HEATSTER: A Database and Web Server for Identification and Classification of Heat Stress Transcription Factors in Plants

Bioinformatics and Biology Insights
Volume 13: 1–5

DOI: 10.1177/1177932218821365
SAGE

Jannik Berz[1], Stefan Simm[1,2], Sebastian Schuster[3], Klaus-Dieter Scharf[1], Enrico Schleiff[1,2] and Ingo Ebersberger[4,5,6]

[1]Department of Biosciences, Molecular Cell Biology of Plants, Goethe University, Frankfurt am Main, Germany. [2]Frankfurt Institute of Advanced Studies, Department of Life Sciences, Frankfurt, Germany. [3]Center for Integrative Bioinformatics Vienna (CIBIV), Max F. Perutz Laboratories, Vienna, Austria. [4]Department of Biosciences, Inst. of Cell Biology and Neuroscience, Applied Bioinformatics Group, Goethe University, Frankfurt am Main, Germany. [5]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany. [6]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany.

ABSTRACT: Heat stress transcription factors (HSFs) regulate transcriptional response to a large number of environmental influences, such as temperature fluctuations and chemical compound applications. Plant HSFs represent a large and diverse gene family. The HSF members vary substantially both in gene expression patterns and molecular functions. HEATSTER is a web resource for mining, annotating, and analyzing members of the different classes of HSFs in plants. A web-interface allows the identification and class assignment of HSFs, intuitive searches in the database and visualization of conserved motifs, and domains to classify novel HSFs.

KEYWORDS: HSF, motif search, database, heat stress

RECEIVED: November 15, 2018. **ACCEPTED:** November 23, 2018.

TYPE: Technical Advances

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was supported by grants of the Deutsche Forschungsgemeinschaft (Schleiff: CRC-SFB902 B9) and from the EU/Marie Curie (project: SPOT-ITN, grant agreement No. 289220, and CALIPSO-ITN, grant agreement No. GA ITN-2013 607 607).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Ingo Ebersberger, Department of Biosciences, Inst. of Cell Biology and Neuroscience, Applied Bioinformatics Group, Goethe University, Biologicum Max-von-Laue Straße 13, D-60438 Frankfurt am Main, Germany. Email: ebersberger@bio.uni-frankfurt.de

Stefan Simm, Department of Biosciences, Molecular Cell Biology of Plants, Goethe University, Biocenter N200, Max-von-Laue Straße 9, D-60438 Frankfurt am Main, Germany. Email: simm@bio.uni-frankfurt.de

Introduction

Plants have evolved a remarkable complexity in their stress response. The transcriptional reprogramming at higher temperatures is controlled by heat stress transcription factors (HSFs) leading to the activation of genes involved in heat stress response (HSR).[1,2] In general, HSFs control the expression of genes responsive to numerous abiotic stresses (eg, heat, drought, and salinity), while recently a function in the developmental regulation was observed as well.[3–6] The abundance and function of HSFs is controlled by various mechanisms like protein degradation, cooperative interactions between distinct HSF members, and the interaction with chaperones.[7]

The HSF gene family comprises between 15 and 50 members depending on the plant species.[8] All HSFs share the presence of 2 conserved functional domains, the N-terminal DNA-binding domain (DBD), containing a helix-turn-helix motif flanked by 2 β-strands on each side, and 2 heptad repeat patterns (HR-A/B) of hydrophobic amino acids (aa) building the oligomerization domain (OD). Based on the length of the insertion in the linker sequence between the 2 HR patterns, HSFs have been differentiated into to class A (21 aa), B (0 aa), and C (7 aa). Furthermore, differences in the primary and secondary structure of the 2 conserved domains have been used for classifying plant HSFs (Figure 1).[8–10]

Each HSF class is further distinguished into sub-classes, eg, HsfA1, based on the characteristic architecture of the functional motifs. These functional motifs regulate the DNA binding (DBD), the oligomerization (HR-A, HR-B), and the intracellular localization (nuclear localization signal/sequence [NLS]; nuclear export signal/sequence [NES]). With respect to the latter, only class A, but not class B and C HSFs harbor a NES. In addition, aromatic, hydrophobic, and acidic sequence stretches can act as activator domains (AHA) or repressor domains (RD). These domains fine-tune the functionality of the individual transcription factors.[10] The RD is generally associated with class B HSFs, while AHA motifs are typically found in class A HSFs. In some cases, a sub-class of HSFs can comprise of up to 5 different factors, annotated by additional letters (eg, HsfA1a).

Overall, HSFs of different classes and sub-classes establish a complex network that controls a fine-tuned program of stress response. Functional analyses in model plants, eg, *Solanum lycopersicum, Oryza sativa*, or *Arabidopsis thaliana*, have greatly contributed to elucidating the function of this regulatory network.[6,11] The information about the function of the individual HSFs in maintenance of homeostasis and recovery after stress cycles serves as basis to investigate the HSR in plants, with a particular focus on crops.[5] Therefore, a comprehensive sequence and structure based mining, annotation, and analysis of plant HSFs in different species will provide a deep understanding of their role in plant abiotic stress responses with a strong focus on HSR, which should lead to a decrease in crop losses worldwide.[4]

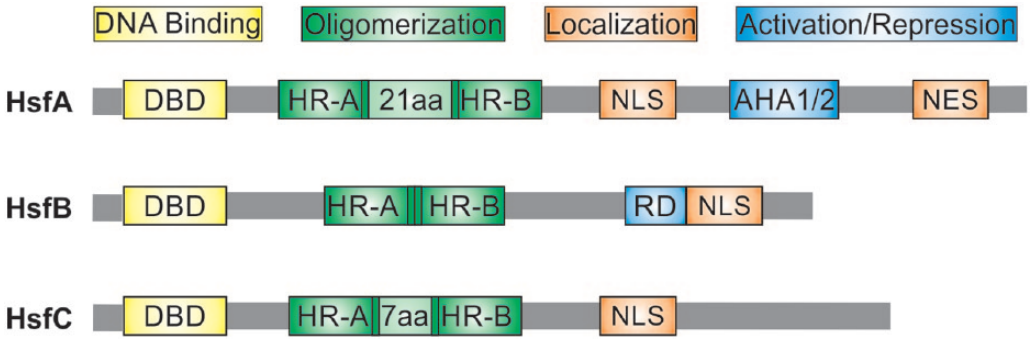

Figure 1. Domain architecture of the 3 major HSF classes in plants. Gray lines represent regions of low conservation, variable length, and without annotated motifs. The domain architecture comprises the conserved DNA-binding (yellow) and oligomerization domains (HR-A/B region, green), the NLS and NES (orange), and the transcriptional activator and repressor domains (blue). AHA indicates activator motifs/domains; DBD, DNA-binding domain; HR-A/B, heptad repeat patterns; HSF, heat stress transcription factors; NES, nuclear export signal/sequence; NLS, nuclear localization signal/sequence; RD, repressor motifs/domains.
Adapted from Scharf et al[10]

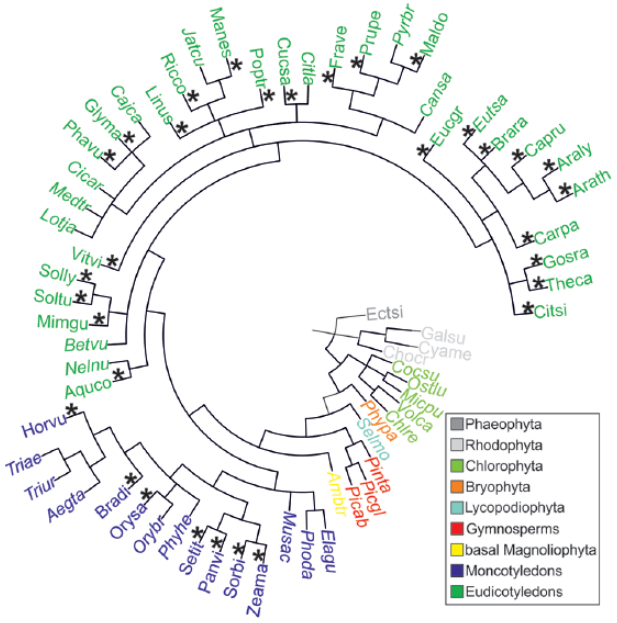

Figure 2. Species included in the HEATSTER databases. The taxonomic tree displays the species collection that is currently included in HEATSTER. The color code corresponds to the taxonomic assignment of the individual species. Species names are abbreviated (first 3 letters of the genus and first 2 letters of the species epithet). Supplementary Table 1 links the short name to the full name of the species. The tree is rooted using the Phaeophyta (dark gray) and Rhodophyta (light gray) as outgroups. Species represented in the manually curated v1.0 of HEATSTER are marked with an asterisk.

HEATSTER is a web-based reference platform for integrating HSF research across plants. The underlying database comprises 848 manually curated HSFs from 32 plant species version 1.0 (v1.0).[10] In version 2.0 (v2.0), this data set is complemented by further 1000 mostly automatically annotated HSFs from additionally 29 plant species of different ranks, 1 Phaeophyta and 3 Rhodophyta. Complementary to the data repository, HEATSTER provides a rich environment for annotation and classification of HSFs in new species. Furthermore, HEATSTER facilitates the analysis of HSFs in a functional and evolutionary context.

Materials and Methods

Deposited data

Full length amino acid (v2.0 and v1.0) and coding sequence (CDS) nucleotide sequences (v1.0) of plant HSFs are deposited in the HEATSTER database. The v1.0 from January 2014 includes 32 manually curated angiosperm species with 26 Eudicotyledons and 6 Monocotyledons. HEATSTER v2.0 from September 2016 extends the database to 65 species including 1 Phaeophyta, 3 Rhodophyta, and 61 species of Viridiplantae (5 Chlorophyta, 1 Bryophyta, 1 Lycopodiidae, 1 basal Magnoliophyta, 3 Gymnosperms, 15 Monocotyledons, and 35 Eudicotyledons) (Figure 2; Supplemental Table 1).

HEATSTER v1.0 features a collection of manually curated HSFs from plant genomes available prior to January 2014. The curation step served to correct sequencing errors and wrong gene models resulting from an automated gene annotation procedure. The curation procedure included the following analysis steps: (1) editing based on homology comparison to model organisms and (2) scanning for conserved signature sequence motifs within the predicted genomic region, as well as in the adjacent 5' and 3' intergenic regions. Although we are confident that the curation procedure removed most annotation errors, an ultimate validation must await experimental evidence. All sequences of the HEATSTER v2.0 are directly extracted from the databases Phytozome, NCBI, Dendrome, Bambogdb, Banana hub, Kazusa Database, and Cucurbit Genomics Database (Supplementary Table 1; September 2016).

Signature motif libraries

The HEATSTER database provides 2 sets of signature motifs for the HSF sub-classes, the first for the manual sequences of v1.0 and the second set for the automatically annotated HSFs of

v2.0. For the creation of the signature motif libraries, training sets of HSF sequences for each sub-class were defined. To compile the training data for the signature motifs of v1.0, the assignment of HSF sub-class sequences was performed in 3 stages using sequences from 32 angiosperm species: (1) Homolog search with known HSF sequences in publicly available EST, cDNA, and protein databases; (2) Refinement of identified HSF sequences by BLAST search in plant genome databases; (3) Classification of newly identified HSF sequences based on conserved functional and signature sequence motifs according to the widely accepted nomenclature for plant HSFs.[8,9]

The detected signature motif library of v1.0 based on the nomenclature of Nover et al[8,9] was used as starting point to classify HSF sub-classes in all 61 plant species of v2.0. Conserved signature motifs within the predicted HSF sequences were performed by MEME, TOMTOM, and MAST.[12,13] The HSF sub-classes were used to perform a motif search via MEME using different parameter sets for motif-width (8-20 aa, 20-35 aa, 35-50 aa) and site coverage (-OOPS, -ZOOPS, -Anr). The option maxiter was set to 100 and nmotifs was set to 20. For each HSF sub-class, we created a decoy database containing the random shuffled sequences of the HSF sub-class and performed the motif search 10 times (×9 shuffling, ×1 original) for each parameter setting. The identification of the same signature motif in the decoy database was counted as false positive (FP) and used to calculate a false discovery rate (FDR). All signature motifs below a threshold of 0.3 were selected as signature motifs for HSF sub-classes. Furthermore, signature motifs with an identity above 95% (identified by CD-hit) were merged via TOMTOM. Therefore, only signature motifs of the single HSF sub-classes were cross-validated by searching the HSF sub-class signature motifs in the other HSF sub-class signature motif libraries with MAST and TOMTOM to detect sub-class specific signature motifs.

Results

HEATSTER platform

The website HEATSTER is free and open to all users without a login requirement. HEATSTER is written in PHP, HTML, and CCS and uses HMMscan from HMMER (see http://hmmer.org/) and MAST from the MEME suite (see http://meme-suite.org/doc/mast.html) for the classification and identification. Furthermore, MySQL (see https://www.mysql.com/de/), jQuery (see https://jquery.com/), sorttable (see https://www.kryogenix.org/code/browser/sorttable/), and CSS Bootstrap (see https://getbootstrap.com/docs/3.3/css/) are included in the website. The HEATSTER website is located at http://applbio.biologie.uni-frankfurt.de/hsf/heatster/. The website is divided into a HSF classification and visualization tool.

Beside classification and visualization, HEATSTER provides downloadable content like logo plots from the HSF motifs and FASTA-files of HSF sequences in the download section. The underlying MySQL database provides comprehensive information about curated HSFs from 26 Eudicotyledons and 6 Monocotyledons of v1.0.[10] HEATSTER v2.0 provides information for 5 Chlorophyta, 1 Bryophyta, 1 Lycopodiidae, 1 basal Magnoliophyta, 3 Gymnosperms, 15 Monocotyledons, and 35 Eudicotyledons (Figure 2; Supplementary Table 1). Furthermore, 3 Rhodophyta and 1 Phaeophyta were included. HEATSTER features the nomenclature of plant HSFs suggested by Nover et al.[8,9] All sequences together with their annotation can be accessed via the web-interface.

Annotation and visualization tool

The sequence analysis routines of the HEATSTER classification tool facilitate the online annotation and classification of novel HSF candidates. A library of signature motifs characterizing the individual HSF classes and sub-classes forms the fundament of these analyses. To generate this library in v2.0, we first compiled HSF sub-classes containing 61 plant species based on the v1.0 nomenclature and used MEME to identify and validate shared motif sets in the individual sub-classes. Signature motifs with a q-value lower than 1.0e-09 and sequence identity of more than 70% were merged via TOMTOM.[12] To arrive at sub-class specific signature motif sets, we removed those motifs that are represented in more than 1 HSF sub-class. The web-logos of the signature motif library can be accessed online.

For the classification of a HSF candidate, HEATSTER first assigns the candidate to the HSF classes A, B, or C based on the characteristic appearance of the DBD and OD domains. Sequences harboring only one of these domains are called HSF-related. Subsequently, HEATSTER maps the signature motif sets of all HSF sub-classes against the candidate with MAST.[13] The sequence is then assigned to the best matching HSF sub-class. For the HSF classification, HEATSTER requires sequences in FASTA format submitted as a file or pasted in the provided textbox as input. HEATSTER also allows batch searches where the entire gene set of an organism can be screened for the presence of HSFs. The output is represented in table format and a modified MAST visualization output. The tables are downloadable in csv- and the sequences in FASTA format.

The visual representation of HSF sequences and their signature motif architecture in HEATSTER is facilitated by the visualization tool and makes comparative studies on HSFs intuitive and straightforward. For visualization of the HSF class-specific motifs, the input is selected via dropdown menus. In the plain mode, the user can display a set of signature motifs for sequences representing a single or multiple HSF sub-classes. However, it is also possible to paste sequences in FASTA format in the extended mode for a comparison to a set of pre-selected HSFs. In the extended mode, the user can upload a query protein for a comparison to a set of pre-selected HSFs. It also allows, if present, the display of signature motifs characteristic for other sub-classes. Thereby, novel motif combination can be detected that may provide first insights into the specific functions of the HSFs and a more fine-grained

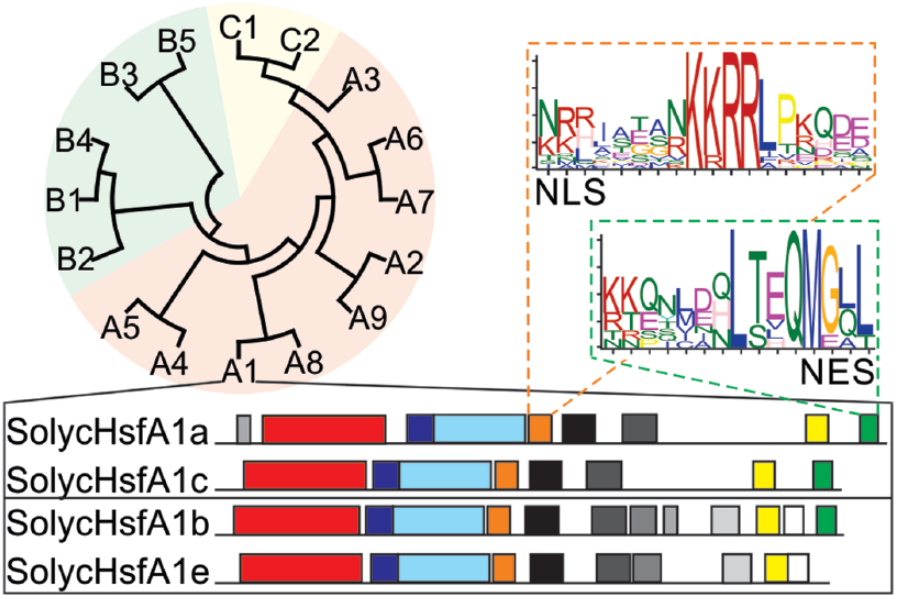

Figure 3. HSF classification in HEATSTER. Candidates are first assigned to the HSF classes A, B, or C, and are subsequently sub-classified based on the signature motif architecture (colored boxes). Logo plots for each motif can be visualized alongside the architecture. Shown are exemplarily NLS and NES. HsfA1a: Solyc08g005170; HsfA1c: Solyc08g076590; HsfA1b: Solyc03g097120; HsfA1e: Solyc08g076590. HSF indicates heat stress transcription factors; NES, nuclear export signal; NLS, nuclear localization signal.

Table 1. HSFs in *Solanum lycopersicum* identified by the HEATSTER.

IDENTIFIER ITAG2.4	V1.0	V2.0	E-VALUE V2.0
Solyc03g097120.2.1	SolycHsfA1b	SollyHsfA1b	4.40E-262
Solyc06g072750.2.1	SolycHsfA1e	SollyHsfA1e	9.70E-256
Solyc08g005170.2.1	SolycHsfA1a	SollyHsfA1a	3.80E-231
Solyc08g076590.2.1	SolycHsfA1c	SollyHsfA1c	1.80E-190
Solyc08g062960.2.1	SolycHsfA2	SollyHsfA2	6.50E-195
Solyc09g009100.2.1	SolycHsfA3	SollyHsfA3	0
Solyc02g072000.2.1	SolycHsfA4c	SollyHsfA4c	3.60E-194
Solyc03g006000.2.1	SolycHsfA4a	SollyHsfA4a	2.50E-257
Solyc07g055710.2.1	SolycHsfA4b	SollyHsfA4b	1.60E-199
Solyc12g098520.1.1	SolycHsfA5	SollyHsfA5	0
Solyc09g065660.2.1	SolycHsfA7	SollyHsfA6	1.40E-160
Solyc09g082670.2.1	SolycHsfA6a	SollyHsfA6a	5.90E-152
Solyc09g059520.2.1	SolycHsfA8	SollyHsfA8	1.00E-210
Solyc02g072060.1.1	SolycHsfl1	SollyHsfA9	1.20E-57
Solyc02g079180.1.1	SolycHsfl2	SollyHsfA9	3.70E-34
Solyc07g040680.2.1	SolycHsfA9	SollyHsfA9	2.40E-220
Solyc02g090820.2.1	SolycHsfB1	SollyHsfB1	1.20E-148
Solyc03g026020.2.1	SolycHsfB2a	SollyHsfB2a	9.50E-158
Solyc08g080540.2.1	SolycHsfB2b	SollyHsfB2b	4.20E-208
Solyc04g016000.2.1	SolycHsfB3a	SollyHsfB3a	1.10E-155

Table 1. (Continued)

IDENTIFIER ITAG2.4	V1.0	V2.0	E-VALUE V2.0
Solyc10g079380.1.1	SolycHsfB3b	SollyHsfB3b	3.00E-162
Solyc04g078770.2.1	SolycHsfB4a	SollyHsfB4a	7.30E-174
Solyc11g064990.1.1	SolycHsfB4b	SollyHsfB4b	1.70E-112
Solyc02g078340.2.1	SolycHsfB5	SollyHsfB5	2.50E-160
Solyc12g007070.1.1	SolycHsfC1	SollyHsfC1	1.90E-176
Solyc06g053960.2.1	SolycHsfA6b	SollyN.C.	
Solyc11g008410.2.1	SolycHsfl3		

The table provides for each gene represented by the ITAG2.4 identifier, the corresponding annotation by the HEATSTER v1.0 and v2.0 as well as the e-value of the classification by the v2.0.

classification of a novel HSF. The output is also represented and downloadable in a modified MAST HTML format.

Applications of HEATSTER

We demonstrate the use of HEATSTER exemplarily on the genome-wide identification of HSFs in the ITAG2.4[14] reference genome of *S lycopersicum* (tomato). This corresponds to the first step in reconstructing the HSF network in a newly sequenced genome (Figure 3 outlines the procedure).

For comparison of the HEATSTER versions 1.0 and 2.0, we analyzed the tomato proteome with the HEATSTER batch search to predict HSFs in multiple sequences. In v1.0, we identified 15 HsfAs, 8 HsfBs, and 1 HsfC. Furthermore, 3 HSF-like sequences are known from literature.[10] The HEATSTER v2.0 could identify 26 of the known 27 HSFs (Table 1).

Furthermore, 2 of the 3 HSF-like sequences could be annotated as HsfA9s and only the very similar HsfA6 and HsfA7 showed differences in the classification between HEATSTER v1.0 and v2.0. As the 4 classified tomato HsfA1 sequences showed distinct patterns of signature motifs (Figure 3), a further, more fine-grained classification of the HSF, as indicated by lower case letters, is possible.

Acknowledgements

We thank Lutz Nover and Arndt von Haeseler for long-term endorsement and critical attendance of the work. We acknowledge Stefan Biermann for the transfer and hosting of the HEATSTER websites. We thank Maik Boehmer, Niclas Fester and Mario Keller for reading and improving the manuscript.

Author Contributions

JB, IE, SeS, and StS designed the bioinformatics analysis. JB implemented the underlying database as well as the website and visualization tools. JB, StS, KDS, and SeS performed the data mining and classification of HSFs. KDS, StS, JB, ES, and IE validate the results of the tools and tested the website. JB, StS, ES, IE, and KDS wrote the manuscript. All authors read and approved the final manuscript. JB and StS contribute equally to the work.

Availability of Data

Freely available at http://applbio.biologie.uni-frankfurt.de/hsf/heatster/.

Supplemental Material

Supplementary data are available at Bioinformatics and Biology insights

REFERENCES

1. Fragkostefanakis S, Roth S, Schleiff E, Scharf KD. Prospects of engineering thermotolerance in crops through modulation of heat stress transcription factor and heat shock protein networks. *Plant Cell Environ*. 2015;38:1881–1895.
2. Fragkostefanakis S, Simm S, Paul P, Bublak D, Scharf KD, Schleiff E. Chaperone network composition in Solanum lycopersicum explored by transcriptome profiling and microarray meta-analysis. *Plant Cell Environ*. 2015;38:693–709.
3. Chidambaranathan P, Jagannadham PTK, Satheesh V, et al. Genome-wide analysis identifies chickpea (Cicer arietinum) heat stress transcription factors (HSFs) responsive to heat stress at the pod development stage. *J Plant Res*. 2017; 131:525–542.
4. Dossa K, Diouf D, Cisse N. Genome-wide investigation of HSF genes in sesame reveals their segmental duplication expansion and their active role in drought stress response. *Front Plant Sci*. 2016;7:1522.
5. Guo M, Liu JH, Ma X, Luo DX, Gong ZH, Lu MH. The plant heat stress transcription factors (HSFs): structure, regulation, and function in response to abiotic stresses. *Front Plant Sci*. 2016;7:114.
6. Mittal D, Chakrabarti S, Sarkar A, Singh A, Grover A. Heat shock factor gene family in rice: genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol Biochem*. 2009;47:785–795.
7. Hahn A, Bublak D, Schleiff E, Scharf KD. Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. *Plant Cell*. 2011;23: 741–755.
8. Nover L, Bharti K, Doring P, Mishra SK, Ganguli A, Scharf KD. Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress Chaperones*. 2001;6:177–189.
9. Nover L, Scharf KD, Gagliardi D, Vergne P, Czarnecka-Verner E, Gurley WB. The HSF world: classification and properties of plant heat stress transcription factors. *Cell Stress Chaperones*. 1996;1:215–223.
10. Scharf KD, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (HSF) family: structure, function and evolution. *Biochim Biophys Acta*. 2012;1819:104–119.
11. Baniwal SK, Bharti K, Chan KY, et al. Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. *J Biosci*. 2004;29:471–487.
12. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202–W208.
13. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*. 1998;14:48–54.
14. Sato S, Tabata S, Hirakawa H, et al. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012;485:635–641.