Kiehlmann E, Englich G, Walch D, Von Koskull-Döring P. 2006. The heat stress transcription factor HsfA2 serves as a regulatory amplifier of a subset of genes in the heat stress response in Arabidopsis. *Plant Molecular Biology* **60**: 759–772.

Tadmor Y, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E. 2002. Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. *Journal of Agricultural and Food Chemistry.* **50**: 2005–2009.

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

Treuter E, Nover L, Ohme K, Scharf KD. 1993. Promoter specificity and deletion analysis of three heat stress transcription factors of tomato. *MGG Molecular & General Genetics* **240**: 113–125.

Wahid A, Gelani S, Ashraf M, Foolad M. 2007. Heat tolerance in plants: an overview. *Environmental and Experimental Botany* **61**: 199–223.

Wang Y, You FM, Lazo GR, Luo MC, Thilmony R, Gordon S, Kianian SF, Gu YQ. 2013. PIECE: a database for plant gene structure comparison and evolution. *Nucleic Acids Research* **41**: 1159–1166.

Wang C, Zhang Q, Shou H. 2009. Identification and expression analysis of OsHsfs in rice. *Journal of Zhejiang University Science B* **10**: 291–300.

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6**, e16765.

Wu T-Y, Juan Y-T, Hsu Y-H, Wu S-H, Liao H-T, Fung RWM, Charng Y-Y. 2013. Interplay between heat shock proteins HSP101 and HSA32 prolongs heat acclimation memory posttranscriptionally in Arabidopsis. *Plant Physiology* **161**: 2075–2084.

Xie K, Zhang J, Yang Y. 2014. Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Molecular Plant* **7**: 923–926.

Yeh CH, Kaplinsky NJ, Hu C, Charng YY. 2012. Some like it hot, some like it warm: phenotyping to explore thermotolerance diversity. *Plant Science* **195**: 10–23.

Yokotani N, Ichikawa T, Kondou Y, Matsui M, Hirochika H, Iwabuchi M, Oda K. 2008. Expression of rice heat stress transcription factor OsHsfA2e enhances tolerance to environmental stresses in transgenic Arabidopsis. *Planta* **227**: 957–967.

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**: 28–36.

Zhang J, Liu B, Li J, Zhang L, Wang Y, Zheng H, Lu M, Chen J. 2015. Hsf and Hsp gene families in *Populus*: genome-wide identification, organization and correlated expression during development and in stress responses. *BMC Genomics* **16**: 1–19.

Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, Huang M, Yao Y, Bassu S, Ciais P *et al.* 2017. Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences, USA* **114**: 9326–9331.

Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C *et al.* 2018. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**: 249–261.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Hypotocyl length of tomato seedlings at 2 d after the stress treatment.

Fig. S2 Splicing of HsfA2 intron 2 in *S. lycoperscicum* cv M82, *S. pennellii* LA716 and the introgression line IL8-1.

Fig. S3 Transactivation activity of domesticated and wild HsfA2.

Fig. S4 Alternative splicing in intron 1 of HsfA2 leads to premature termination codon and targeting of HsfA2-III transcripts for nonsense mRNA decay.

Fig. S5 Localization of GFP-HsfA2-I and GFP-HsfA2-II in tomato mesophyll protoplasts isolated from transgenic A1CS plants.

Fig. S6 Transcriptional activation activity of HsfA2 isoforms.

Fig. S7 Splicing of HsfA2 and contribution of SNPs to splicing efficiency of HsfA2 minigenes.

Fig. S8 Linkage disequilibrium of HsfA2 polymorphisms.

Fig. S9 Phylogenetic tree of tomato and frequency of nonreference SNPs in vintage tomatoes in the region of HsfA2.

Fig. S10 Alternative splicing of HsfA2 in Solanaceae species.

Table S1 Accessions used in thermotolerance assays.

Table S2 Primers used in this study.

Table S3 Accessions classified by HsfA2 allele and phylogenetic group.

Table S4 Genes identified by GWAS as being associated with thermotolerance, including flanking genes in a 20 kb radius.

Table S5 Calculation of nucleotide diversity (π).

Table S6 HsfA2 orthologs in plant species.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.