Heterogeneous expression pattern of tandem duplicated sHsps genes during fruit ripening in two tomato species

Arce DP$^{1,2}$ Krsticevic FJ$^{3}$ Ezpeleta J$^3$ Ponce SD$^2$ Pratta GR$^1$ Tapia E$^3$

$^1$Cátedra de Genética. Departamento de Biología. Facultad de Ciencias Agrarias. Universidad Nacional de Rosario. 
$^2$Grupo de Análisis, Desarrollos e Investigaciones Biomédicas. Facultad Regional San Nicolás. Universidad Tecnológica Nacional.
$^3$Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas. CONICET- UNR- AMU.

*These authors contributed equally to this work
Corresponding author’s: debora.arce@gmail.com

Abstract. The small heat shock proteins (sHSPs) have been found to play a critical role in physiological stress conditions protecting proteins from irreversible aggregation. To characterize the gene expression profile of four sHsps with a tandem gene structure arrangement in the domesticated Solanum lycopersicum (Heinz 1706) genome and its wild close relative Solanum pimpinellifolium (LA1589), differential gene expression analysis using RNA-Seq was conducted in three ripening stages in both cultivars fruits. Gene promoter analysis was performed to explain the heterogeneous pattern of gene expression found for these tandem duplicated sHsps. *In silico* analysis results contribute to refocus wet experiment analysis in tomato sHsp family proteins.

1. Introduction
Small heat shock proteins (sHsps) are low molecular weight (12-40 kDa) proteins that play a critical role in physiological stress conditions protecting proteins from irreversible aggregation and stress tolerance. Furthermore, some sHsps are expressed during certain stages of development, such as embryogenesis, germination, pollen development, and fruit maturation [1]. The special importance sHsp family in plants is suggested by their unusual abundance and diversity. Small Hsps are found in distinct cellular compartments like cytosol, chloroplast, endoplasmic reticulum, mitochondria and membranes [2]. The sHsp family possess a conserved domain of approximately 80 to 100 amino acids called alpha-crystalline domain (ACD), flanked by N- and C-terminal regions. The number of genes in the sHsp family is overestimated because the same conserved domain IPR008978 (ACD or HSP20 domain) has become incorporated into a large number of unrelated proteins, known as ACD-like or HSP20-like [3].

Tomato has become the model system to study organ expansion, maturity, ripening, shelf-life and fruit nutritional quality [4]. Tomato, the centerpiece of the Solanaceae family, has emerged as a model of fleshy fruit development, primarily because this is the species for which the genetic and molecular toolkits are most advanced. The inbred tomato S. lycopersicum cultivar ‘Heinz 1706’ and its wild relative S. pimpinellifolium LA1589 were sequenced and estimated divergence between the wild and domesticated genomes is 0.6% (5.4 million single nucleotide polymorphisms distributed along the chromosomes). Consequently, due ancestrally similar gene structure and expression profiles are expected for sHsp family in both cultivars. Transcriptome analysis by next generation sequencing technology (NGS), such as high-throughput RNA sequencing (RNA-Seq) have led to the
characterization of previously unrecognized mechanisms of gene regulation, in-depth sequence analysis, transcriptional structure and expression level for each gene in these two tomato cultivars [5]–[7].

Analysis of the tomato *S. lycopersicum* cv Heinz 1706 transcriptome indicated a total of 2459 loci that could be annotated as transcription or transcriptional regulators of which 223 show expression patterns that change substantially between mature unripe and ripening fruit and thus potentially influencing ripening processes [7]. From a preliminary analysis combining phylogeny and RNA-Seq expression patterns in the sHsp family, we found a subset of four sequences. These related sequences were originated in two independent events at chromosome 1 (Soly01g098790 and Soly01g098810) and 4 (Soly04g082720 and Soly04g082740) by tandem duplication event, showed an heterogeneous pattern of expression (unpublished Krsticevic et al, personal communication) and maybe the promoter regulation could be explain this fact.

The aim of this work was to analyze gene promoter region to further characterize sHsps transcriptional regulation during maturation in the genome of domesticated *S. lycopersicum* cv. Heinz 1706 and its wild close relative *S. pimpinellifolium*. Furthermore, we compare gene expression profile of four sHsp in *S. pimpinellifolium* with the already data obtained for the commercial variety. Differential sHsp gene expression analysis using RNA-Seq was conducted in three ripening stages (G, BR and RR) in *S. lycopersicum* cv Heinz 1706 and *S. pimpinellifolium* LA1589 fruits. Finally, 1000 bp upstream of gene promoter region for each gene was performed and putative transcriptional regulation during maturation was explained.

2. **Materials and methods**

*S. lycopersicum* and *S. pimpinellifolium* transcriptome Illumina datasets

The Illumina RNA-Seq reads files from two biological replicates of *S. lycopersicum* (cv Heinz 1706) mature green or G (SRR404324, SRR404325), breaker or BR (SRR404326, SRR404327), red ripening fruits or RR (SRR404328, SRR404329) and *S. pimpinellifolium* (LA 1589) G (SRR404331, SRR404333), BR (SRR404334, SRR404336) and RR (SRR404338, SRR404339) were downloaded from the Sequence Read Archive (SRA from DDBJ, [http://trace.ddbj.nig.ac.jp/dra/index_e.html](http://trace.ddbj.nig.ac.jp/dra/index_e.html)). These data were used by The Tomato Genome Consortium (Supplementary material, 2.3.5 Transcriptome sequencing pp 46). The SRA files were converted to FASTQ files using the fastq-dump utility of the SRA toolkit ([http://eutils.ncbi.nlm.nih.gov/Traces/sra/?view=software](http://eutils.ncbi.nlm.nih.gov/Traces/sra/?view=software)). The nomenclature used to describe the different stages of tomato fruit development is G: mature green, BR: mature orange-breaker, RR: mature red ripening stages [9].

**Bowtie alignment and counting**

The Bowtie short read aligner (v. 1.0.1) was used to map the reads against a set of reference genomic sequences [10]. To assess the expression levels of each gene in the different samples (G, BR, RR), the obtained raw counts (Table of counts) after Bowtie software alignment were used for differential gene expression analysis.

**Alignment**: The 4 sHsp tandem duplicated sequences plus the HSP70 as positive control of maturation RR stage and other two sHsps sequences which have previous experimental evidences for up or down-regulation at RR stage were used to build the reference library. Whenever possible, we tried to include ~1 kb up- and down-stream of the CDC region (~3 kb), taking care to prevent overlapping with other CDC gene regions. This procedure helped us ensure that all the sHsps reads (even those aligned to the 5' and 3' ends) were included in our analysis. The genomic regions contained sHsps CDC of Soly01g098790, Soly01g098810, Soly04g082720, Soly04g082740, Soly12g042830, Soly08g062450, Soly09g011710, HSP70 and their orthologs in *S. pimpinellifolium* genomic sequences. HSP70 gene region was used as a positive control for the *in silico* experiment, because is known to be up-regulated during the fruit maturation process [11]. The other sHsps used as controls with experimental evidences included were mitochondrial Soly12g042830 (up-regulated in *S.
lycopersicum RR stage), class II Solyc08g062450 (up-regulated in S. lycopersicum RR) and class I Solyc09g011710 (down-regulated in S. lycopersicum RR). The nomenclature used to describe the sHsps in S. lycopersicum is S01g08790_SL, S01g098810_SL, S04g082720_SL and S04g082740; in S. pimpinellifolium sHsps the letter “L” change for a “P”. The number of reads that aligned to each gene provided a quantification of how many RNA transcripts of that gene were present in the sample. Mapped reads were quantified from SAM output files using GNU commands to obtain a Table of counts to further statistical analysis.

Transcript abundance
The Fragments Per Kilobase of transcript per Million (FPKM) mapped fragments were calculated manually after Bowtie alignment to determine transcript abundance (mRNA absence or presence) for each gene analyzed. Mapped reads were obtained as described above and average FPKM were calculated for the two biological replicates of each sample and for each gene. Genes were considered expressed when the average FPKM of both biological replicates was greater than or equal to 2, as described by Steijger et al [12] calculated that fragments per kilobase of gene model per million mapped reads (FPKM) values of 1 to 2 correspond to ~1 mRNA molecule per cell [13].

Statistics for differential gene expression analysis
Descriptive and RNA-Seq statistics were calculated using the edgeR Bioconductor package in R, with two modifications [14]. The standard edgeR pipeline calculates library sizes as the sum of read counts for each gene, assuming that the library is mapped to the entire genome. In our work, reads were instead mapped to a reference constructed with a set of genomic regions which were extracted using the Solgenomics tomato genome browser (Jbrowse, genome version S. lycopersicum ITAG2.40 release and S. pimpinellifolium draft genome WGS Contigs cshl_1.00). Therefore, to prevent edgeR from incorrectly assuming that the library sizes were equal to the sum of read counts for each gene, we manually input the correct library sizes, which were first calculated by aligning the reads to the complete tomato reference genome. In addition, the calcNormFactors function was omitted because it is also intended for genome-wide expression studies, as it performs a normalization which assumes that most genes are not differentially expressed. Instead, the present study focuses on a subset of genes, a significant number of which are expected to be differentially expressed, which makes the above correction not only unnecessary but also counterproductive. The G stage was taken as baseline and edgeR’s exact test was applied to calculate the log₂ fold changes (log FC) of the BR and RR stages relative to this baseline. Positive log₂ FC values indicate up-regulation, negative values indicate down-regulation and zero indicates constant gene expression relative to the G stage. Transcripts whose absolute log₂ fold change (log₂ FC) were greater than 1 and whose p-value was lower than 0.01 for either BR or RR were included in the plots.

Gene promoter analysis
Gene promoter sequences were extracted 1000 pb upstream of each gene transcription start site (TSS) Sequences were obtained from Solgenomics (http://solgenomics.net/). For gene promoter extraction in S. pimpinellifolium genome bed tools were also used (http://bedtools.readthedocs.org/en/latest/). Plant CARE database was used to find specific TFBS for tomato and Arabidopsis plants [15].

3. Results

Transcript abundance: FPKM quantification
Transcript abundance from four S. lycopersicum (SL) and S. pimpinellifolium (SP) sHps (S04g082720, S04g082740, S01g098790, and S01g098810) was estimated by counting Illumina reads from mature green (G), breaker (BR) and mature red ripening (RR) fruit stages aligned to their genomic regions using short read aligner Bowtie1. Results are shown as FPKM quantification in Table 1. Variability in
transcript levels was observed from G to RR maturing stages. Small Hsps with previous experimental evidences with different subcellular localization were also used as controls (up-regulated: mitochondrial S12g042830, class II S08g062450; or down-regulated: class I S09g011710). The highest FPKM values were observed in S04g082720 in *S. lycopersicum* and S08g062450 (control) in both cultivars. The S04g082740 sequence showed an opposite transcript profile with a decreased in FPKM levels as development advanced in *S. lycopersicum* and *pimpinellifolium*. The columns RR-G in Table 1 represent differences in transcript abundances between RR and G ripening stages, showing positive values as up-regulated and negative values as down-regulated genes.

Table 1. Small Hsps FPKM quantification from RNA-Seq in ripening tomato fruits. Expression transcript levels were quantified as FPKM for three ripening fruit stages: mature green (G), breaker (BR) and red ripe (RR) in *S. lycopersicum* and *pimpinellifolium* tomato fruits.

| Gene ID      | *S. lycopersicum* | *S. pimpinellifolium* |
|--------------|-------------------|-----------------------|
|              | G | BR  | RR  | RR-G   | G        | BR   | RR   | RR-G   |
| S01g098790   | 0.26709592 | 0.37646615 | 0.33152983 | 0.06443391 | 1.10390168 | 0.61207048 | 0.34932573 | -0.75457595 |
| S01g098810   | 9.28909591 | 8.97380435 | 5.98045634 | -3.30863957 | 4.37726509 | 2.98985227 | 2.11367002 | -2.26359508 |
| S04g082720   | 44.0794412 | 52.4421379 | 103.57674  | 59.4972988  | 29.2965039 | 48.006047  | 78.6167481 | 49.3202442 |
| S04g082740   | 5.50757926 | 2.29629047 | 0.55631675 | -4.95126251 | 4.75406219 | 2.86131773 | 1.16123017 | -3.59283201 |
| S08g062450   | 64.853396  | 421.062959 | 819.664401 | 754.811005  | 72.2412884 | 113.097962 | 369.092026 | 441.333315 |
| S09g011710   | 186.614016 | 115.135764 | 58.9809129 | -127.705103 | 270.269463 | 161.215473 | 99.9142164 | -170.355247 |
| S12g042830   | 8.47584812 | 44.4285468 | 87.3254453 | 79.0495972  | 58.2405858 | 293.787118 | 202.543895 | 144.303309 |

**Differential gene expression analysis**

To evaluate if these four sHsps might be involved specifically in any of the three ripening stages, we compared sHsps expression levels in RR and BR stages using G ripening stage as baseline levels. Figure 1 show sHsps transcript levels in BR and RR stages using edgeR package. Genes with log2FC values lower than 1 mean that transcript levels were not differentially expressed (NDE) in BR or RR stages compared to G stage. Tandem chromosome 1 sHsps transcripts (S01g098790 and S01g098810) were NDE in the BR and RR stages relative to the G fruit stage in *S. lycopersicum* and *pimpinellifolium* except for S01g098810 which was down-regulated at both stages BR and RR in *S. pimpinellifolium* (Fig. 1). Chromosome 4 sHsp (S04g082740) was also NDE in *S. pimpinellifolium* but it was down-regulated at BR and RR stages in *S. lycopersicum*. Only one sHsp (S04g082720) from these four tandem genes analyzed showed up-regulation in BR and RR ripening stages compared to G fruits in both cultivars. Because tandem genes are expected to show the same expression pattern, gene promoter regions might explain this distinct expression profile (see below in Discussion). Small Hsps used as controls with previous experimental evidences showed the expected differential gene expression profiles: S08g062450 and S12g042830 were up-regulated in both cultivars. HSP70 was up-regulated except at BR stage in *S. pimpinellifolium*. 
Small Hsps gene promoter analysis

Results from gene promoter analysis are shown in Table 2. Note that column with differential gene expression results comparing RR versus G stage (DGE RR) from Fig. 1 were included showing up-regulation (Up), down-regulation (Do), and not differentially expressed (NDE) genes in Table 2. Two gene relative to chromosome 1 (S01g098810_SL, S01g098810_SP), and two from chromosome 4(S04g082720_SL, and S04g082720_SP) showed the presence of heat stress motif (heat shock element, HSE). Interestingly, tandem chromosome 1 sHsps showed NDE despite the presence or absence of HSE in S. lycopersicum. Contrarily, the promoter of the down-regulated S01g098810 showed the presence of HSE in S. pimpinellifolium. Mitochondrial S12g042830 in S.pimpinellifolium does not show HSE. All gene promoters of sHsps used as controls have HSE motifs. We also looked for motifs which belong to different stress and hormonal regulation pathways like abscisic acid or ABA (ABRE), ethylene (ethylene response, ERE), gibberellins (GA) and transcription factors (MYB, WRKY) in sHsps gene promoter sequences. These motifs were searched because they are over-represented in gene promoter regulatory sequences from genes differentially expressed during stress, maturation and development processes [16], [17]. Variability in these last motifs was observed in the sequences analyzed.

Pairwise sequence analysis between promoters of orthologues sHps in chromosome 1 and 4 were also analyzed with Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html) and identity and similarity (as percentage) are shown in column “ID/Sim (%).” The highest values were observed for pairs S01g098790_SL/S01g098790_SP, S04g082740_SL/S04g082740_SP and S04g082720_SL/S04g082720_SP. In opposite, pair S01g098810_SL/ S01g098810_SP showed the lowest ID/Sim (%) with a value of 43.8%.
Table 2. sHsps gene promoter analysis in *S. lycopersicum* (SL) and *pimpinellifolium* (SP)

| Promoter         | ERE | MYB | ABRE | AUXIN | GA | WRKY | HSE | DGE RR | ID/Sim (%) |
|------------------|-----|-----|------|-------|----|------|-----|--------|------------|
| S01g098790_SL    | -   | +   | +    | +     | +  | -    | +   | NDE    | 99.1/99.1  |
| S01g098790_SP    | -   | +   | +    | -     | +  | -    | +   | NDE    | 43.8/43.8  |
| S01g098810_SL    | +   | +   | -    | -     | -  | +    | -   | Do     | 81.2/81.2  |
| S01g098810_SP    | -   | -   | +    | -     | -  | +    | -   | Do     | 99.6/99.6  |
| S04g082720_SL    | +   | +   | -    | -     | -  | +    | -   | Up     | ND         |
| S04g082720_SP    | -   | -   | -    | -     | -  | +    | -   | Up     | ND         |
| S04g082740_SL    | -   | +   | -    | -     | -  | +    | -   | Do     | ND         |
| S04g082740_SP    | -   | -   | +    | +     | -  | -    | +   | Do     | ND         |
| S08g062450_SL    | -   | -   | -    | -     | -  | +    | +   | Up     | ND         |
| S08g062450_SP    | -   | -   | -    | -     | -  | +    | -   | Up     | ND         |
| S12g042830_SL    | +   | -   | +    | -     | -  | +    | -   | Up     | ND         |
| S12g042830_SP    | +   | -   | -    | +     | +  | -    | -   | Up     | ND         |
| S09g011710_SL    | -   | +   | -    | +     | -  | +    | -   | Do     | ND         |
| S09g011710_SP    | -   | +   | +    | +     | -  | -    | +   | Do     | ND         |

4. Discussion
In the present work differential gene expression and promoter analysis was conducted for four sHsps with tandem gene structure. This gene group has heterogeneous pattern of expression. The chromosome 1 gene member S01g098790 is NDE in both cultivars and has not HSE in the promoter regions. These members might represent pseudogenes because its transcript abundance levels are very low in all tissues assayed as few reads were observed in the Bowtie alignment. Anyway, they show a similar expression profile (NDE) as is expected for copies with high percentage of identity. The typical ACD domain is absent in their protein structure (data not shown) which does not mean that a pseudogene cannot be expressed. Contrarily, some pseudogenes can also be expressed and their expression levels tend to be lower compared with annotated genes in Arabidopsis and rice [18]. Furthermore, tandem chromosome 1 sHsps in *S. pimpinellifolium* showed a different pattern of expression where S01g098810 gene promoter showed a combination of HSE, GA and ABRE motifs that might be enough to trigger down-regulation of gene expression.

Contrarily, S01g098810_SL and S01g098810_SP have heterogeneous expression pattern since one member is NDE and the other is down-regulated. Differences in gene expression could be explained by pairwise gene promoter alignment showing an ID/Sim of 43.8%. This value is low and might be indicating gene sub functionalization due to differences at their promoter regions.

Interestingly, S04g082720_SL and S04g082720_SP show similar up-regulation expression profiles, as was expected by their high percentage of identity in their genomic sequences. These two chromosome 4 gene copies have HSE in their promoter regions and present similar expression profiles (Up-regulation) like it observed by the positive controls used in this work (S08g062450_SL and S12g042830 in both cultivars). All these findings support the hypothesis that despite S04g082720_SL and S04g082720_SP don not have the typical ACD domain, they act as if they were functional shsps in the RR stage, indicating that S04g082720 could be specifically involved in *S. lycopersicum* and *pimpinellifolium* fruit ripening process. These results are according to previous findings by Matas et al.
who showed that S04g082720 referred as TU143174 is differentially expressed specifically in the collenchyma and inner epidermis of the *S. lycopersicum* fruit (cv Ailsa Craig) [19].

On the other hand, S04g082740_SL and S04g082740_SP show down-regulation and NDE in RR stage compared to G stage expression profiles. Although gene promoter analysis showed a high percentage of identity (99.6%) they do not have HSE. MYB motif TFBS might explain this results since its presence in S04g082740_SL is the only difference between these two promoter sequences.

HSE were identified and these TFBS are always present in the promoter region of heat shock genes. The HSE is the binding target site for recognition of the heat shock factors (Hsfs) [21]. The essential role of HSE for heat-dependent transcriptional activation in plants has been established by promoter deletion assays in tomato [20], [21]. Moreover, HSE present in ascorbate peroxidase 1 (apx1) gene promoter is involved in antioxidative response in Arabidopsis indicating that heat and oxidative pathways could be related and share transcriptional regulation mechanisms [22]. Note in our results that none specific combination of ERE, ABRE, GA, MYB, WRKY and HSE motifs is necessary to trigger gene expression but suggest that heat response and hormone mediated stress signaling may share common pathways where oxidative stress is implicated. Moreover, recent study with a similar strategy (promoter analysis and regulatory network prediction of non-redundant DREB genes stored in RefSeq database in Arabidopsis) was accomplished by use of Plant CARE program and GeneMANIA web tool, respectively. The results indicated that DREB genes have diverse motifs involved in abiotic stress responses and cooperation of phytohormones like abscisic acid, ethylene, salicylic acid and methyl jasmonate signaling is crucial for the regulation of the expression of abiotic stress (drought and cold responses) through *DREB* transcription factors [23].

Further promoter analysis, protein structure analysis, phylogenetic analysis and experimental evidences should be conducted to elucidate the functionality of these tandem genes in chromosome 1 and 4 in tomato genomes. There is a need to exploit the transcription factors associated with sHsps and its regulation and differential expression in order to use it for developing more resistant cultivars to stress and better fruit quality.

5. **Conclusions**

RNA-Seq transcriptome analysis combined with structural protein and gene promoter analysis carried out in this study contributed to interpret the transcriptional regulation cascade of 4 sHsps in tandem arrangement during ripening in *S. lycopersicum* (Heinz 1706) and *S. pimpinellifolium* (LA1589). Chromosome 4 S04g082720 act as a functional sHsp in the RR stage in both cultivars, indicating that it could be specifically involved in *S. lycopersicum* and *S. pimpinellifolium* fruit ripening process.

Promoter analysis conducted in this work might explain some gene regulation aspects of sHsps but wet experiments like Chip-Seq should be considered to clarify the functionality of sHsps and its gene regulation mechanism. Gene network during fruit ripening is a complex process mediated by different counterparts and probably the sHsp family might have some role in this networking process.

**References**

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

[2] W. X. Wang, B. Vinocur, O. Shoseyov, and A. Altman, “Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response,” *Trends Plant Sci.*, vol. 9, pp. 244–252, 2004.
[3] R. L. van Montfort, E. Basha, K. L. Friedrich, C. Slingsby, and E. Vierling, “Crystal structure and assembly of a eukaryotic small heat shock protein,” Nat. Struct. Biol., vol. 8, pp. 1025–1030, 2001.

[4] H. J. Klee and J. J. Giovannoni, “Genetics and control of tomato fruit ripening and quality attributes,” Annu. Rev. Genet., vol. 45, pp. 41–59, Jan. 2011.

[5] K. Aoki, K. Yano, A. Suzuki, S. Kawamura, N. Sakurai, K. Suda, A. Kurabayashi, T. Suzuki, T. Tsugane, M. Watanabe, K. Ooga, M. Torii, T. Narita, T. Shin-I, Y. Kohara, N. Yamamoto, H. Takahashi, Y. Watanabe, M. Egusa, M. Kodama, Y. Ichinose, M. Kikuchi, S. Fukushima, A. Okabe, T. Arie, Y. Sato, K. Yazawa, S. Satoh, T. Omura, H. Ezura, and D. Shibata, “Large-scale analysis of full-length cDNAs from the tomato (Solanum lycopersicum) cultivar Micro-Tom, a reference system for the Solanaceae genomics,” BMC Genomics, vol. 11, p. 210, Jan. 2010.

[6] K. Aoki, Y. Ogata, K. Igarashi, K. Yano, H. Nagasaki, E. Kaminuma, and A. Toyoda, “Functional genomics of tomato in a post-genome-sequencing phase,” Breed. Sci., vol. 63, no. 1, pp. 14–20, Mar. 2013.

[7] T. Tomato and G. Consortium, “The tomato genome sequence provides insights into fleshy fruit evolution,” Nature, vol. 485, pp. 635–41, 2012.

[8] F. Zambelli, G. Pesole, and G. Pavesi, “Motif discovery and transcription factor binding sites before and after the next-generation sequencing era,” Brief. Bioinform., vol. 14, no. 2, pp. 225–37, Mar. 2013.

[9] P. Pesaresi, C. Muzzetti, M. Colombo, and S. Masiero, “Genetic regulation and structural changes during tomato fruit development and ripening,” Front. Plant Sci., vol. 5, no. April, p. 124, Jan. 2014.

[10] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, “Ultrafast and memory-efficient alignment of short DNA sequences to the human genome,” Genome Biol., vol. 10, no. 3, p. R25, Jan. 2009.

[11] S. Fragkostefanakis, S. Simm, P. Paul, D. Bublik, K.-D. Scharf, and E. Schleiff, “Chaperone network composition in S.olanumlycopersicum explored by transcriptome profiling and microarray meta-analysis,” Plant Cell Environ., vol. 38, no. 4, pp. 693–709, Apr. 2015.

[12] T. Steijger, J. F. Abril, P. G. Engström, F. Kokocinski, M. Akerman, T. Alioto, G. Ambrosini, S. E. Antonarakis, J. Behr, P. Bertone, R. Bohnert, P. Bucher, N. Cloonan, T. Derrien, S. Djebali, J. Du, S. Dudoit, M. Gerstein, T. R. Gingeras, D. Gonzalez, S. M. Grimmond, R. Guigó, L. Habegger, J. Harrow, T. J. Hubbard, C. Iseli, G. Jean, A. Kahles, J. Lagarde, J. Leng, G. Lefèbvre, S. Lewis, A. Mortazavi, P. Niermann, G. Rätsch, A. Reymond, P. Ribeca, H. Richard, J. Rougemon, J. Rozowsky, M. Sammeth, A. Sboner, M. H. Schulz, S. M. J. Searle, N. D. Solarzcano, V. Solovyev, M. Stanke, B. J. Stevenson, H. Stockinger, A. Valsesia, D. Weese, S. White, B. J. Wold, J. Wu, T. D. Wu, G. Zeller, D. Zerbino, and M. Q. Zhang, “Assessment of transcript reconstruction methods for RNA-seq,” Nat. Methods, vol. 10, pp. 1177–84, 2013.

[13] T. Hart, H. K. Komori, S. Lamere, K. Podshivalova, and D. R. Salomon, “Finding the active genes in deep RNA-seq gene expression studies,” 2013.

[14] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data,” Bioinformatics, vol. 26, no. 1, pp. 139–40, Jan. 2010.
[15] B. Mohanty, S. P. T. Krishnan, S. Swarup, and V. B. Bajic, “Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species,” *Ann. Bot.*, vol. 96, no. 4, pp. 669–81, Sep. 2005.

[16] V. Petrov, V. Vermeirssen, I. De Clercq, F. Van Breusegem, I. Minkov, K. Vandepoele, and T. S.Gechev, “Identification of cis-regulatory elements specific for different types of reactive oxygen species in Arabidopsis thaliana,” *Gene*, vol. 499, no. 1, pp. 52–60, May 2012.

[17] E. Picot, P. Krusche, A. Tiskin, I. Carré, and S. Ott, “Evolutionary analysis of regulatory sequences (EARS) in plants,” *Plant J.*, vol. 64, no. 1, pp. 165–76, Oct. 2010.

[18] C. Zou, M. D. Lehti-Shiu, F. Thibaud-Nissen, T. Prakash, C. R. Buell, and S.-H.Shiu, “Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice,” *Plant Physiol.*, vol. 151, no. 1, pp. 3–15, Sep. 2009.

[19] A. J. Matas, T. H. Yeats, G. J. Buda, Y. Zheng, S. Chatterjee, T. Tohge, L. Ponnala, A. Adato, A. Aharoni, R. Stark, A. R. Fernie, Z. Fei, J. J. Giovannoni, and J. K. C. Rose, “Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation,” *Plant Cell*, vol. 23, no. 11, pp. 3893–910, Nov. 2011.

[20] L. Nover, K. D. Scharf, D. Gagliardi, P. Vergne, E. Czarnecka-Verner, and W. B. Gurley, “The Hsf world: classification and properties of plant heat stress transcription factors,” *Cell Stress Chaperones*, vol. 1, pp. 215–223, 1996.

[21] S. K. Baniwal, K. Bharti, K. Y. Chan, M. Fauth, A. Ganguli, S. Kotak, S. K. Mishra, L. Nover, M. Port, K.-D. Scharf, J. Tripp, C. Weber, D. Zielinski, and P. von Koskull-Döring, “Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors,” *J. Biosci.*, vol. 29, pp. 471–487, 2004.

[22] S. Storozhenko, P. De Pauw, M. Van Montagu, D.Inzé,S.Kushnir,“The heat-shock element is a functional component of the Arabidopsis APX1 gene promoter” *Plant Physiol.*, vol. 118, no. 3, pp. 1005–1014, 1998.

[23] S. Sazegari, A. Niazi, and F. S. Ahmadi, “A study on the regulatory network with promoter analysis for Arabidopsis DREB -genes,” vol. 11, no. 2, pp. 2–7, 2015.