

SCIENTIFIC REPORTS



OPEN

A comprehensive analysis of the Korean fir (*Abies koreana*) genes expressed under heat stress using transcriptome analysis

Jung Eun Hwang¹, Yun Jeong Kim¹, Myung Hwan Shin¹, Hwa Ja Hyun², Hans J. Bohnert³ & Hyeong Cheol Park¹

Korean fir (*Abies koreana*), a rare species endemic to South Korea, is sensitive to climate change. Here, we used next-generation massively parallel sequencing technology and *de novo* transcriptome assembly to gain a comprehensive overview of the Korean fir transcriptome under heat stress. Sequencing control and heat-treated samples of Korean fir, we obtained more than 194,872,650 clean reads from each sample. After *de novo* assembly and quantitative assessment, 42,056 unigenes were generated with an average length of 908 bp. In total, 6,401 differentially expressed genes were detected, of which 2,958 were up-regulated and 3,443 down-regulated, between the heat-treated and control samples. A gene ontology analysis of these unigenes revealed heat-stress-related terms, such as "response to stimulus". Further, in depth analysis revealed 204 transcription factors and 189 Hsps as differentially expressed. Finally, 12 regulated candidate genes associated with heat stress were examined using quantitative real-time PCR (qRT-PCR). In this study, we present the first comprehensive characterisation of Korean fir subjected to heat stress using transcriptome analysis. It provides an important resource for future studies of Korean fir with the objective of identifying heat stress tolerant lines.

Korean fir (*Abies koreana*), a valuable tree species for ornamental purposes, is an endemic and rare species in Korea¹. It grows in the upper regions of Mounts Halla, Chiri, Mudung, Kaji and Dukyu, which are located in the southern part of the Korean peninsula¹. Recently, the Korean fir population underwent a large dieback, resulting in a severe decline. This dieback may have been the result of complex interactions among multiple environmental factors caused by global warming². Because this species is susceptible to climate changes^{2,3}, it was designated an indicator species illustrating climate change by the Korean government⁴. The resistance to high temperature has been investigated in few tree species, including *Acer rubrum*⁵, *Picea glauca*⁶. For most trees, tolerance to high temperatures remains largely unstudied. Thus, it is essential to uncover and understand molecular response mechanisms of the Korean fir to heat stress, with the objective of gaining an understanding of mechanisms that impede heat tolerance in this species.

High temperatures can reduce plant growth and development, which may become a major issue in the coming years owing to global warming⁷. Global temperatures are predicted to rise by an additional at least 2 °C, and possibly more, by the end of this century. Therefore, a better understanding of plant heat tolerance mechanisms is crucial for the sustainable preservation of forest. Plants respond to high temperatures by altering the expression levels of thousands of genes, thereby changing cellular, physiological, and biochemical processes. When experiencing stress, the expression of a number of genes changes, enabling plants to respond to the abiotic stressor and consequently to control or mitigate deleterious effects. Several transcriptional regulatory networks are involved in stress-induced changes in gene expression⁸. In a hierarchical mode, stress-induced genes can up-regulate the expression levels of many downstream genes that then may provide abiotic stress tolerance to extreme temperatures, drought, or high salinity^{9,10}. These include various transcription factors (TFs) such as dehydration responsive element-binding (DREB) and heat shock transcription factors (HSFs), suggesting the roles of various

¹Division of Ecological Conservation, Bureau of Ecological Research, National Institute of Ecology, Seocheon, Republic of Korea. ²National Institute Forest Science Warm Temperate and Subtropical Forest Research Center, Jeju, Republic of Korea. ³Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA. Correspondence and requests for materials should be addressed to H.C.P. (email: hcpark@nie.re.kr)

Sample	Raw reads	Clean reads	Clean bases	GC (%)	Q20 (%)	Q30 (%)
Control_1	237,953,246	236,917,754	23,844,037,594	44.28	95.61	97.52
Control_2	207,999,230	207,072,292	20,854,653,490	44.33	95.42	97.44
Control_3	228,337,054	227,273,940	22,875,259,422	44.99	95.41	97.43
Heat-treated_1	234,835,376	233,574,546	23,512,736,318	45.22	95.48	97.47
Heat-treated_2	196,808,520	194,872,650	19,606,902,313	45.83	95.16	97.25
Heat-treated_3	224,823,004	223,806,200	22,542,923,111	45.31	95.52	97.48

Table 1. Quality of Korean fir's sequencing.

Type	Contigs	Unigenes
Total trinity transcripts	406,207	42,056
Minimum length (bp)	201	201
Maximum length (bp)	19,314	19,314
Average length (bp)	472.74	908
N50 (bp)	532	1,608
Total length (bp)	192,031,706	51,118,088

Table 2. Length distributions of the assembled *Abies koreana* transcripts.

transcriptional regulatory mechanisms in the stress-signal transduction pathways⁹. These TFs regulate the expression of heat shock proteins (HSPs), which play a central role in the heat-stress response¹¹. The tolerance conferred by HSPs then reinforces physiological functions, such as photosynthesis, assimilate partitioning, water and nutrient use efficiency levels, and membrane stability⁷. However, there are differences in responses to heat stress among different species and often also genotypes of one species¹². For the vast majority of species, their transcriptomes are even now largely uncharacterised, and even in species for which substantial information is available, it is often in the form of partially characterized transcriptomes only.

The analysis of gene expression levels is a valuable tool in understanding the transcriptome dynamics and the potential for manipulating gene expression patterns in plants. The development of a high-throughput sequencing technology, RNA sequencing (RNA-seq), has been successfully used for gene expression profiling and other transcriptome studies in an increasing number of plants, including *Arabidopsis*¹³, rice¹⁴, and poplar¹⁵. In the past, the *de novo* assembly of very short-read sequences without a known reference posed often severe difficulties¹⁶. However, the recent development and optimisation of a *de novo* short-read assembly methods now allows for rapid and cost-effective assembly of transcriptomes also of non-model organisms with unknown genomes, opening the door for performing numerous and substantial new analysis opportunities^{17,18}. This method made possible sequencing of transcriptomes in species not previously approachable, such as *Picea abies*^{19,20}. However, no comparative transcriptomic analyses have been performed using next-generation sequencing technologies in the genus *Abies* under an environmental stress simulation.

In our study, the genome-wide analysis of gene expression levels during a heat treatment was performed with Korean fir in the genus *Abies* using a next generation sequencing-based Illumina paired-end platform. The data were used to create a reference transcriptome, containing an extensive set of genes expressed under heat treatment. Moreover, we sought and identified genes with important roles in heat tolerance. The results will provide understanding of the adaptive heat-stress mechanisms in Korean fir at the molecular level, and increase the genomic information available for the genus *Abies*.

Results

Transcriptome sequencing and *de novo* assembly. To elucidate the molecular responses to heat stress in Korean fir, we prepared six libraries from heat-treated and control samples for sequencing. In total, 1,330,756,430 raw reads (ranging from 196,808,520 to 237,953,246 for each sample) were obtained from the control and heat-treated samples (Table 1). From these, 236,917,754; 207,072,292; 227,273,940; 233,574,546; 194,872,650 and 223,806,200 clean reads were obtained from each sample, respectively (Table 1). Of the clean reads, the Q20 percentage (sequencing error rate < 1%) was over 95% and the G + C content was approximately 45% for all libraries (Table 1). The six libraries reads were combined to assemble the Korean fir transcriptome. Transcriptome *de novo* assembly was performed using Trinity software, which generated 406,207 transcripts with average length of 472.74 bp and an N50 of 532 bp for the merged assembly of six libraries (Table 2). The N50 value improved to 1,608 bp after running CD-HIT and filtered raw read count with total number of transcripts reduced to 42,056 (Table 2). The high-quality reads produced in this study have been deposited in the NCBI SRA database (accession number: SRR6781816, SRR6781817, SRR6781818, SRR6781819, SRR6781820 and SRR6781821).

Functional annotation and classification of the Korean fir transcriptome. For annotation purposes, the 42,056 assembled unigenes were analysed for gene ontology (GO) terms using gene ontology database and WEGO. Altogether, 22,880 unigenes, 54.40% of the total unigenes, were annotated using the GO database. The annotated Korean fir genes were functionally categorised based on the GO classification system, which contains three major functional categories, biological processes, molecular functions, and cellular components

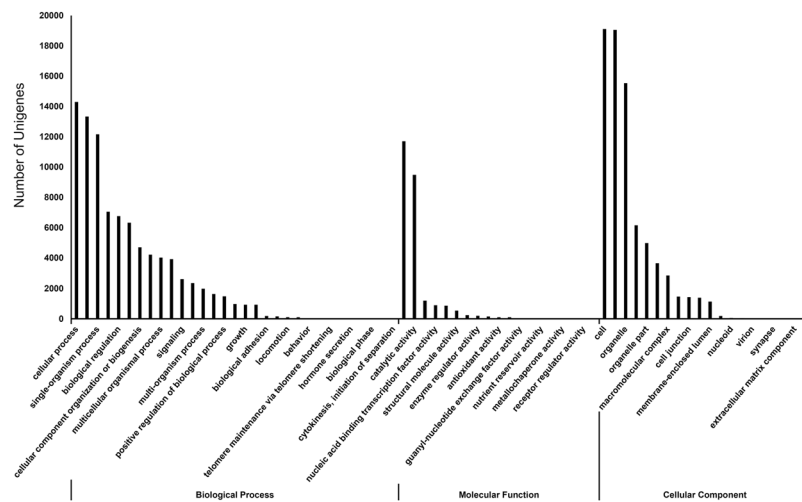


Figure 1. Gene Ontology (GO) classification of *Abies koreana* unigenes. A total of 42,056 unigenes were functionally classified into three main functional categories: biological processes, molecular functions, and cellular components.

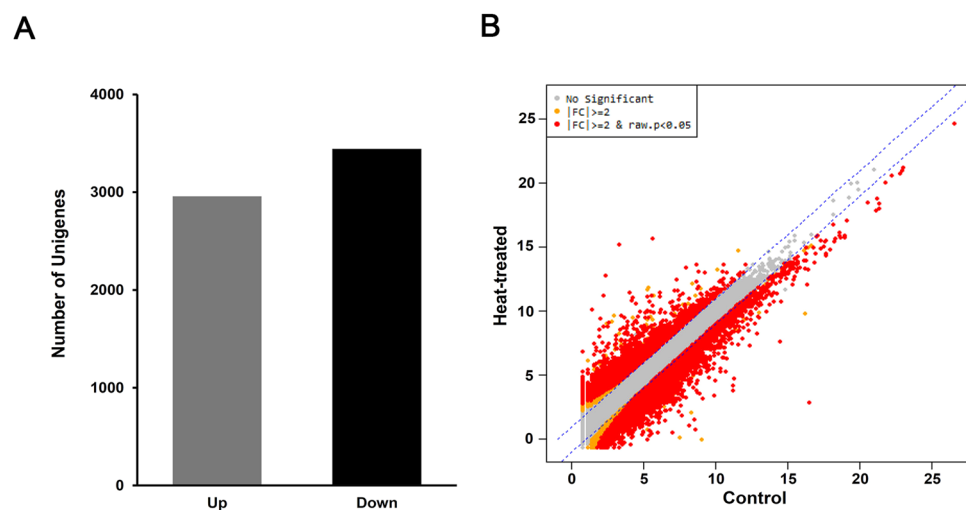


Figure 2. Distribution of differentially expressed *Abies koreana* unigenes in heat-treated samples compared with control conditions. (A) Distributions of up-regulated and down-regulated unigenes. The up-regulated and down-regulated unigenes indicate two fold change ($FC \geq 2$) and P value < 0.05 in comparison with under control conditions. (B) Scatter plot of the normalised expression levels of all unigenes under control and heat-treated conditions. Each point represents the mean expression level of a gene under control and heat-treated conditions. Red dots represent differentially expressed genes between the conditions.

(Supplemental Table S1 and Fig. 1). For the category of biological process, the most abundant groups were “cellular process” (14,297 genes), “metabolic process” (13,336 genes), “single-organism process” (12,160 genes), “response to stimulus” (7,053 genes), and “biological regulation” (6,758 genes). In the molecular function category, composed of 19 functional groups, “binding” (11,702 genes) and “catalytic activity” (9,483 genes) were the most highly represented groups. In cellular component, “cell” (19,112 genes) “cell part” (19,056 genes) and “organelle” (15,537 genes) were the most represented groups.

Differentially expressed genes (DEGs) involved in the heat-stress responses of Korean fir. To identify potential heat-stress-responsive genes in Korean fir, the gene expression profiles were compared between control and heat-treated samples. For each gene of the assembly, the number of mapped reads was compared between the control and the heat-treated samples (Fig. 2). As a result, 6,401 were found to be DEGs (fold change ≥ 2 , $p < 0.05$), with 2,958 up-regulated genes and 3,443 down-regulated genes in heat-treated sample compared (Supplemental Table S2). The distribution of transcript changes is shown in Fig. 2.

To investigate the biological roles of genes regulated by heat stress in Korean fir, we identified DEGs among the enriched GO terms, which were separated into the three main categories, biological processes, molecular

Functional groups	GO-id	Unigene number
Biological process		
cellular process	GO:0009987	2578
metabolic process	GO:0008152	2424
single-organism process	GO:0044699	2294
response to stimulus	GO:0050896	1305
biological regulation	GO:0065007	1212
regulation of biological process	GO:0050789	1121
localization	GO:0051179	782
cellular component organization or biogenesis	GO:0071840	747
developmental process	GO:0032502	676
multicellular organismal process	GO:0032501	659
signaling	GO:0023052	500
Molecular functions		
binding	GO:0005488	2102
catalytic activity	GO:0003824	1815
Cellular components		
cell	GO:0005623	3215
cell part	GO:0044464	3199
organelle	GO:0043226	2435
membrane	GO:0016020	1244
organelle part	GO:0044422	847
membrane part	GO:0044425	760
macromolecular complex	GO:0032991	414

Table 3. Top 20 most enriched functional groups in the gene ontology categories.

functions, and cellular components (Supplemental Table S3). The top 20 most enriched functional groups are shown in Table 3. Among these, 11 functional groups (55%), including “cellular process”, “metabolic process”, “single-organism process”, “response to stimulus”, “biological regulation”, and “regulation of biological process”, were significantly enriched within the biological processes category. Two functional groups (10%), including “binding” and “catalytic activity” were significantly enriched within the molecular functions category, and seven functional groups (35%) “cell”, “cell part”, “organelle”, “membrane”, “organelle part”, “membrane part” and “macromolecular complex”, were significantly enriched within the cellular component category. Thus, changes in the biological processes category appear to be of great significance in the response to heat stress in Korean fir.

Identification of transcription factors (TFs) involved in heat stress. The transcriptional regulation of heat stress has been widely documented in model plants²¹. To identify the TFs involved in heat-stress responses, we surveyed the putative TFs that were differentially expressed in Korean fir under heat stress. The TFs in this study were compared with *P. abies* transcriptome sequences obtained from publicly available datasets (E-value < 1e⁻¹⁰). A total of 204 TFs (81 up-regulated and 123 down-regulated) showed more than twofold stronger signals in the heat treated sample than in the control (Supplemental Table S4 and Fig. 3). Most abundantly represented, we identified the gene family involved in the basic helix-loop-helix family (bHLH), followed by ethylene-responsive element-binding factors (ERF), MYB, C2H2 family, and the NAC family. Of these TF families, bHLH, including 24 unigenes (5 up- and 19 down-regulated), ERF, including 22 unigenes (14 up- and 8 down-regulated), and MYB, including 18 unigenes (7 up- and 11 down-regulated), were the three most enriched TF families. Almost of the NAC TF family unigenes (12 of 14) were up-regulated under heat-treated conditions (Fig. 3). Taken together, the analysis provided a profound illustration of the roles of TFs under heat stress.

Identification of heat shock proteins (Hsps). To begin to elucidate the molecular basis of heat-stress tolerance in Korean fir, we sought to identify sequences in the transcriptome that encoded authentic Hsps. Based on sequence conservation (E-value < 1e⁻¹⁰), we identified 189 putative Hsp genes (Supplemental Table S5). Most of these Hsps were significantly up-regulated during the heat treatment (Supplemental Table S5). Of these unigenes, Trans Decoder identified 30 complete open reading frames with putative start and stop codons (Table 4).

Validation of DEGs using qRT-PCR. To confirm the accuracy of the RNA-seq results, 12 DEGs were randomly selected for a qRT-PCR-based comparison of their expression levels between the control and heat-treated samples (Fig. 4). The primer sequences are listed in Supplemental Table S6. All 12 DEGs in the control and heat-treated samples showed the same expression patterns in the qRT-PCR (Fig. 4). The unigenes included six putative heat-related TFs. The heat treatment up-regulated c173884_g1_i1 (bHLH), c194935_g1_i1 (ERF), c151473_g1_i1 (MYB), and c189572_g2_i1 (NAC) and down-regulated c96987_g1_i1 (bHLH), and c85122_g1_i1 (MYB) (Fig. 4A). The remaining six unigenes encoded Hsps. The expression levels of c212080_g4_i1

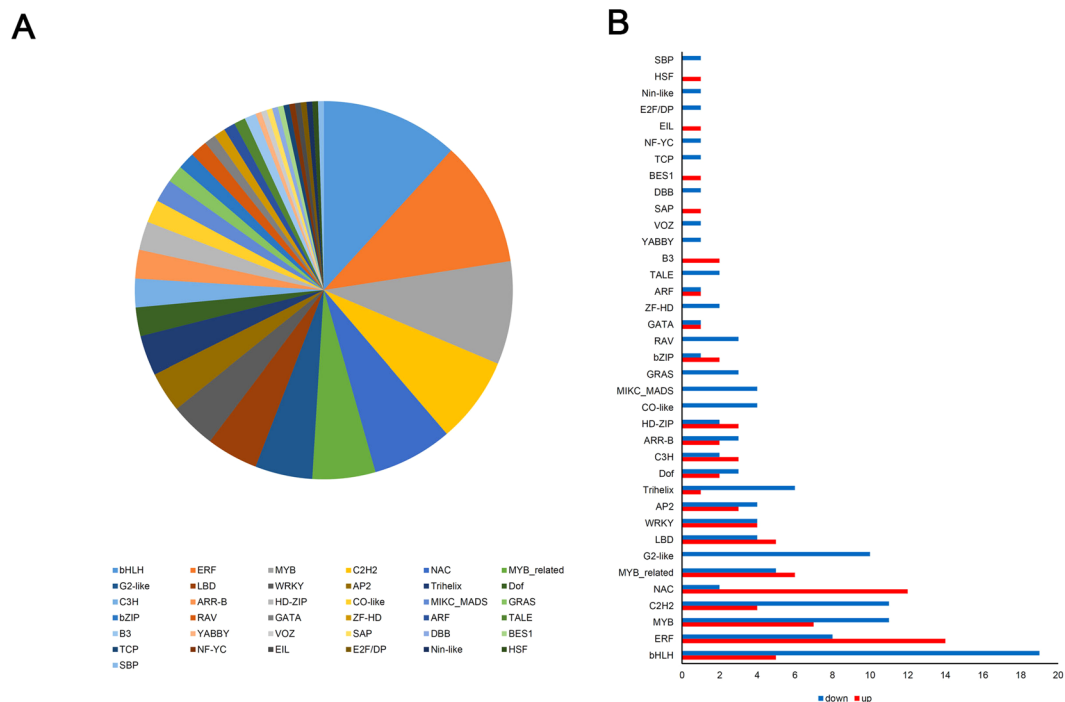


Figure 3. Family distribution of the transcription factors in the Korean fir transcriptome. (A) The number of each transcription factor family's member. (B) Up- or down-regulated transcripts from every transcription factor family involved in transcription.

(Hsp90), c149565_g1_i1 (Hsp70), c199303_g3_i1 (Hsp60), and c207145_g1_i1 (sHsp) were up-regulated by heat-treatment (Fig. 4B), while the expression levels of c149639_g1_i1 (Hsp70), and c178445_g1_i1 (Hsp60) were down-regulated by heat-treatment (Fig. 4B). This independent evaluation confirmed the reliability of the RNA-seq data. Also, this strongly supported the involvement of these 12 unigenes in responses to heat.

Furthermore, eight up-regulated genes involved in heat tolerance were further investigated by qRT-PCR at different time points to confirm the expression profiles under heat stress in Korean fir (Fig. 5). Transcription factors ERF had the highest expression levels at 7-day, and bHLH, MYB and NAC had their highest expression levels at 21-day (Fig. 5A). The three Hsp genes (c212080_g4_i1, c149565_g1_i1 and c207145_g1_i1) showed similar fold changes at all of the time points (Fig. 5B). In most cases, the expression changes were significant after heat treatments (Fig. 5). Thus, these genes might be important in the response to heat stress in Korean fir.

Discussion

In the absence of a whole genome sequence, RNA-seq provides an effective application tool for comprehensive studies of gene expression and the detection of novel transcripts associated with valuable traits²². In this study, we implemented a *de novo* RNA-seq technology to obtain insights into the transcriptomic responses induced by heat stress in Korean fir.

A whole-transcriptome analysis was performed in both heat-stressed and unstressed plants. For each sample, more than 160 M high-quality clean reads were obtained, which were *de novo* assembled into 406,207 contigs with an average length of 472.74 bp (Tables 1 and 2). These sequences produced longer contigs than those assembled in previous studies, such as blueberry (227 bp)²³, *Gossypium aridum* (229 bp)²⁴, *Hevea brasiliensis* (141 bp)²⁵, *Paeonia ostii* (307 bp)²⁶, pakchoi (343 bp)²⁷ and sweet potato (202 bp)²⁸.

Functional annotation and classification provide predicted information on cellular metabolic pathways and the condition-specific behaviour of gene expression. GO is an internationally standardised gene functional classification system that offers a dynamically-updated controlled vocabulary and a strictly defined structure to describe the properties of genes and their products in any organism²⁹. Among the unigenes, 22,880 (54.40%) known proteins were assigned to GO classes. However, a large proportion of unigenes (45.60%) failed to match these databases owing to the paucity of gene information for genus *Abies*. According to the GO classification, “cellular process”, “binding”, and “cell” comprised the largest groups in the three main GO categories of biological processes, cellular components, and molecular functions, respectively (Fig. 1). Our GO classifications of the annotated unigenes provide a general gene expression profile signature for Korean fir (*A. koreana*) that will facilitate further studies not only in this species but also in other species of genus *Abies*.

Subsequently, we analysed the differentially expressed genes in the heat-treated and control samples. Under heat stress, the GO category of biological processes was enriched (Table 3). The large proportion of the terms were included the “cellular process”, “metabolic process”, and “single-organism process” (Table 3), indicating that wide-ranging changes in Korean fir gene expression levels occurred in response to heat treatment. These findings indicated significant changes in category of biological process as a response to heat stress. Additionally, many

Classification	Contigs	Annotation	Fold change
Hsp90	c212080_g4_i1	HSP90_0281 Vitis vinifera	4.66
	c188934_g1_i1	HSP90_0278 Vitis vinifera	3.82
	c273438_g1_i1	HSP90_0266 Vitis vinifera	3.25
	c218689_g1_i1	HSP90_0266 Vitis vinifera	2.18
	c218750_g1_i1	HSP90_0266 Vitis vinifera	2.17
	c218389_g1_i1	HSP90_0266 Vitis vinifera	2.10
Hsp70	c151899_g1_i2	HSP70_1149 Vitis vinifera	7.20
	c149565_g1_i1	HSP70_1146 Vigna radiata	5.80
	c210065_g1_i1	HSP70_1095 Solanum lycopersicum	3.96
	c201565_g1_i2	HSP70_1078 Ricinus communis	3.48
	c194240_g1_i1	HSP70_1154 Vitis vinifera	2.70
	c207016_g1_i1	HSP70_0928 Arabidopsis thaliana	2.61
	c203374_g1_i1	HSP70_1077 Ricinus communis	2.29
	c149639_g1_i1	HSP70_1102 Sorghum bicolor	-5.06
Hsp60	c199303_g3_i1	HSP60_1249 Vitis vinifera PN40024	18.04
	c208087_g1_i3	HSP60_1095 Oryza sativa Indica Group	7.49
	c195910_g3_i3	HSP60_1095 Oryza sativa Indica Group	7.39
	c40393_g1_i1	HSP60_1095 Oryza sativa Indica Group	6.19
	c195910_g4_i1	HSP60_1095 Oryza sativa Indica Group	6.10
	c208445_g2_i2	HSP60_1095 Oryza sativa Indica Group	5.46
	c214017_g2_i1	HSP60_1044 Arabidopsis thaliana	4.41
	c196455_g1_i1	HSP60_1171 Ricinus communis	4.30
	c208445_g3_i1	HSP60_1095 Oryza sativa Indica Group	2.37
	c175049_g1_i1	HSP60_1249 Vitis vinifera PN40024	-2.94
	c194574_g1_i2	HSP60_1095 Oryza sativa Indica Group	-3.28
	c195155_g1_i1	HSP60_1249 Vitis vinifera PN40024	-3.74
	c178445_g1_i1	HSP60_1249 Vitis vinifera PN40024	-4.30
	sHsp	c207145_g1_i1	sHsp_0659 Physcomitrella patens subsp. patens
c201988_g1_i1		sHsp_0819 Vitis vinifera PN40024	4.85
c213666_g1_i3		sHsp_0673 Ricinus communis	3.17
Average expression levels			3.39

Table 4. Differentially expressed *Abies koreana* transcripts identified as heat shock protein (Hsp) families.

unigenes were over-represented as belonging to the category “response to stimulus” in the heat-treated sample (Table 3), and these unigenes represented the most important components directly involved in protecting plants from stress.

Apart from the list of unigenes that strongly responded to the changing environment (Fig. 2 and Supplemental Table S2), we focused on TFs, sequence-specific DNA-binding proteins that interact with *cis*-elements in the promoter regions of target genes and thus modulate gene expression. These TFs regulate gene transcription in response to biotic and abiotic stresses, such as cold, high temperatures, high salinity, drought, and pathogen attacks^{7,10}. In our results, several TF families were identified as being involved in heat-stress responses, including bHLH, ERF, MYB, NAC, and C2H2 (Fig. 3).

The highest number of up-regulated genes are ERF family TFs, and an ERF coactivator gene is synergistically expressed with ERFs under heat stress³⁰. The expressions of *AtERF53* and *ERF1* are induced by heat treatment in Arabidopsis and pakchoi, respectively^{27,31}. The DREB2's TF group belongs to the AP2/ERF family, and it has been characterised in the heat regulatory pathway³². The induced DREB2 functions to enhance heat tolerance in various plants^{33,34}. Other TFs, including bHLH, MYB, and C2H2 families, were also altered during heat treatments, and many members of these gene families function in heat tolerance^{35,36}. The *ERF*, *bHLH*, *MYB*, and *C2H2* pathways are conserved in Korean fir's responses to heat stress. The plant-specific NAC TF family has been implicated in the regulation of diverse processes, including hormone signalling, defence, and stress tolerance. NAC TFs in plants are mainly involved in osmotic stresses, including drought and high salinity³⁷. However, some NACs (RD26) function in response to cold stress³⁸. Morishita *et al.*³⁹ also reported that ANAC078 in the NAC group TIP is responsive to a combination of high light and heat stress. We found 14 genes encoding NAC TF domains, and almost the genes were up-regulated and showed significant expression levels by RNA-seq and qRT-PCR (Table S4, Figs 4 and 5). The results can provide insights that will explain critical functions of the NAC family of genes in the heat responses of Korean fir.

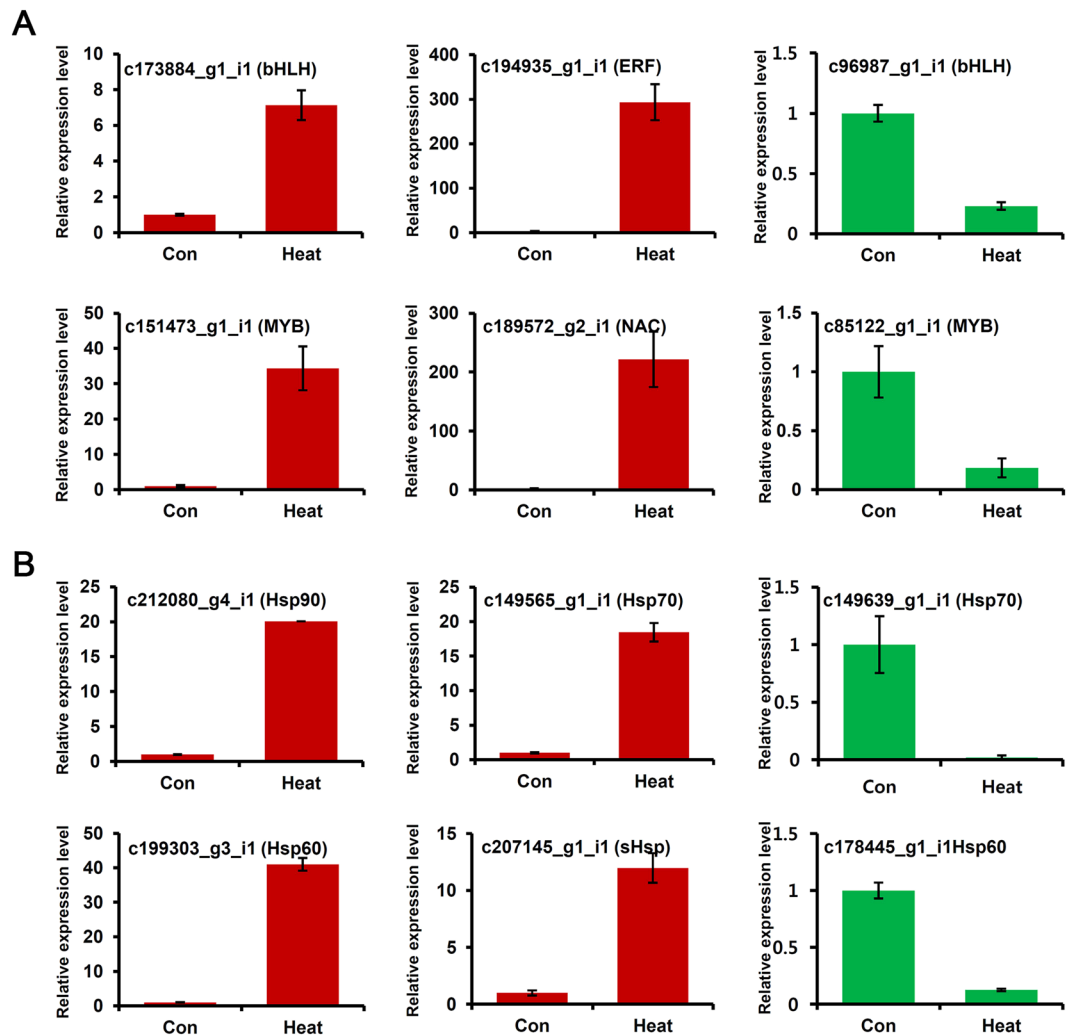


Figure 4. qRT-PCR expression analyses of 6 transcription factor and 6 heat shock proteins in response to heat stress. qRT-PCR was performed to validate the results of the RNA sequencing analysis using cDNAs prepared from 3-year-old needles of Korean fir exposed for 21-day to control (22 °C) or heat-treatment (30 °C) conditions. Error bars denote standard errors of biological replicates. Expression values of each gene are normalised against the expression of *Actin*¹⁸.

Heat shock factor (HSF) TFs are key regulators involved in responses to heat stress⁴⁰. Most plant HSFs are, mostly induced, regulated by heat stress. Among 23 rice *OsHsf* genes, 16 *OsHsfs* were up-regulated, however, *OsHsfC1a* was noted to be down-regulated by heat stress⁴¹. Similarly, many HSF genes from different plant species, such as Arabidopsis, cotton, maize, apple and pepper showed up- and down-regulation by heat stress⁴². However, we found only one transcriptional HSF, which was up-regulated in our data set (Fig. 3 and Supplemental Table S4). The small number of expressed HSFs (Fig. 3) might indicate that the heat-response pathway responded, at least partially, in an HSF-independent manner in the signalling network in Korean fir. Future studies should reveal such a manner of regulation.

In addition, several seemingly novel TF families responding to heat shock (ARR-B, AP2, C3H, and G2-like; Fig. 3) were also identified. Homologs in other plant species have not yet been reported in the response to heat stress, suggesting that the behaviour of these genes might be *Abies* species specific, making them attractive targets for further functional characterisation. The findings should expedite studies focusing on the interactions of different TFs in the regulation of heat stress. Thus, our analysis combines considerably conserved responses in addition to seemingly novel, possibly conifer specific, components involved in the heat-stress response mechanisms across plant species.

The analyses of transcriptome profiles in plants after heat treatment have amply documented the central role of the various HSP families in heat-stress responses^{30,43,44}. Hsp families, including Hsp100, Hsp90, Hsp70, Hsp60, and small Hsps, are involved in folding and assembling proteins, maintaining protein stabilisation, activating proteins, and degrading proteins in many normal cellular processes and under stress conditions⁴⁵. In our study, the expression levels of most Hsp genes in Korean fir were up-regulated after heat stress (Table 4, Figs 4 and 5). Thus, inductions of Hsps are critical for acclimating to heat stress in this species.

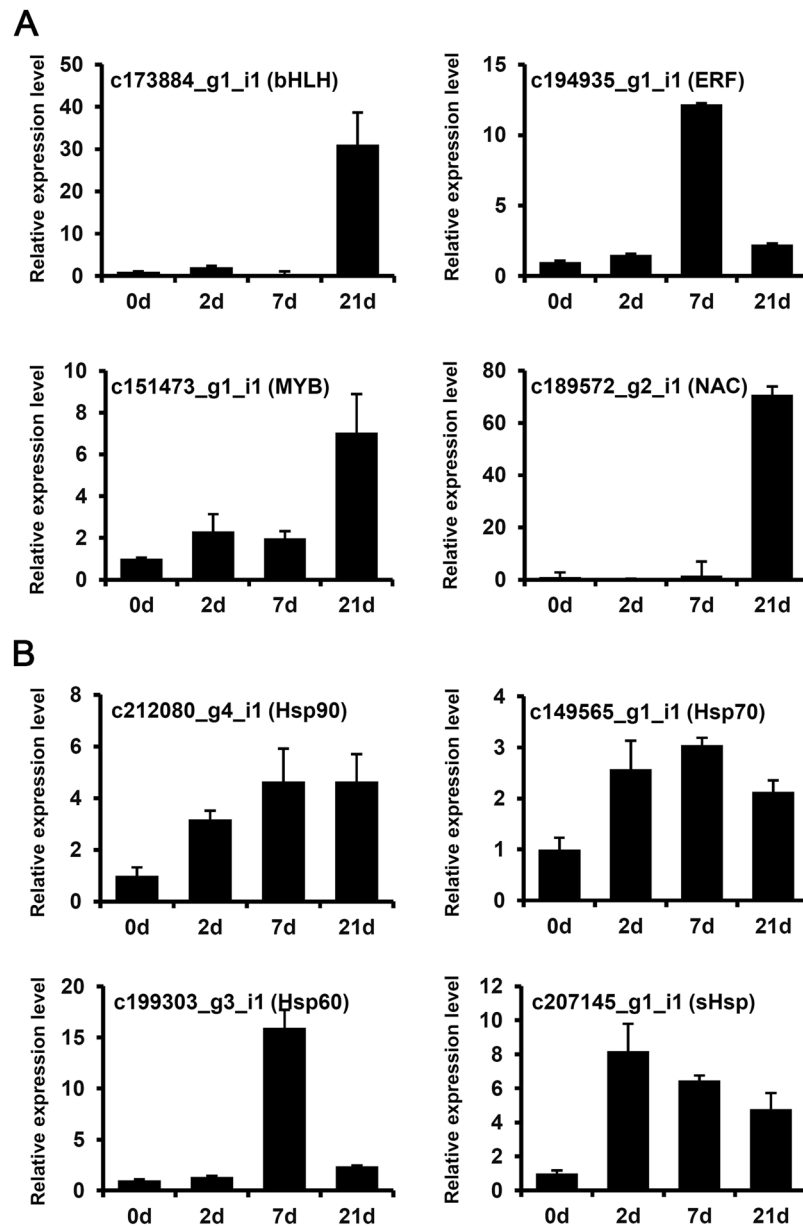


Figure 5. qRT-PCR expression analysis of 8 up-regulated genes at different time point in response to heat stress. Relative expression levels of 8 genes at 0, 2, 7 and 21-day after heat treatment (30°C) were examined by qRT-PCR. Error bars denote standard errors of biological replicates. Expression values of each gene are normalised against the expression of *Actin*¹⁸.

This first comprehensive transcriptomic analysis of Korean fir provides a valuable genomic resource for further studies, including as well in other *Abies* species. Additionally, our study provides new insights into heat-stress adaptation; it will facilitate further studies on Korean fir genes, their regulation and their functions. Our study represents a fully characterised transcriptome and provides valuable resources for genomic studies in Korean fir under heat stress and beyond.

Methods

Plant material and treatments. Korean fir (*Abies koreana* Wilson) seeds were collected from Mount Halla on Jeju Island, Korea (33° 13–36′ N, 126° 12–57′ E). Seeds were sown in seedling trays with soil after breaking dormancy at 4°C for three months. A single 1-year-old seedling was transplanted into each pot filled with the same soil. Plants were grown in a greenhouse under natural sunlight conditions.

The 3-year-old seedlings were exposed to normal growth conditions (22°C) and heat stress (30°C) under 16 h of 100 $\mu\text{E s}^{-1} \text{m}^{-2}$ light and 8 h dark. For RNA-sequencing, all of the needles were harvested after a 21-day heat treatment at which time controls were harvested as well. Three biological replicates were harvested. For the

quantitative RT-PCR (qRT-PCR) at different time points, plants were harvested at 0, 2, 7 or 21-days after the treatment. All of the needles were frozen in liquid nitrogen and stored at -80°C for RNA extraction.

Library preparation and RNA-sequencing. RNA samples were extracted from the needles of 21-day heat-treated and control plants. Total RNA was isolated using TRIzol reagent according to the manufacturer's protocol (GibcoBRL, Cleveland, OH, USA). The RNA was analysed for quality and concentration using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). A total of $3\ \mu\text{g}$ of RNA for each sample was used in library construction with the Truseq Stranded Total RNA sample Preparation Kit (Illumina, Inc. San Diego, CA, USA) per the manufacturer's instructions. Briefly, mRNA was enriched using magnetic beads containing poly-T molecules. Following purification, the enriched mRNA was broken into small fragments. Random oligonucleotides and SuperScript II were used to synthesize the first-strand cDNA. The second-strand cDNA was subsequently synthesized using DNA Polymerase I and RNase H. Finally, end repair was carried out on these cDNA fragments, and they were ligated with Illumina adapters. Libraries were amplified using PCR according to Illumina guidelines. Libraries with insert sizes from 290 to 330 bp were constructed, and sequenced using the Illumina HiSeq. 2500 to generate 101-bp paired-end reads.

De novo transcriptome assembly and annotation. The raw reads were first cleaned by filtering out adaptor sequences and low-quantity reads using Trimmomatic (version 0.32). Trimmed reads were combined across all conditions and *de novo* assembled with Trinity software (version R20140717)⁴⁶ using default settings to build a suitable set of reference contigs. Then, CD-HIT was used for further clustering with a 200-bp sequence length and 90% similarity cut-off values to obtain non-redundant transcripts⁴⁷. High-quality reads were mapped back to the assembled transcriptome sequences for validation. Reads were aligned using Bowtie with default parameters⁴⁸. RSEM software was subsequently employed to calculate the abundance for each transcript, and filtered raw read count were retained for further processing⁴⁹. These sequences were named as unigenes. Gene functions were annotated using BLASTX (BLAST 2.6.0+) based on the NCBI non-redundant protein sequences, GO and KEGG with default parameters. A functional enrichment analysis of unigenes using the GO categories molecular functions, biological processes, and cellular components was performed based on the protein sequence similarity in the GO database²⁹. The number of genes assigned in each GO term was counted using WEGO website (<http://wego.genomics.org.cn/>).

Identification of differentially expressed genes (DEGs). We identified differentially expressed genes with DEseq⁵⁰ using per gene expected counts for each sample generated by RSEM⁴⁹. A twofold change ($\text{FC} \geq 2$) and P value < 0.05 were used to define the significant DEGs between treatment and control. GO enrichment analyses were performed using GO database and WEGO website.

Transcription factors (TFs) and heat shock proteins (Hsps) analysis. To identify the putative TFs and decipher their roles during heat-stress responses, plant-specific TFs were downloaded from the Plant Transcription Factor Database (<http://plntfdb.bio.uni-potsdam.de/v3.0/>). A BLASTX algorithm-based search was performed with DEG sequences using the E-value cut off $\leq 1e^{-10}$ and classified unigenes according to the gene family's information. Similarly, to identify canonical Hsps represented in our samples, DEG sequences were queried against a list of Hsp domain sequences from the HSRIP (<http://pdslab.biochem.iisc.ernet.in/hspir>) database. A stringent domain matching criteria of E-value $\leq 1e^{-10}$ was used to select Hsp domains. Following alignments, TransDecoder (<http://transdecoder.sourceforge.net/>) was used to predicate optimal open reading frame information.

Quantitative RT-PCR (qRT-PCR). To determine the reliability of the RNA-seq data, qRT-PCR was performed on the same RNA pools previously used for RNA-seq. Twelve DEGs, six from four groups having the largest TF transcript numbers and six DEGs from each group of Hsps, were randomly selected for qRT-PCR analysis. Furthermore, to verify the differential expression, qRT-PCR was performed on a new set of 3 replicates for each sample at different time points.

Total RNAs ($1\ \mu\text{g}$) of each sample were reverse transcribed using a Power cDNA Synthesis Kit (Intron Biotech Inc., Sungnam, Korea). The specific primers used for qRT-PCR are listed in Supplemental Table S1. qRT-PCR was carried out on a Bio-Rad CFX qRT-PCR detection system (Bio-Rad Laboratories Inc., CA, USA) using iQTM SYBR[®] Green supermix (Bio-Rad). The reaction was performed under the following conditions: 95°C for 10 min, followed by 45 cycles of 95°C for 10 s and 60°C for 30 s. Their relative expression levels were calculated using the relative $2^{-\Delta\Delta\text{Ct}}$ method⁵¹. The qRT-PCR reactions were repeated in three biological and three technical replications.

References

- Lee, T. B. Endemic plants and their distribution in Korea. *Bull. Nat Acad Sci.* **21**, 169–218 (1986).
- Woo, S. Y. Forest decline of the world: A linkage with air pollution and global warming. *Afr J Biotechnol.* **8**, 7409–7414 (2009).
- Hong, Y. P. *et al.* Genetic variation of nSSR markers in natural populations of *Abies koreana* and *Abies nephrolepis* in South Korea. *J Korean For Soc.* **100**, 577–584 (2011).
- Lee, B. Y. *et al.* Biological indicators to monitor responses against climate change in Korea. *Korean J Pl Taxon.* **40**, 202–207 (2010).
- Weston, D. J. & Bauerle, W. L. Inhibition and acclimation of C_3 photosynthesis to moderate heat: a perspective from thermally contrasting genotypes of *Acer rubrum* (red maple). *Tree Physiology* **27**, 1083–1092 (2007).
- Bigras, F. J. Selection of white spruce families in the context of climate change: heat tolerance. *Tree Physiology* **20**, 1227–1234 (2000).
- Wahid, A., Gelani, S., Ashraf, M. & Foolad, M. R. Heat tolerance in plants: An overview. *Environ Exp Bot* **61**, 199–223 (2007).
- Yamaguchi-Shinozaki, K. & Shinozaki, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu Rev Plant Biol.* **57**, 781–803 (2006).
- Agarwal, P. K. & Jha, B. Transcription factors in plants and ABA dependent and independent abiotic stress signalling. *Biologia Plantarum.* **54**, 201–212 (2010).

10. Lata, C. & Prasad, M. Role of DREBs in regulation of abiotic stress responses in plants. *J Exp Bot.* **62**, 4731–4748 (2011).
11. Kotak, S. *et al.* Complexity of the heat stress response in plants. *Curr Opin Plant Biol.* **10**, 310–316 (2007).
12. Craita, E. B. & Tom, G. Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front plant sci.* **4**, 273 (2013).
13. Filichkin, S. *et al.* A genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* **20**, 45–58 (2010).
14. He, F. *et al.* RNA-Seq analysis of rice roots reveals the involvement of post-transcriptional regulation in response to cadmium stress. *Front Plant Sci.* **6**, 1136 (2015).
15. Qui, Q. *et al.* Genome-scale transcriptome analysis of the desert poplar. *Populus euphratica*. *Tree Physiol.* **31**, 452–61 (2011).
16. Schuster, S. C. Next-generation sequencing transforms today's biology. *Nat Methods* **5**, 16–18 (2008).
17. Gibbons, J. G. *et al.* Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol Biol Evol.* **26**, 2731–2744 (2009).
18. Surget-Groba, Y. & Montoya-Burgos, J. I. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Res.* **20**, 1432–1440 (2010).
19. Chen, J. *et al.* Sequencing of the needle transcriptome from Norway spruce (*Picea abies* Karst L.) reveals lower substitution rates, but similar selective constraints in gymnosperms and angiosperms. *BMC Genomics.* **13**, 589 (2012).
20. Uddenberg, D. *et al.* Early cone setting in *Picea abies* acrocona is associated with increased transcriptional activity of a MADS box transcription factor. *Plant Physiol.* **161**, 813–23 (2013).
21. Ohama, N., Sato, H., Shinozaki, K. & Yamaguchi-Shinozaki, K. Transcriptional regulatory network of plant heat stress response. *Trends Plant Sci.* **22**, 53–65 (2017).
22. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* **10**, 57–63 (2009).
23. Li, X. *et al.* *De novo* sequencing and comparative analysis of the blueberry transcriptome to discover putative genes related to antioxidants. *Gene* **511**, 54–61 (2012).
24. Xu, P. *et al.* *De novo* transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress. *Gene* **525**, 26–34 (2013).
25. Xia, Z. *et al.* RNA-Seq analysis and *de novo* transcriptome assembly of *Hevea brasiliensis*. *Plant Mol Biol.* **77**, 299–308 (2011).
26. Wang, Y. *et al.* *De novo* transcriptome sequencing and discovery of genes related to copper tolerance in *Paeonia ostii*. *Gene* **576**, 126–135 (2016).
27. Xu, H. *et al.* *De novo* transcriptome sequencing of pakchoi (*Brassica rapa* L. *chinensis*) reveals the key genes related to the response of heat stress. *Acta. Physiol. Plant* **38**, 252 (2016).
28. Wang, Z. *et al.* *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics* **11**, 726 (2010).
29. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29 (2000).
30. Zhang, X. *et al.* Expression profile in rice panicle: insights into heat response mechanism at reproductive stage. *Plos one.* **7**, e49652 (2012).
31. Hsieh, E. J., Cheng, M. C. & Lin, T. P. Functional characterization of an abiotic stress-inducible transcription factor AtERF53 in *Arabidopsis thaliana*. *Plant Mol Biol.* **82**, 223–237 (2013).
32. Mizoi, J., Shinozaki, K. & Yamaguchi-Shinozaki, K. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim Biophys Acta* **1819**, 86–96 (2011).
33. Diaz-Martín, J., Almoquera, C., Prieto-Dapena, P., Espinosa, J. M. & Jordano, J. Functional interaction between two transcription factors involved in the developmental regulation of a small heat stress protein gene promoter. *Plant Physiol.* **139**, 1483–1494 (2005).
34. Sakuma, Y. *et al.* Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc Natl Acad Sci USA* **103**, 18822–18827 (2006).
35. Ambawat, S., Sharma, P., Yadav, N. R. & Yadav, R. C. MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants.* **19**, 307–321 (2013).
36. Wang, S., Wei, X. L., Cheng, L. J. & Tong, Z. K. Identification of a C2H2-type zinc finger gene family from *Eucalyptus grandis* and its response to various abiotic stresses. *Biol Plant.* **58**, 385–390 (2014).
37. Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K. & Yamaguchi-Shinozaki, K. NAC transcription factors in plant abiotic stress responses. *Biochim Biophys Acta.* **1819**, 97–103 (2012).
38. Fujita, M. *et al.* A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J.* **39**, 863–876 (2004).
39. Morishita, T. *et al.* Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant Cell Physiol.* **50**, 2210–2222 (2009).
40. Pirkkala, L., Nykänen, P. & Sistonen, L. Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J* **15**, 1118–1131 (2001).
41. 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.* **47**, 785–795 (2009).
42. Guo, M. *et al.* The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Front Plant Sci.* **7**, 114 (2016).
43. Qin, D. D. *et al.* Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array. *BMC Genomics.* **9**, 432 (2008).
44. Liu, G. T. *et al.* Transcriptomic analysis of grape (*Vitis vinifera* L.) leaves during and after recovery from heat stress. *BMC plant Biology.* **12**, 174 (2012).
45. Huang, B. & Xu, C. Identification and characterization of proteins associated with plant tolerance to heat stress. *J Integr Plant Biol.* **50**, 1230–1237 (2008).
46. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
47. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
48. Langmead, B. *et al.* Ultrafast and memoryefficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
49. Li, B. & Dewey, C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
50. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
51. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCt} method. *Methods* **25**, 402–408 (2001).

Acknowledgements

We thank Jin Ho Park and Haksob Shin for sample preparation. This research was supported by the National Institute of Ecology (NIE-C-2018-15).

Author Contributions

J.E.H. and H.C.P. designed the study and wrote the manuscript. J.E.H., Y.J.K., M.H.S. and H.C.P. performed the experiment, collected the samples, performed RNA extractions, and analyzed the data. H.J.H. prepared plant material and H.J.B. analysed the data and edited the manuscript. All authors have read and approved the manuscript for publication.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-28552-1>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018