OMICS A Journal of Integrative Biology
Volume 19, Number 10, 2015

DOI: 10.1089/omi.2015.0097

Harnessing Next Generation Sequencing in Climate Change: RNA-Seq Analysis of Heat Stress-Responsive Genes in Wheat (*Triticum aestivum* L.)

Ranjeet R. Kumar,[1,†] Suneha Goswami,[1,†] Sushil K. Sharma,[1] Yugal K. Kala,[2] Gyanendra K. Rai,[3] Dwijesh C. Mishra,[4] Monendra Grover,[4] Gyanendra P. Singh,[5] Himanshu Pathak,[6] Anil Rai,[4] Viswanathan Chinnusamy,[7] and Raj D. Rai[1]

Abstract

Wheat is a staple food worldwide and provides 40% of the calories in the diet. Climate change and global warming pose a threat to wheat production, however, and demand a deeper understanding of how heat stress might impact wheat production and wheat biology. However, it is difficult to identify novel heat stress associated genes when the genomic information is not available. Wheat has a very large and complex genome that is about 37 times the size of the rice genome. The present study sequenced the whole transcriptome of the wheat *cv*. HD2329 at the flowering stage, under control (22° ± 3°C) and heat stress (42°C, 2 h) conditions using Illumina HiSeq and Roche GS-FLX 454 platforms. We assembled more than 26.3 and 25.6 million high-quality reads from the control and HS-treated tissues transcriptome sequences respectively. About 76,556 (control) and 54,033 (HS-treated) contigs were assembled and annotated *de novo* using different assemblers and a total of 21,529 unigenes were obtained. Gene expression profile showed significant differential expression of 1525 transcripts under heat stress, of which 27 transcripts showed very high (>10) fold upregulation. Cellular processes such as metabolic processes, protein phosphorylation, oxidations-reductions, among others were highly influenced by heat stress. In summary, these observations significantly enrich the transcript dataset of wheat available on public domain and show a *de novo* approach to discover the heat-responsive transcripts of wheat, which can accelerate the progress of wheat stress-genomics as well as the course of wheat breeding programs in the era of climate change.

Introduction

Wheat is a staple food in most regions worldwide and provides 40% of the calories in the diet (Kumar et al., 2014a). The terminal heat stress (HS), especially during the reproductive stage, has detrimental effect on the growth and yield of wheat (Maestri et al., 2002); high temperature during the critical stages reduces the yield of wheat by as much as 15% (Wardlaw et al., 1994). Heat stress causes drying of fluid on stigmatic surface and defunct pollen, leading to 'pseudo-seed' setting problems in wheat (Kumar et al., 2013). High temperature, especially during flowering and grain-filling stages in wheat, reduces the activity of enzymes associated with photosynthesis and starch biosynthesis pathways, causing reduction in kernel weight and number (Hays et al., 2007; Plaut et al., 2004).

Plants exposed to the HS showed significant changes in cellular and metabolic processes such as denaturation/aggregation of enzymes with decreased activity, triggering of signaling molecules like protein kinases, and overexpression of heat shock proteins (HSPs) (Goswami et al., 2014; Kumar et al., 2012). Studies have demonstrated that the reprogramming of expression of genes and proteins under stresses are important to initiate a series of networking mechanisms involved in modulating the cellular and whole plant tolerance (Guy, 1999). Although the research on elucidating the mechanism of HS tolerance in plants has considerably progressed in the recent years in model plants, limited information is available on heat-responsive genes and proteins associated with tolerance of pathways such as carbon assimilation and starch biosynthesis in wheat (Kumar et al., 2013).

Divisions of [1]Biochemistry, [2]Genetics, [6]CESCRA, [7]Plant Physiology, Indian Agricultural Research Institute, New Delhi, India.
[3]Sher-e-Kashmir University of Agricultural Sciences and Technology, Jammu, India.
[4]Centre for Agricultural Bio-Informatics (CAB-in), Indian Agricultural Statistics Research Institute (IASRI), New Delhi, India.
[†]These authors contributed equally to this work.

Plant omics and new biotechnologies such as massively parallel sequencing and microarray analysis were preferred tools to identify and characterize the differentially expressed genes (DEGs) under HS in different modal species (Gong et al. 2014; Qian et al. 2014; Rabara et al. 2014), but Next-Generation Sequencing (NGS) is now the most preferred technique used for transcriptome study. The transcriptome data generated through NGS provide very useful information on the regulatory pathway networks operating in plants under different conditions (Rensink and Buell, 2005). NGS data help to identify DEGs, along with their functional annotation under different stresses, which has been used in the past to elucidate the different mechanisms as well as pathways associated with HS tolerance (Mochida et al., 2006).

RNA-Sequencing (RNA-Seq) is now the preferred technology used for global genome identification and *de novo* transcriptome of various organisms (Mousavi et al., 2014). In this technology, small stretches of cDNAs are sequenced at a very high coverage and assembled using different programs to reconstruct the contigs. It has been successfully used for the relative expression study, editing 5' and 3' ends of annotated genes and functional gene identification with their respective exons and introns (Nagalakshmi et al., 2010). In wheat, RNA-Seq has been mostly used for the identification of novel and conserved stress-responsive genes, especially associated with biotic stress tolerance and nutrient-responsive regulation (Kugler et al., 2013; Oono et al., 2013).

Wheat has a very large and complex genome ($\sim$16,000 Mbp), which is about 37 times the size of the rice genome (Arumuganathan and Earle, 1991). The Chromosome Survey Sequence (CSS) of the wheat genome has been recently released and published (Marcussen et al., 2014). Projects, however, are underway under the international consortium (International Wheat Genome Sequencing Consortium; www.wheatgenome.org) to sequence the whole wheat genome. 6474 Expressed Sequence Tags (ESTs) associated with HS tolerance have been deposited in the NCBI database for the ESTs (dbEST; http://www.ncbi.nlm.nih.gov/dbEST/). As of on February, 2015, the number of ESTs is quite less, as compared to the other agriculturally important crops.

Global transcriptome expression levels in model species are analyzed using the reference genome sequences available, which is not feasible in cases of non-model organisms lacking the genome information (Grabherr et al., 2011). *De novo* assembly of short sequences, however, may prove to be a powerful tool to reconstruct the entire transcriptome of non-model organism and profile the expression of genes at the temporal and spatial basis (Duan et al., 2012; Martin and Wang, 2011).

Here we have used combined NGS technologies (Illumina HiSeq and Roche GS-FLX 454) to pursue RNA-Seq analysis of global gene expression changes in a heat sensitive wheat *cv.* HD2329 exposed to HS during anthesis stage, and have generated a wheat transcript dataset for *de novo* assembly.

Materials and Methods

Plant material and stress treatment

HD2329, a popular wheat *cv.* (*T. aestivum* L.) but sensitive to HS was used in the present investigation. Pre-treated seeds (Bavistin @ 0.5%) were sown in six pots (having equal quantity of perlite farmyard manure mixture) and were kept inside the plant growth chamber (22° ± 3°C, relative humidity of 75% and 8 h light with intensity of 250 μmol m^{-1} s^{-1} Photosynthetically Active Radiation) in the National Phytotron Facility at Indian Agricultural Research Institute (IARI), New Delhi.

Irrigation was done at regular intervals. Plants (three pots) at the flowering stage (Feekes scale—10.53) (Large, 1954) were exposed to HS (42°C for 2 h), whereas three other pots served as control (22° ± 3°C). The HS was given in a sinusoidal mode using a microprocessor-regulated controller with an increase of 1°C/10 min until the temperature reached 42°C and was held at 42°C for 2 h. Thereafter, normalcy (22° ± 3°C) was brought in the same fashion. Samples (root, stem, flag leaf, and spike) were collected and frozen in liquid nitrogen for the transcriptome analysis. For the validation of randomly selected differentially expressed transcripts (DETs), samples were collected separately in triplicates after the HS-treatment and were stored at −80°C. Flow-chart of the work plan is depicted in the supplementary file (Supplementary Fig. S1; supplementary material is available online at www.liebertpub.com/omi).

RNA isolation and sequencing

Total RNA was isolated from 100 mg of each tissue (root, stem, flag leaf, and spike) separately using the Trizol method (Invitrogen, UK); DNase treatment was given to the isolated RNA and further purified using the RNeasy kit (Qiagen, Valencia, CA, USA). The quality of RNA was analyzed on 1% denaturing agarose gel. Total RNAs isolated from different tissues were pooled in equimolar concentration to make 60 μg of total RNA for one sample to represent the whole plant.

Equal quantities of total RNA were used for the cDNA library construction. The pair-end cDNA sequencing libraries were prepared using Illumina TruSeq RNA Library Preparation Kit (Illumina, San Diego, CA, USA) as per the protocol. In brief, pooled samples were subjected to mRNA fragmentation, followed by reverse transcription, second-strand synthesis, pair-end adapter ligation, and index PCR amplification. The quantitation of the Library was performed on the Agilent bioanalyzer (Agilent Technologies, USA) using high sensitivity DNA chip. The libraries of all samples were in the size range of 250 to 400 bp.

The DNA of each cluster was sequenced using the TruSeq Sequencing By Synthesis (SBS) v3-HS kits on the Illumina HiSeq 2000 platform. For Roche GS-FLX 454 platform, cDNA synthesis was performed using SMART II™ cDNA synthesis kit (Clontech Laboratories, Mountain View, CA). cDNA with length more than 100 bp was separated and eluted from 2% agarose gel; concentration was estimated using the Bioanalyzer 2100 (Agilent Technologies, Germany).

Further, Rapid Library preparation kit was used to process 5 μg of sheared cDNA sample for fragment end repair and adaptor ligation. Small fragment removal step was performed to remove the DNA fragments below 300 bp. A single-stranded library was quantitated on TBS 380 Fluorometer and a RL curve was prepared to generate the standard curve of fluorescence readings to be used for the calculation of the library sample concentration for preparation of emulsion PCR. The emulsion-based clonal amplification of cDNA library of all the samples was performed by following the

protocol of emPCR Method–Lib-L MV. Clonal amplified library beads obtained from emPCR amplification were deposited on Pico Titre Plate (PTP) for sequencing using the pyrosequencing chemistry on Roche GS-FLX 454 according to the manufacturer recommendations.

De novo *assembly and sequence clustering*

Raw data generated from Illumina HiSeq and Roche GS-FLX 454 pyrosequencing were pre-processed to remove a) adaptor-only reads (that were added for reverse transcription and 454 sequencing), b) primers, c) very short (<50 bp) sequences, and d) low-quality reads (reads mean Q-value 20) using the Newbler program v 2.5. The high quality reads for each sample were assembled using the assembly programs Cap3 (http://seq.cs.iastate.edu/cap3.html), MIRA v 3.2.1 (http://sourceforge.net/), GS *de novo* Assembler (Newbler v 2.5.3 (http://www.454.com/products-solutions/analysis-tools/gs-denovoassembler.asp), and Velvet-Oases assembly program v 0.2.04 (Zerbino and Birney, 2008) with optimal parameters.

To generate a nonredundant set of reference transcripts (Unigenes, Plant Unique Transcripts), we merged the assemblies using cd-hit-EST/cd-hit-454 (http://weizhongli-lab.org/cd-hit/) with 100% identity. TIGR gene indices clustering (TGICL) software (http://www.tigr.org/tdb/tgi/software/) was used to generate single contigs by combining overlapped contigs with minimum length of 50 bp and identity of 99% in order to remove redundancy (Pertea et al., 2003).

Sequence annotation

Assessment and annotation of unigenes were performed by sequence similarity search on five different databases: NCBI nonredundant (http://www.ncbi.nlm.nih.gov/refseq/), UniProtKB/Swiss-Prot (http://www.ebi.ac.uk/uniprot), Triticeae full-length CDS database (TriFLDB—http://trifldb.psc.riken.jp/v3/index.pl), PlantCyc (http://www.plantcyc.org/), and Pfam (http://pfam.xfam.org/) using Basic Local Alignment Search Tool (BLAST—http://blast.ncbi.nlm.nih.gov/Blast.cgi) with E-value < 1e-5. Gene function was assigned to each unigenes based on the best BLAST hit (lowest E-value). To increase computational speed, such search was limited to the first 10 significant hits for each query.

Calculation of RPKM value and clustering of differentially expressed unigenes

We quantified transcript levels in Read per Kilobase of Transcript per Million Reads Mapped (RPKM) value. The RPKM value was calculated using the equation

$$\mathrm{RPKM} = (10\text{^}9 \ ^{*} \ \mathrm{C})/(\mathrm{N} \ ^{*} \ \mathrm{L})$$

where C = Number of reads mapped to a gene, N = Total mapped reads in the experiment, and L = Exon length in base-pairs for a gene.

The RNA-Seq reads from each sample were aligned back to the unigenes, and the transcripts were assembled using Bowtie v 1.0.0 (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead et al., 2014), with the parameters '-k 2 –no-mixed -q'. Digital gene expression profiles of all unique transcripts mapped onto the unigenes were quantified using edgeR package (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html; Robinson et al., 2010). Each RPKM value was incremented by 1 to avoid division by 0 following the fold change calculations. Differential expression (DE) was detected using edgeR Bioconductor package with log fold change threshold of 2 and FDR value cut-off <0.01. A system of cluster analysis for expression data was performed that uses k-means clustering algorithm implemented in edgeR to arrange genes according to similarity in their pattern of expression.

Gene ontology (GO) and pathway analysis

The assembled sequences were annotated with Gene Ontology (GO) terms such as biological processes, molecular functions, and cellular components by importing the NCBI-nr BLAST results into Blast2GO software package (http://www.blast2go.com/b2ghome) that retrieves the GO terms, allowing gene functions to be determined and compared (Altschul, 1993). The unigenes sequences were also aligned to the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database using the online KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/kegg/kaas/). The bi-directional best hit (BBH) method (http://www.nmpdr.org/) was used to obtain KEGG Orthology (KO) system (http://www.genome.jp/kegg/ko.html).

GO pathway enrichment analysis

GO functional enrichment analysis was carried out to identify GO terms significantly enriched in DETs, as compared to the genome background. Pathway analysis was carried out by mapping the DETs to terms in the KEGG and MapMan database (Thimm et al., 2014). Using the Mercator sequence annotation pipeline produced by MapMan (mapman.gabipd.org), we assigned MapMan ''Bins'' to the identified DETs by searching six reference databases: three based on BLAST such as The Arabidopsis Information Resource (TAIR10), plant proteins from Swiss-Prot and UniRef90, two based on RPSBLAST such as Conserved Domain Database (CDD; http://www.ncbi.nlm.nih.gov/cdd/), EuKaryotic Orthologous Groups (KOG; http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/kog/), and InterProScan 5 (http://www.ebi.ac.uk/Tools/pfa/iprscan5/). The output was used directly as a mapping file for the high-throughput data visualization software MapMan.

Chromosomal localization of identified DETs

The identified DETs were mapped on to the recently available draft genome sequence of wheat downloaded from Ensembl Plants (ftp.ensemblgenomes.org/pub/release-24 / plants / fasta / triticum_aestivum / dna). We downloaded 742742 scaffolds with average size of 6021.66 including gaps with N50 value 6856 (Supplementary Table S1). The fasta sequences for the DETs were extracted and aligned against the wheat scaffolds using BLASTN search tool of NCBI (http://blast.ncbi.nlm.nih.gov/blast/).

Novel transcript discovery and gene prediction

The unigenes that were not annotated by GenBank (http://www.ncbi.nlm.nih.gov/genbank/) or Universal Protein Resource (Uniport; http://www.uniprot.org/) were classified as novel transcripts. However, the novel transcripts that showed

significant regulations were further used for the gene prediction. The gene prediction was performed using ab initio gene prediction methods (Issac et al., 2004) implemented in Augustus v 2.7 software (Stanke et al., 2008) (http://bioinf.uni-greifswald.de/augustus/datasets/), utilizing *A. thaliana* and *Z. mays* gene annotation as training sets.

Validation of identified novel transcripts

Designing of primers. The validation of identified transcripts (novel and conserved) was carried out using the quantitative real-time RT-PCR expression analysis. Gene specific forward and reverse primers were designed using Genefisher2 primer designing software (http://bibiserv.techfak.uni-bielefeld.de/genefisher2/). The quality of the primers was checked using OligoAnalyzer v 3.1 (IDT, USA; https://eu.idtdna.com/calc/analyzer). The list of primers used for the expression analysis is presented in Supplementary Table S2.

Quantitative real-time RT-PCR. Total RNA was isolated from the control and HS-treated samples using Trizol method (Invitrogen, UK) and further real-time RT-PCR was carried out following the protocol as mentioned in Kumar et al., (2014). All the expression levels were normalized to the arithmetic mean of the selected β-actin gene (accession no. AF282624). The relative expression was calculated using comparative C_t method (Pfaffl, 2001).

Results

Transcriptome sequencing and read assembly

cDNA libraries prepared from the total RNA of control and HS-treated pooled samples (root, stem, flag leaf, and spikes) of wheat *cv*. HD2329 were sequenced and a total of 63.8 million raw reads were generated. The raw reads were subjected to stringent quality assessment and data filtering (read mean quality score >=20), and 52.1 million high quality reads were further selected for the analysis (26.3 million high quality reads in control and 25.6 million reads in case of HS-treated); Information of the sequencing data is presented in Table 1.

The assembly of high-quality reads generated 76,556 and 54,033 contigs with mean transcript contig length of 544 bp and 518 bp for the control and HS-treated samples, respectively. The N50 value observed for the control and HS-treated samples were 678 and 640 bp. The merged contigs were subjected to cluster and assembly analysis. A total of 21,529 unigenes were obtained, among which 6,712 genes (31.1%) were greater than 1 kb (Fig. 1). The mean length of the unigenes identified was 886 bp with minimum length of 201 bp and maximum length of 6596 bp. The N50 value observed for the contig was 1021. The summary and length distributions of unigenes are shown in Table 2 and Figure 1. The raw datasets were deposited in the Short Read Archive (SRA) repository of National Centre for Biotechnology Information (NCBI) under the accession number PRJNA171754 (http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA171754).

Table 1. Summary of the Raw and Cleaned Reads Obtained from Control and HS-Treated Wheat cv. HD2329 by NGS

	Illumina HiSeq 2000		*Roche GS-FLX 454*	
Samples	*Raw reads*	*Clean reads*	*Raw reads*	*Clean reads*
HD2329 control	31,495,619	26,265,216	138,649	137,985
HD2329 treated	32,078,392	25,552,763	150,845	148,975

Functional annotation and classification of transcriptome

Functional annotation of wheat RNA-Seq data is an arduous task because of limited information available on the reference genome/gene sequences in the public databases. However, to maximize the annotation percentages, we have mined five different databases: NCBI, UniProt: Swiss-Prot, TriFLDB, PlantCyc, and Pfam with an E-value <= 1e-5. Out of 21,529 unigenes identified, 16,531 (76.7%) were annotated by NCBI-nr database. Uniprot BLAST analysis resulted in 10,816 (50.2%) annotations. Some of the unigenes associated with thermotolerance and identified using the NCBI-nr database are, for example, 17.6 kDa HSP (CL1Contig9), WRKY4 transcription factor (CL1Contig22), HSP90 (CL20Contig3), HSP70 (CL283Contig1), and Rubisco activase A (CL804Contig1). A list of unigenes identified using NCBI-nr database is in Supplementary Table S3.

Similarly, fkb70_wheat 70 kDa peptidyl-prolyl isomerase (CL1Contig40), HS17A_ORYSJ kDa class I heat shock protein (CL17Contig1), and HSP83_ORYSJ heat shock protein 81-3 (CL20Contig1), associated with HS-tolerance, were identified using UniProt: Swiss-Prot database (Supplementary Table S4). Similarly, using the TriFLD, we identified SOD1.2 (CL1Contig15), WRKY71 (CL1Contig22), and TaHSP70d (CL16Contig1), to be associated with HS-tolerance (Supplementary Table S5). Annotation of unigenes using PlantCyc database identified the presence of SOD (CL1Contig29), NADH: ubiquinone reductase (CL1Contig34), and phosphoglycerate kinase (CL2Contig4) (Supplementary Table S6).

We identified 12 unigenes to have at least one enzyme hit and 3,177 unigenes to have at least one GO hit (Supplementary Fig. S2). Pfam database mining showed 1693 sequences to have at least one domain with domain coverage >50%; 1328 unigenes showed hit with the entire Enzyme, Domain and GO databases (Supplementary Fig. S2). Of the 21,529 unigenes, 12,962 (60.2%) and 8,809 (40.9%) could be aligned to the full length wheat cDNAs and plant metabolic pathway genes, respectively.

Unigenes identified were subjected to Gene Ontology (GO) analysis for their categorization under biological, molecular and cellular functions (Fig. 2). Under the category of biological processes, most of the identified unigenes were predicted to be associated with metabolic process, cellular, response to stimulus, and biological regulation (Fig. 2). Similarly, under molecular category, most of the unigenes were involved in binding followed by catalytic activity, transporter activity, and structural molecular activities.

Identification of DETs under the HS

The uniquely mapped reads observed were used for the expression analysis of each sample using empirical analysis

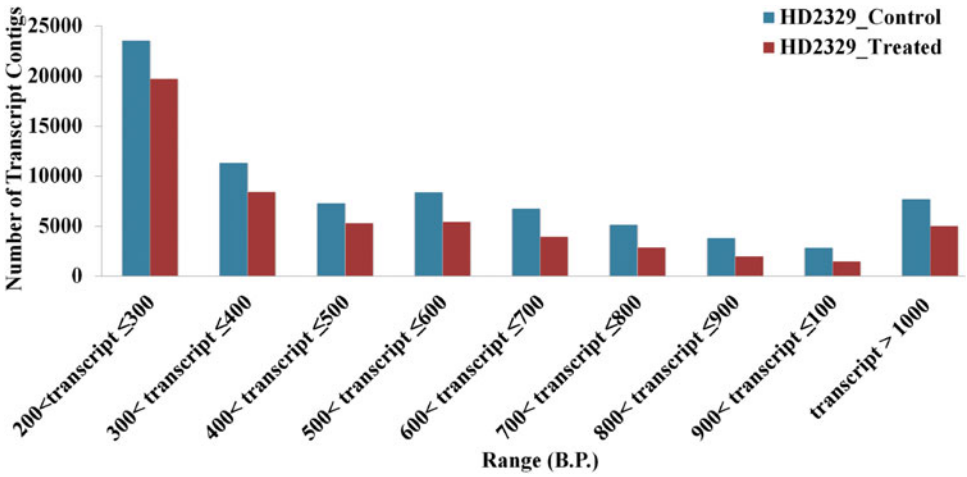

FIG. 1. Distribution of identified unigenes length from wheat *cv*. HD2329 under heat stress. Unigenes were generated by different assemblers such as Cap3, MIRA v 3.2.1, Newbler v 2.5.3, and Velvet-Oases v 0.2.04.

of digital gene expression (DGE) data in R (edgeR) (Mortazavi et al., 2008). Average RPKM values were 34.45 and 49.21 in control and HS-treated samples, respectively. The maximum RPKM values reported for control and HS-treated samples were 9,235.41 and 45,106.53, respectively. The RPKM values of 1,355 and 1,331 unigenes were greater than 100 for control and HS-treated samples, suggesting higher number of modulated genes under the HS.

In order to have a quick overview of expression data, MA plot (Dudoit et al., 2002) was drawn to visualize the expression pattern of the transcripts (Fig. 3a). We observed numerous variable points on the MA plot, as seen by the spread-up of points around the plots which shows the difference between the expression of control and HS-treated samples. We observed 716 transcripts to be upregulated and 809 transcripts to be downregulated in HS-treated sample, as compared to nonstressed sample (Fig. 3b). A list of DETs (upregulated and downregulated) has been presented in Supplementary Table S7. Some of the transcripts that showed more than 10-fold upregulation are bag family molecular chaperone regulator 6-like, HSF a-6a-like, and CDPK 30. Twenty seven transcripts showed >10-fold downregulation, such as pollen-specific protein SF3-like and major allergen Phl p5 (Table 3). The variation in the expression level of DETs was shown in the heat map with red (upregulation) and green (downregulation) colors (Fig. 4).

Table 2. De Novo Assembly Statistics for Whole Transcriptome Analysis of Wheat cv. HD2329 (Control and HS-Treated) Performed on ILLUMINA HISEQ 2000 and Roche GS-FLX 454 Platforms

Total number of unigenes	21,529
Total base pairs (bp)	19,073,024
Number of unigenes with >1 kb length	6,712
Mean length of unigenes (bp)	886
Minimum length of unigenes (bp)	201
Maximum length of unigenes (bp)	6,596
Median length of unigenes (bp)	781
N50 value	1,021
GC (%)	48.58

Functional classification of DETs

DETs were grouped into different functional categories based on their functions identified by BLAST: metabolic pathways, protein phosphorylation, oxidation-reduction process, and response to heat and cold. The largest group of genes with known function was associated with protein phosphorylation, followed by oxidation-reduction process, response to stress, regulation of DNA transcription, and response to heat stress (Fig. 5). Similarly, under molecular function, most of the DEGs were associated with ATP binding, followed by serine threonine kinase activity, zinc ion binding, and metal ion binding.

DETs mapped on to KEGG database using BLAST showed pathways such as metabolic, and biosynthesis of secondary metabolites to be highly influenced by HS. Most of the upregulated transcripts were associated with metabolic pathways, biosynthesis of secondary metabolites, and oxidative phosphorylation, whereas downregulated transcripts were associated with starch and sucrose metabolism, ribosome etc. (Supplementary Fig. S3).

MapMan enrichment plot for DETs

De novo transcriptome analysis revealed a significant fold change in the expression of 1525 genes (716 upregulated and 809 downregulated) in HS-treated samples, as compared to control. To study the effect of HS on different metabolic pathways, significantly altered genes ($p < 0.05$) were mapped on the pathway bins of MapMan 3.5.1R2 (http://mapman.gabipd.org). The functions and trend in MapMan were consistent with the GO enrichment results. The HS treatment significantly repressed the expression of genes associated with different pathways, such as (i) genes associated with protein synthesis; (ii) genes associated with amino acid metabolism; (iii) genes encoding enzymes associated with nitrogen, sulfur, and nucleotide metabolism. The HS treatment was observed to have mixed effects (induction/repression) on

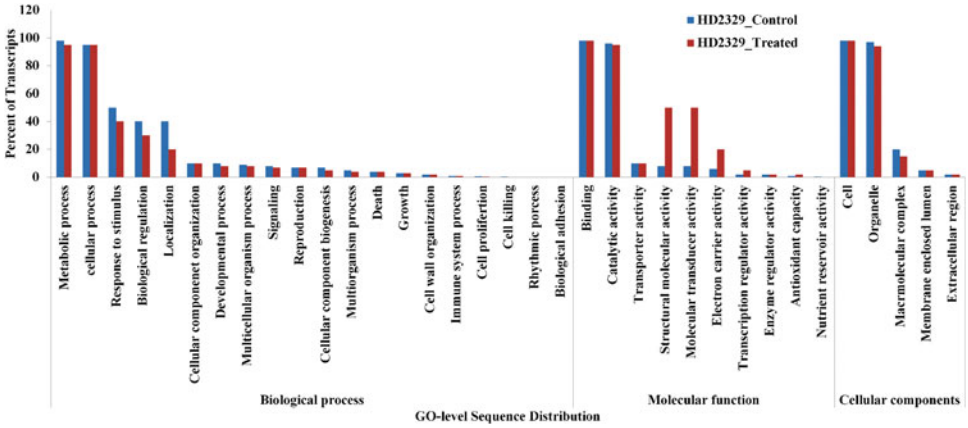

FIG. 2. Distribution of identified unigenes from wheat using Gene Ontology analysis. Biological process, molecular function, and cellular categories were identified using the GO database (http://www.geneontology.org/).

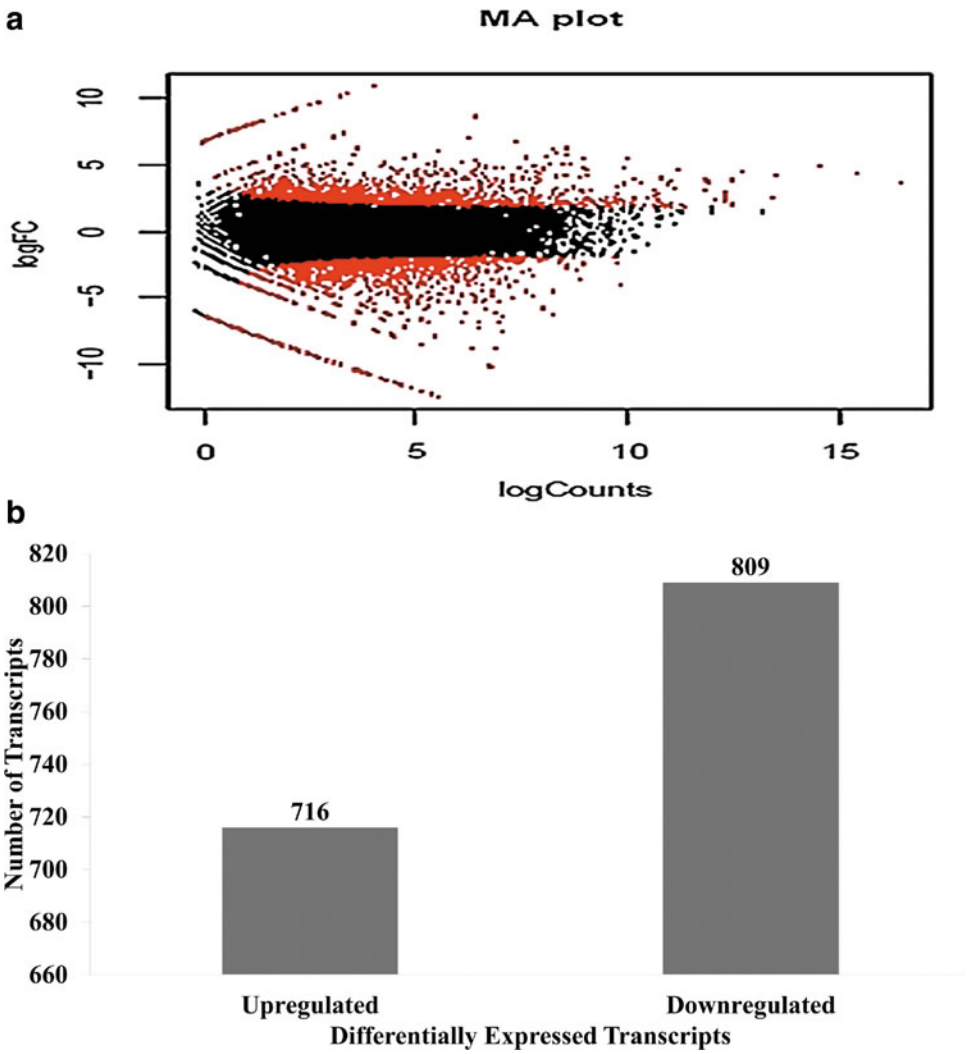

FIG. 3. Distribution of unigenes from wheat based on HS-responsive expression pattern. **(a)** MA plot showing the expression pattern of the transcripts. **(b)** Upregulated and downregulated transcripts observed in wheat *cv*. HD2329 with log fold value threshold 2 and FDR value cut-off <0.01.

Table 3. Differentially Expressed Transcripts (DETS) Identified in Wheat cv. HD2329 Showing Upregulation* and Downregulation** under Heat Stress

Transcript ID	*Functional Annotation*	*Length*	*Log FC*	*FDR*
Upregulated				
CL921Contig1	Bag family molecular chaperone regulator 6-like	3012	10.98	3.03E-10
CL5691Contig1	Heat stress transcription factor a-6a-like	1074	10.97	3.30E-10
CL1025Contig1	NADH-plastoquinone oxidoreductase subunit 7	1488	10.20	1.88E-08
TrANsCripT_17144	Hypothetical protein F775_00566	757	9.72	2.25E-07
CL14453Contig1	kDa class IV heat shock protein precursor	660	8.64	4.79E-14
TrANsCripT_5361	Cytochrome C-type biogenesis CCDA-like chloroplastic protein 2-like	1014	8.26	0.000268
TrANsCripT_35857	IRR receptor-like serine threonine-protein kinase	433	8.05	0.000598
TrANsCripT_6584	Receptor-like protein kinase feronia-like	852	8.05	0.000598
TrANsCripT_15253	Oxidoreductase GLYR1-like	576	7.35	0.006304
CL9762Contig1	Amino acid binding	522	7.26	0.007867
TrANsCripT_10398	Succinate dehydrogenase	531	7.26	0.007867
TrANsCripT_10584	Heat shock cognate 70 kDa	752	7.09	5.12E-08
CL5052Contig1	Heat-shock protein	681	7.04	7.15E-12
Tr_CoNTiG_32171	Heat shock protein 17	585	6.78	6.87E-12
CL9999Contig1	kDa heat-shock protein	792	6.28	3.22E-11
TrANsCripT_10661	ATP synthase CF0 subunit III	529	6.21	5.38E-11
TrANsCripT_10575	Peroxisomal FA beta-oxidation multifunctional protein	698	6.01	1.52E-09
CL895Contig1	Heat shock cognate 70 kDa	2220	5.89	2.70E-10
CL17371Contig1	Heat shock protein 17	715	5.77	5.22E-10
CL5557Contig1	L-ascorbate peroxidase 3	888	5.70	1.14E-08
CL11394Contig1	Heat shock protein 90	750	5.58	1.45E-09
CL4517Contig1	Photosystem I assembly protein YCF4	2699	5.48	2.06E-09
CL6059Contig1	Rubisco large partial	563	5.47	2.02E-09
CL874Contig1	Peroxidase 52	1909	5.34	1.23E-08
TrANsCripT_34599	Expansin EXPA11	307	5.07	0.000711
CL9811Contig1	Cell wall-associated hydrolase	428	5.06	1.48E-06
Downregulated				
CL5507Contig1	Pollen-specific protein SF3-like	616	−12.4931	6.45E-13
CL2138Contig1	Major allergen PHL P 5	1156	−12.3272	1.56E-12
CL2649Contig1	ATPase plasma membrane-type-like	3070	−11.8172	2.21E-11
CL6812Contig1	Hypothetical protein TRIUR3_21511	441	−11.4901	1.24E-10
CL12055Contig1	Protein	510	−11.4014	2.00E-10
CL3697Contig1	Sucrose transporter	938	−10.6412	1.23E-08
CL6973Contig1	Probable purine permease 11-like	1125	−10.5989	1.55E-08
CL9972Contig1	Predicted protein	931	−10.4773	2.78E-08
Tr_CoNTiG_18983	Histone H3	562	−10.4571	3.07E-08
CL17784Contig1	ABC transporter G family member 11-like	732	−10.0447	2.49E-07
CL11833Contig1	Pyruvate decarboxylase isozyme 2	792	−9.5688	4.39E-07
CL14409Contig1	Transcription initiation factor TFIID subunit 3-like isoform X1	816	−9.4916	5.95E-07
CL13739Contig1	Catalytic hydrolase	658	−9.39597	9.82E-07
Tr_CoNTiG_18808	AUX IAA protein 5-methyltetrahydropteroyltriglutamate-	784	−9.30865	1.54E-06
Tr_CoNTiG_16515	Homocysteine expressed	452	−9.08162	4.60E-06
CL12123Contig1	Protein FAR1-related sequence 5	577	−9.06395	5.04E-06
CL3307Contig1	Hypothetical protein TRIUR3_10580	618	−9.06395	5.04E-06
Tr_CoNTiG_34766	Fibre protein FB34	605	−9.04607	5.57E-06

*log2, fc ≥ 5; **log2, fc ≤ −9; FC, Fold change; FDR, False discovery rate.

genes involved in several other processes such as biotic stress, cell wall integrity, starch biosynthesis, lipid metabolism, ascorbate cycle, glutathione cycle, and secondary metabolism pathway.

Biological processes such as secondary metabolite, photosynthesis, photorespiration, amino acid metabolism, carbohydrate metabolism, and ATP synthesis were observed to be significantly enriched with upregulated transcripts in response to HS (Fig. 6). Lipid, sucrose, nitrogen, and amino acid metabolism were observed to be enriched with downregulated transcripts under the HS. Similarly, the cellular processes of abiotic stresses, redox processes, and cell cycle were enriched with upregulated transcripts, while biotic stress, developmental process, and drought-associated were enriched with downregulated transcripts in response to the HS (Fig. 6).

Regulatory pathways such as transcription factors, phytohormone synthesis, calcium-based regulation, and mitogen-activated protein kinase (MAPK) were observed to be enriched with upregulated transcripts in response to HS. Similarly, regulatory pathways associated with protein modifications, regulation by receptor kinases, and few TFs were enriched with downregulated transcripts under the HS. Metabolic pathways characterization showed majority of

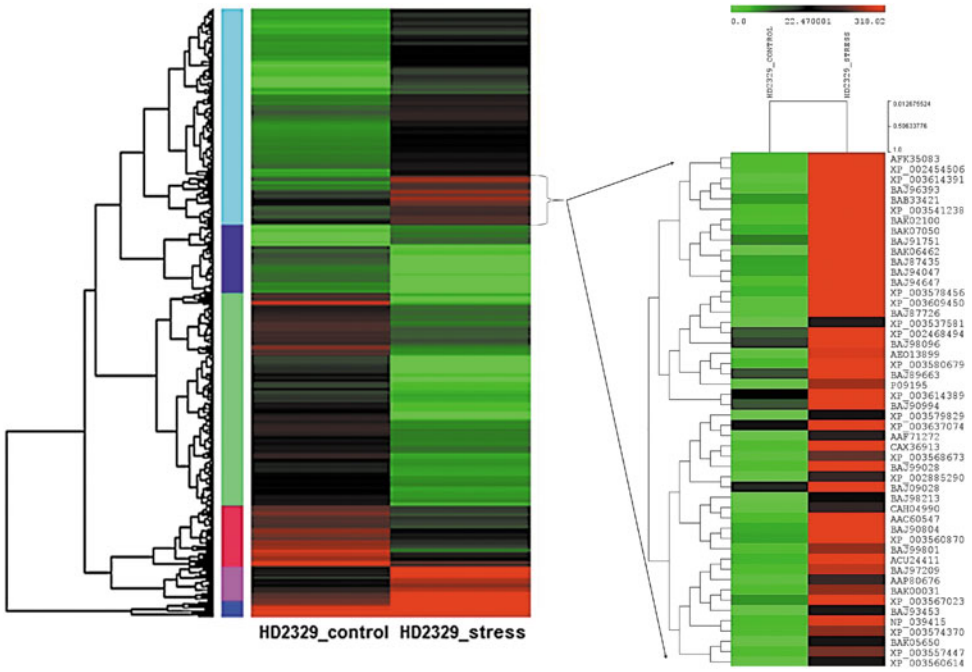

FIG. 4. Clustering and heat map of differentially expressed transcripts (DETs) in wheat under the heat stress. DETs were identified based on the RPKM value. Set of genes displayed in *red* and *green color* show the high fold upregulation and downregulation in response to heat stress (42°C, 2 h).

HS-responsive genes to be associated with photosynthesis, cell wall, and lipids bins. Downregulation of genes associated with cell wall synthesis showed disintegration of cell wall. We also observed downregulation of genes involved in carbohydrate (CHO) metabolism under the HS (Fig. 6).

Distribution of DETs on T. aestivum chromosomes

The DETs identified from wheat *cv*. HD2329 at anthesis stage were mapped on the *T. aestivum* chromosomes based on the recent survey sequence of wheat downloaded from Ensembl Plants (http://plants.ensembl.org/). Data statistics for the Chromosome Survey Sequence (CSS) of wheat are presented in Supplementary Table S1. We downloaded 742742 scaffolds with 792588 contigs having an average size 5642.52 and N50 value of 6651. We observed 1355 (control) and 1362 (HS) transcripts to be localized on the chromosome 3B. Similarly, chromosome 5BL harbors 898 and 878 transcripts of control and HS-treated samples (Fig. 7). To conclude, a maximum number of DETs was observed to be

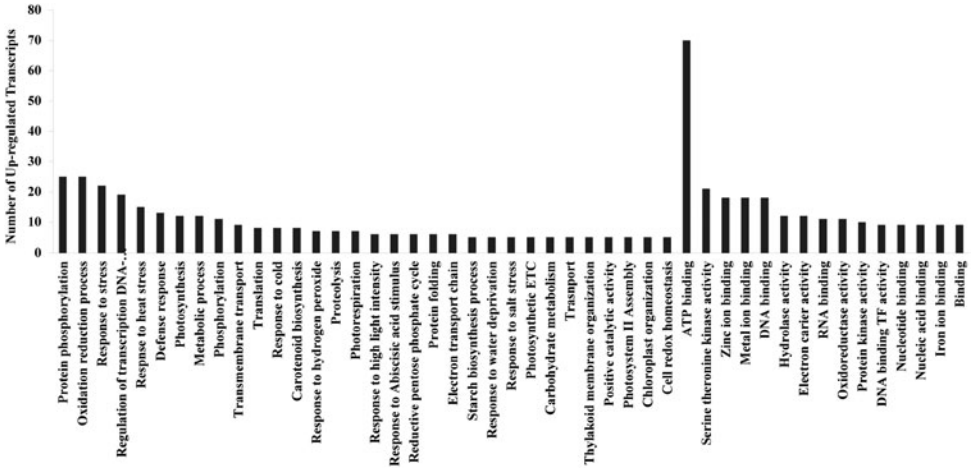

FIG. 5. Distribution of differentially expressed transcripts (DETs) based on Gene Ontology analysis. Upregulated transcripts were grouped based on their biological and molecular functions and involvement in different pathways using the GO database (http://www.geneontology.org/).

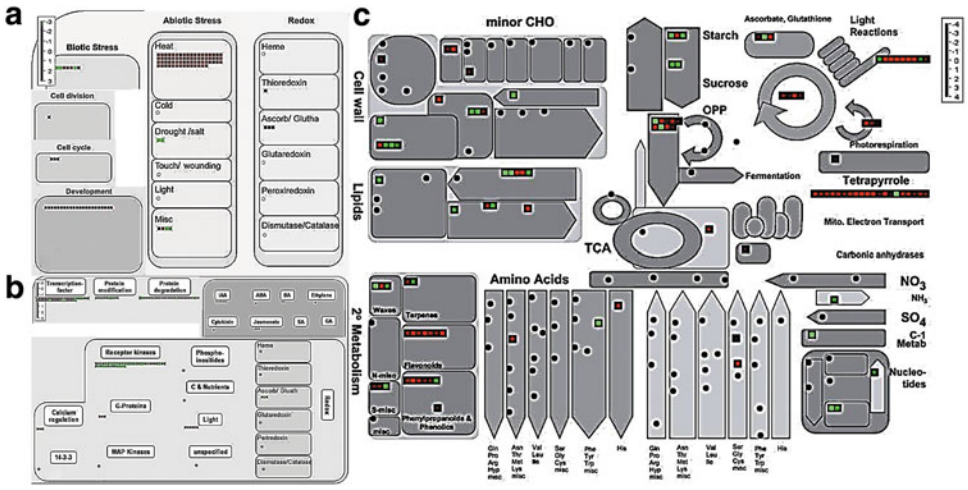

FIG. 6. Overview of the MapMan visualization of differentially expressed transcripts (DETs) in wheat under heat stress. **(a)** Effect of HS on stress-associated genes, cell cycle, and redox-regulated pathways. **(b)** Effect of HS on genes associated with photosynthetic pathway, hormonal synthesis and regulation, and protein metabolism. **(c)** Effect of HS on genes associated with carbohydrate, lipid, amino acid metabolism, and other metabolite pathways. Pathways were analyzed by the MapMan software (http://mapman.gabipd.org/web/guest/mapman). Fold changes (Log2) are indicated by the color scale, *red* and *green* represent increase and decrease of expression, respectively, in the HS-treated sample as compared to the control sample.

located on the chromosome 3B in both the samples. A list of unigenes along with their chromosomal localization are presented in Supplementary Table S8.

Expression cluster analysis

Genes having similar pattern of expression (co-expressed) can be easily identified by cluster analysis; we used K-Means clustering for co-expression analyses (Dembele and Kastner, 2003), and identify ten expression clusters (with six of them having >100 genes each) (Supplementary Fig. S4; Supplementary Table S9). Subclusters 1, 2, 3, 5, 7, and 8 showed more than 100 genes having same pattern of expression under HS in wheat *cv*. HD2329. Subcluster 3 showed more than 420 genes to have similar pattern of expression under the HS. Most of the genes in subcluster 1 show downregulation under HS such as pollen-specific protein SF3-like (log2, FC-5.04), starch branching enzyme 3 (log2, FC-3.44), and IAA17-auxin-responsive AUX IAA family member (log2, FC-2.79).

Similarly, subcluster 2 has downregulated genes such as nitrate transporter-like (log2, FC-5.07), NAC domain-containing protein 72 (log2, FC-2.8), and sucrose transporter (log2, FC-6.9). Subcluster 3 has upregulated genes, to name a few, NADH-plastoquinone oxidoreductase subunit 7 (log2, FC-10), MYB transcription factor (log2, FC-2.2), and heat stress transcription factor a-6a-like (log2, FC-4.3). Similarly,

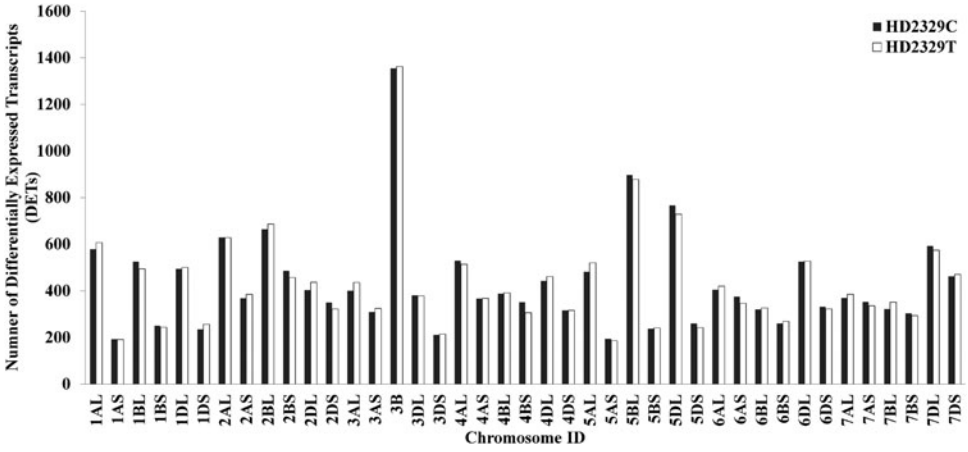

FIG. 7. Chromosomal localization of identified differentially expressed transcripts (DETs) from wheat under heat stress. Recently published chromosome survey sequence (CSS) of wheat was downloaded from Ensembl Plants (ftp.ensemblgenomes.org/pub/release-24 / plants / fasta / triticum _ aestivum / dna) and was used for the characterization.

Table 4. Differentially Expressed Transcripts (DETS) Involved in Photosynthetic Carbon Assimilatory Pathway in Wheat cv. HD2329 under Heat Stress

Sequence Description (NCBI NR)	*Transcript ID*	*Log FC*
Upregulated		
NADH-plastoquinone oxidoreductase subunit 7	CL1025Contig1	10.20
Cytochrome c-type chloroplastic protein 2-like	TrANsCripT_5361	8.26
Oxidoreductase glyr1-like	TrANsCripT_15253	7.35
Heat-shock protein	CL5052Contig1	7.04
ATP synthase cf0 subunit iii	TrANsCripT_10661	6.21
L-ascorbate peroxidase 3	CL5557Contig1	5.70
Photosystem I assembly protein YCF4	CL4517Contig1	5.48
Ribulose-bisphosphate carboxylase oxygenase large partial	CL6059Contig1	5.47
Cytochrome B6F complex subunit IV	CL6623Contig1	4.79
Chloroplast ORF137	TrANsCripT_12855	4.78
Photosystem II 47 kDa protein	CL17230Contig1	4.77
Cytochrome F	CL9720Contig1	4.74
Photosystem II Cytochrome B559 alpha subunit	Tr_CoNTiG_21470	4.72
Photosystem I P700 apoprotein A1	CL3240Contig1	4.69
photosystem II protein d1	CL387Contig2	4.54
Hypothetical protein F775_01096	CL11828Contig1	4.33
ATP-dependent clp protease adapter protein clps	CL670Contig1	4.20
Stromal 70 kDa heat shock-related chloroplastic-like	Tr_CoNTiG_28872	4.16
Downregulated		
Sucrose transporter	CL3697Contig1	−10.64
Histone H3	Tr_CoNTiG_18983	−10.46
Hypothetical protein F775_01182	CL11961Contig1	−9.26
5-methyltetrahydropteroyl triglutamate-homocysteine expressed	Tr_CoNTiG_16515	−9.08
Phytosulfokine receptor 2	CL1701Contig1	−8.87
Probable sodium metabolite cotransporter chloroplastic-like	Tr_CoNTiG_27348	−8.25
3-oxoacyl-synthase I	CL12702Contig1	−7.86
SCY1-like protein 2-like	CL14372Contig1	−7.59
Carbonic anhydrase	TrANsCripT_10432	−7.48
Predicted protein	Tr_CoNTiG_36436	−7.32
Not2 Not3 Not5 family expressed	Tr_CoNTiG_24711	−7.12
Hypothetical protein F775_07261	CL3495Contig1	−6.16
Bidirectional sugar transporter sweet11-like	CL9466Contig1	−5.95
Cytochrome P450 86B1-like	CL5167Contig1	−5.63
Protein notum homolog	Tr_CoNTiG_28662	−5.47
ADP-glucose pyrophosphorylase large subunit	CL12887Contig1	−5.39
Predicted protein	CL12640Contig1	−4.76
Histone H4	CL6976Contig1	−4.61
ATP binding	CL10449Contig1	−3.82
MADS-box transcription factor	CL5071Contig1	−3.82
Early light-induced protein	TrANsCripT_9324	−3.59
Protein executer chloroplastic-like	CL2899Contig1	−3.50

subcluster 4 harbour genes reveals upregulation under HS such as stromal 70 kDa heat shock-related chloroplastic-like (log2, FC-2.31), and Rubisco small subunit (log2, FC-2.66).

Subcluster 5 shows both upregulated as well as downregulated genes such as hypothetical protein, expressed (log2, FC-8.1), triosephosphate isomerase (log2, FC-4.74), and zinc finger CCCH-domain containing protein 44 (log2, FC-4.45). Subcluster 6 shows upregulated genes such as chlorophyll a-b binding protein chloroplastic-like (log2, FC-2.72), HSP90 (log2, FC-5.5), and class I heat shock protein (log2, FC-4.9). Subcluster 7 has upregulated transcripts such as class IV HSP precursor (log2, FC-8.64), and heat-shock protein (log2, FC-7.09), HSC 70 kDa (log2, FC-5.89).

Downregulated transcripts are observed in subcluster 8: fructose-bisphosphate aldolase (log2, FC-2.45) and sucrose synthase 2 (log2, FC-4.5). Subcluster 9 shows upregulated transcripts such as Rubisco activase (log2, FC-2.5), 23s ribosomal RNA (log2, FC-3.94), whereas subcluster 10 has downregulated transcripts of phospholipase A2 (log2, FC-2.33) and adenosylhomocysteinase (log2, FC-2.33). Genes present in subcluster 1, 2, 5, 8, and 10 show downregulation, whereas subcluster 3, 4, 6, 7, and 9 show upregulation under HS in heat-sensitive HD2329 cultivar during flowering stage (Supplementary Fig. S5; Supplementary Table S9).

Discovery of novel transcripts and their regulation under HS

We identified 658 novel transcripts, of which 311 were upregulated and 347 downregulated under the HS. To understand its biological relevance, an attempt was made to predict gene on these novel transcripts using ab initio gene prediction methods (Fickett, 1996) implemented in Augustus, utilizing *A. thaliana* and *Z. mays* gene annotation as

training sets. A list of gene prediction made for the identified novel upregulated and downregulated transcripts are presented in the Supplementary Tables S10 and S11.

We predicted 5'UTR and 3'UTR including introns of most of the identified DETs and observed that many of the genes have predicted CDS regions with exons, suggesting that unannotated genes play key roles in stress tolerance. 244 (37.0%) of the novel transcripts were observed to have complete CDS (Fig. S1). We observed 238 transcripts (based on *A. thaliana* training set) and six transcripts (based on the *Z. mays* training set) with CDS. For example, the functional gene structure of some of the annotated transcripts (51, 59, 71, and 74) are depicted in Supplementary Figure S6.

Identification of DETs involved in photosynthetic carbon assimilatory process

The photosynthetic carbon assimilatory process plays a very important role in fixing the carbon into sucrose, which is further mobilized to the endospermic tissue of the grain and is accumulated as starch. Here, we observed significant variations in the expression of DETs associated with photosynthesis in wheat *cv.* HD2329 under HS (Table 4). Some of the unigenes (associated with the photosynthetic carbon assimilatory pathway) that showed upregulation under HS are NADH-plastoquinone oxidoreductase subunit 7 (log2, FC-10.2), cytochrome C-type chloroplastic protein 2-like (log2, FC-8.26), heat-shock protein (log2, FC-7.04), ATP synthase CF_0 subunit III (log2, FC-6.21), Photosystem I assembly protein YCF4 (log2, FC-5.48), and Rubisco LSU partial (log2, FC-5.47).

NADH-plastoquinone oxidoreductase, which showed the highest upregulation under HS, is involved in oxidation-reduction reaction and is present in chloroplastic thylakoid (Table 4). Similarly, some of the unigenes that showed downregulation under HS are sucrose transporter (log2, FC-10.64), and histone H3 (log2, FC-10.46) (Table 4). Sucrose transporter, which plays very important role in transporting the photosynthates from source to sink, showed maximum downregulation under HS.

Validation of identified DETs

Ten DETs associated with HS-tolerance, photosynthesis, metabolism, and oxidoreductase activity were randomly selected for the validation in control (22° ± 3°C) and HS-treated (42°C, 2 h) samples. The transcripts selected for the validation were *HSFA6A* (CL13901Contig1), *HSP90* (CL20Contig3), ATP-dependent Clp protease proteolytic subunit (*CLPP*; transcript_15008), ribulose bisphosphate carboxylase/oxygenase activase A (*RCA-A*; CL2092Contig1), cytochrome c oxidase polypeptide III (*COX3*; CL4709Contig1), glyceraldehyde-3-phosphate dehydrogenase chloroplastic (*GAPDH*; CL118Contig1), cytochrome P450 (CL6044Contig1), heat shock protein 101 (CL5711Contig1), Novel transcript (transcript_11676), and small heat shock protein *HSP17.8* (CL1Contig9).

Expression profiling of *HSFa-6a* showed 3.5-fold upregulation in response to HS, while HS-associated genes like HSP90 and HSP101 showed 2.2- and 1.7-fold upregulation under HS, as compared to control. Similarly, cytochrome c oxidase (*COX3*) and cytochrome P450 showed upregulation in response to HS (Fig. 8). RCA which acts as catalytic chaperone for converting the inactive form of Rubisco into active form showed upregulation (8-fold) under HS, which is quite significant for photosynthetic carbon assimilatory process. The expression of novel transcript and *HSP17.8* (small HSP) showed 26.2- and 27.5-fold upregulation in HS-treated sample, compared to control (Fig. 8).

Tissue-specific expression of DETs under HS

The validated DETs were analyzed for their expression in different tissues such as root, stem, and spike of heat-sensitive wheat *cv*. HD2329 exposed to HS during anthesis.

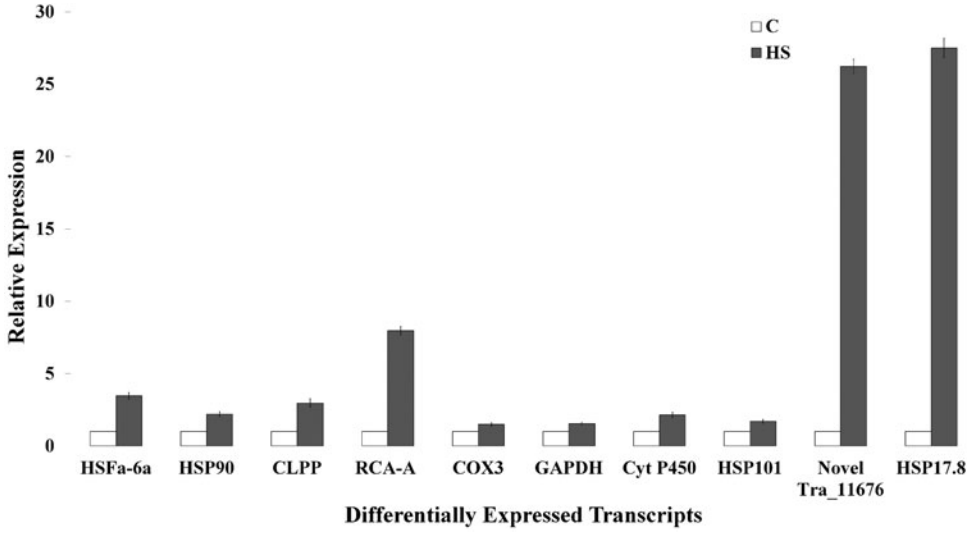

FIG. 8. Validation of identified unigenes in wheat under heat stress using quantitative real time RT-PCR (qRT-PCR). Ten DETs selected for the validation are *HSFA6A*, *HSP90*, *CLPP*, *RCA-A*, *COX3*, *GAPDH*, cytochrome P450, *HSP101*, Novel transcript _11676 and *HSP17.8*. β-actin gene (accession no. AF282624) was used for normalizing the data. Relative fold expression was calculated by Pfaffl method. *Vertical bars* indicate *s.e* ($n = 3$).

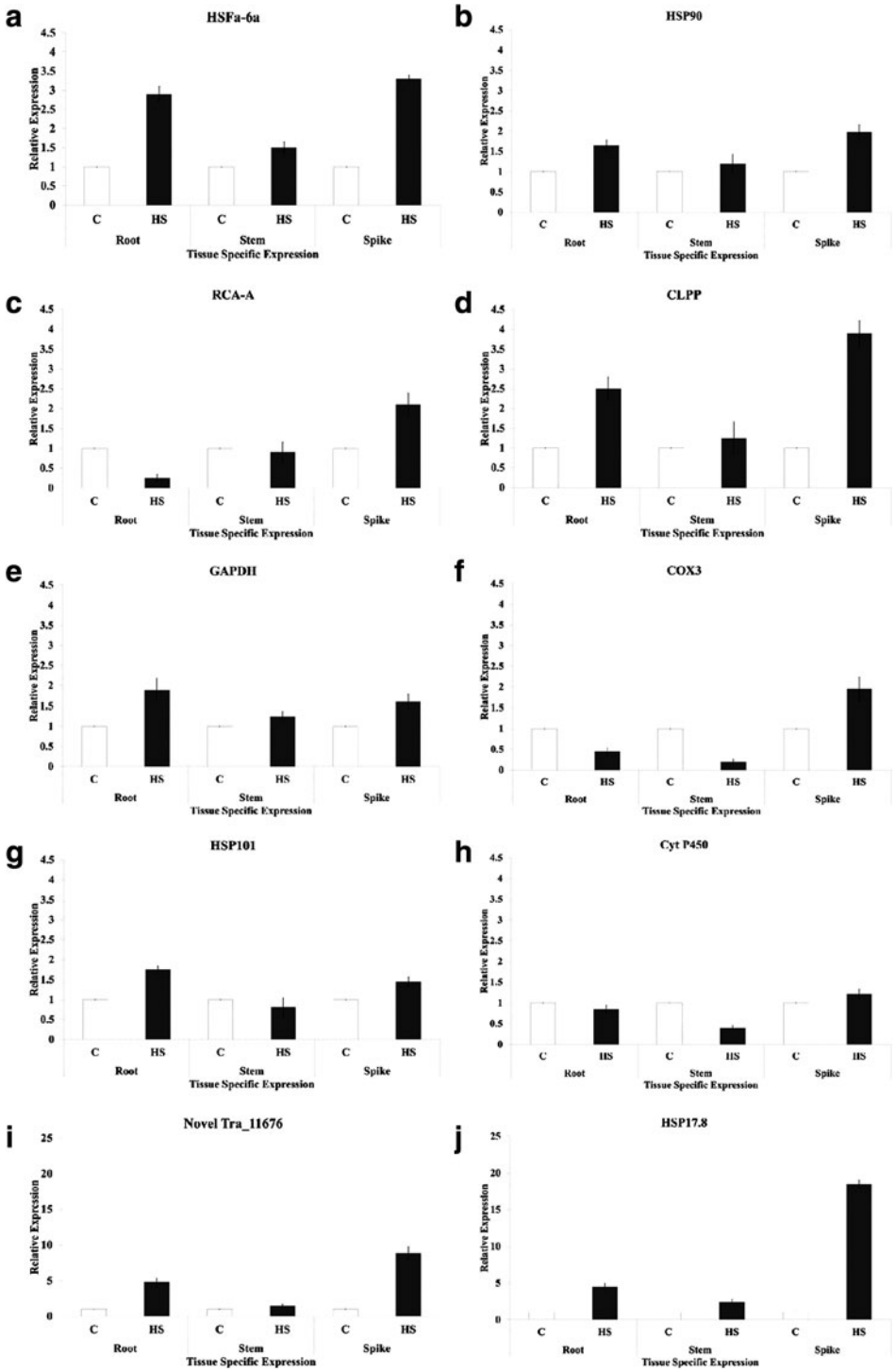

FIG. 9. Tissues specific expression profiling of selected unigenes in wheat under heat stress. Ten transcripts selected for the validation are *HSFA6A*, *HSP90*, *CLPP*, *RCA-A*, *COX3*, *GAPDH*, cytochrome P450, *HSP101*, Novel transcript _11676 and *HSP17.8*. Root, stem, and spikes collected in triplicates were used for the validation. The expression of β-actin gene (accession no. AF282624) was used for normalizing the data. Relative fold expression was calculated by Pfaffl method. *Vertical bars* indicate *s.e* ($n = 3$).

Novel transcript (transcript_11676) showed 4.8-fold upregulation, whereas *RCA-A* was highly downregulated (0.25-fold) in the root under the HS (Fig. 9). We also observed abundance of *HSP17.8*, *HSFa-6a,* and ATP-dependent Clp protease proteolytic subunit in root under HS. Similarly, *HSP17.8* was 2.45-fold upregulated in HS-treated stem, as compared to control. *RCA-A*, *COX3,* and *Cyt P450* showed downregulation in the stem under HS. Expression profiling of selected genes in spike showed *HSP17.8* with very high fold expression followed by novel transcript under HS. Other

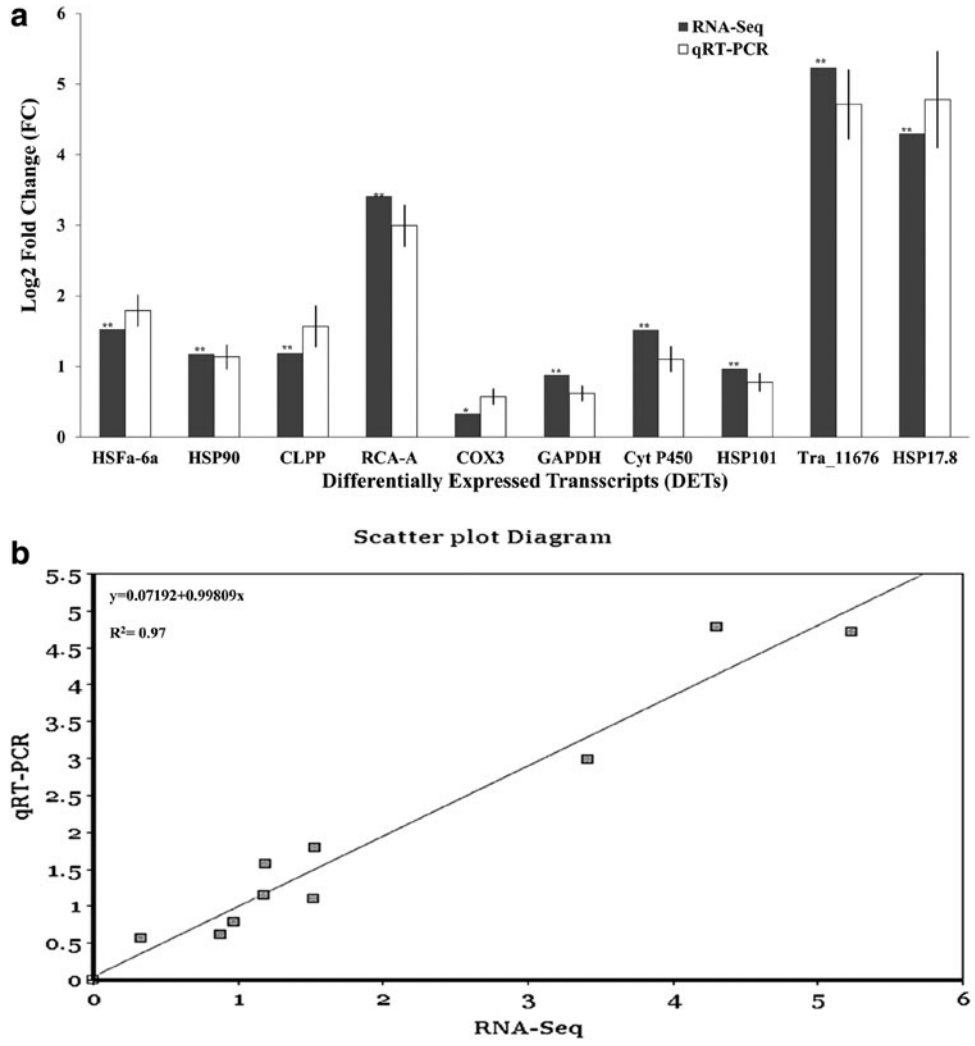

FIG. 10. Quantitative reverse transcriptase PCR validations of RNA-Seq results. **(a)** Comparison of differential expression of ten unigenes determined by RNA-Seq (*dark gray*) and qRT-PCR (*white*). **(b)** Scatter plot diagram showing the relationship between the log2 fold changes (FC) of unigenes observed from RNA-Seq and qRT-PCR. Scatter plots were generated by the log2 expression ratios from RNA-Seq (x-axis) and qRT-PCR (y-axis). Results of ANOVA are shown (**$p < 0.05$; *$p > 0.05$). *Vertical bars* indicate *s.e* ($n = 3$).

selected DETs also showed upregulation in response to HS; fold increase was significant in different tissues (Fig. 9).

Validation of digital gene expression (DGE) of identified unigenes through quantitative real-time RT-PCR

In order to confirm the accuracy and reproducibility of the transcriptome analysis results, the DGE (log2 FC) of 10 randomly selected DETs observed through RNA-Seq was validated using the qRT-PCR. The log2 FC value of *HSFA6A* (CL13901Contig1), *HSP90* (CL20Contig3), ATP-dependent clp protease proteolytic subunit (transcript_15008), *RCA-A* (CL2092Contig1), *COX3* (CL4709Contig1), *GAPDH* (CL118Contig1), cytochrome P450 (CL6044Contig1), *HSP101* (CL5711Contig1), Novel transcript (transcript_11676), and *HSP17.8* (CL1Contig9) observed by RNA-Seq and qRT-PCR was compared by histogram and scatter plot diagram (Fig. 10). The expression (log2, FC) of selected unigenes under HS was consistent with their DGE derived from RNA-Seq (Fig. 10a). The DGE (as observed by RNA-Seq) and log2, FC of the gene (as estimated by qRT-PCR) showed significantly positive correlation (Fig. 10b), which validates our transcriptome findings.

Discussion

Among various abiotic stresses, HS is considered as the most detrimental for agriculturally important crops in general, and wheat in particular. Predictions on yield reductions in wheat on account of the climate change are frightening. However, it is difficult to identify novel SAGs, when the genomic information is not available (Kumar et al., 2014b). High-throughput NGS is, however, currently contributing significantly in the identification of novel genes associated with the biotic and abiotic stresses.

Various crops such as chickpea (Molina et al., 2011), rice (Mizuno et al., 2010; Zhang et al., 2012), sorghum (Dugas et al., 2011), soybean (Hao et al., 2011), and parsley (Li et al.,

2014) have been subjected to high-throughput NGS for identification of abiotic stress-related genes. RNA-Seq is virtually changing the area of sequencing and transcriptome profiling (Wang et al., 2009). Recently, *de novo* assembly has been used in some of the agriculturally important crops like chilli pepper (Liu et al., 2013), coconut (Fan et al., 2013), rice (Zhang et al., 2013), and sugarcane (Cardoso-Silva et al., 2014) for identification of novel genes.

We used Illumina HiSeq™ 2000 and Roche GS-FLX 454 for high-throughput deep sequencing of a heat-sensitive wheat *cv.* HD2329 under the control and HS-treated conditions. Earlier, Roche GS-FLX 454 was the most widely used platform for *de novo* transcriptome sequencing, due to its long read length in different organisms, for example, ginseng (Sun et al., 2010), *A. thaliana* (Wall et al., 2009), and maize (Vega-Arreguin et al., 2009). The Illumina transcriptome was mainly used for the sequencing of organisms whose genome was sequenced (Li et al., 2010). It was confirmed in due course of time that the relatively short reads can be effectively assembled by the Illumina transcriptome, or whole genome *de novo* sequencing and assembly with the advantage of paired-end sequencing (Maher et al., 2009).

The experiment was optimized for *de novo* transcriptome assembly, taking into consideration the suggestions given by Duan et al. (2012). With a view to maximize annotation percentage, five different databases (NCBI nonredundant, UniProt: Swiss-Prot, TriFLDB, PlantCyc, and Pfam) were used, as reported earlier by Wu et al. (2014) for *de novo* transcriptome of common bean. Previous studies have identified approximately 165,499 unigenes in wheat infested with Fusarium blight (Xiao et al., 2013), and 14,550 DETs in wheat at the anthesis period (Wan et al., 2008). DETs in response to inorganic phosphate (P_i) deprivation were identified using *de novo* assembly in wheat (Misson et al., 2005).

Similarly, 120,911 and 117,969 transcripts were identified in DV92 and G3116 (*Triticum monococcum*) using a *de novo* transcriptomic approach (Fox et al., 2014). We observed response to heat stress, protein folding, and oxidation reduction process with higher number of upregulated genes. Heat shock proteins molecular chaperones showed a significant percentage in the upregulated group, which is in conformity with the observations of Chen et al. (2014) in *Populus euphratica*.

Through cluster mapping, we identified 10 expression clusters having similar pattern of expression under HS. Genes involved in response to HS, protein folding, and responses to high light intensity were highly upregulated. Pathways such as metabolic, photosynthesis, and secondary metabolites synthesis, were highly influenced by the HS. Our observations are in close conformity with the findings in poplar (Chen et al., 2014; Song et al., 2014). Many genes responsive to abiotic stress has been reported to be co-expressed across large transcriptome datasets of Barley (Mochida et al., 2011).

Similar observations were reported by Mizuno et al. (2012) in sorghum and Sarkar et al. (2014) in rice under HS using transcriptomic approaches. Under HS, enrichment of binding sites for HSFs, bZIPs and DREBs were observed, along with the increase in the expression of different families of TFs (Sarkar et al., 2014). Similarly, we used ab initio method for the functional annotation of transcripts using the dataset of *Arabidopsis and Z. mays*; now-days, ab initio method is frequently used to assemble a transcriptome for interpreting the annotated transcripts (Denoeud et al., 2008; Guttman et al., 2010; Trapnell et al., 2010).

MapMan-based analyses were observed to be consistent with the GO enrichment results. Metabolism overview showed that most of the heat-responsive genes were involved in the abiotic stress responses, development, secondary metabolism, carbohydrate metabolism, and starch biosynthesis, which are in conformity with the findings of Zhang et al. (2014) in rice post-meiosis panicle under the HS. GO enrichment analysis too revealed that DETs were significantly enriched in signaling, biological, and metabolic process regulation in poplar under HS (Song et al., 2014).

We observed sets of unigenes associated with photosynthesis (source) and grain-filling (sink) showing significant upregulation and downregulation in response to HS; the findings were in consistent with the observations of Song et al. (2014) in poplar under HS. Heat stress reduces the activity of transporters and enzymes associated with source and sink disturbing its ratios; the effect is more pronounced in endospermic tissues.

Validation of a few randomly selected DETs showed very high fold increase in the expression of *HSP17.8* and a novel transcript_11676, followed by *RCA-A* and *HSFa-6a*, etc. under the HS. Very high levels of upregulation of *RCA-A* has been reported in response to HS in wheat (Wan et al., 2008), which is concurrent with our observations. Different mechanisms involving signaling molecules, chaperones, TFs, and other SAGs has been reported to be the reason behind the wide-thermotolerance diversity in wheat (Qin et al., 2008), similar to our present findings.

The findings in the present investigation significantly enrich the transcript dataset of wheat available on public domain. It shows a *de novo* approach to discover the heat-responsive transcripts of wheat, and the information generated could accelerate the progress of wheat stress-genomics as well as breeding programs.

Acknowledgments

We acknowledge the financial assistance received from Indian Council of Agriculture Research under the NICRA project (TG3079) and extra-mural fund from SERB, Department of Science and Technology (sanction order no. SERB/SB/SO/PS/07/2014).

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Altschul SF. (1993). A protein alignment scoring system sensitive at all evolutionary distances. J Mol Evol 36, 290–300.

Arumuganathan K, and Earle ED. (1991). Nuclear DNA content of some important plant species. Plant Mol Biol Rep 3, 208–218.

Cardoso-Silva CB, Costa EA, Mancini MC, et al. (2014). De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLoS One 9, e88462.

Chen J, Yin W, and Xia X. (2014). Transcriptome profiles of *Populus euphratica* upon heat shock stress. Curr Genomics 15, 326–340.

Dembele D, and Kastner P. (2003). Fuzzy C means method for clustering microarray data. Bioinformatics 19, 973–980.

Denoeud F, Aury JM, Silva CD, et al. (2008). Annotating genomes with massive-scale RNA sequencing. Genome Biol 9, R175.

Duan J, Xia C, Zhao G, Jia J, and Kong X. (2012). Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genomics 13, 392.

Dudoit S, Yang YH, Callow MJ, and Speed TP. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Stat Sin 12, 111–139.

Dugas DV, Monaco MK, Olsen A, et al. (2011). Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. BMC Genomics 12, 514–535.

Fan H, Xiao Y, Yang Y, et al. (2013). RNA-Seq analysis of *Cocos nucifera*: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. PLoS One 8, e59997.

Fickett JW. (1996). Finding genes by computer: The state of the art. Trends Genet 12, 316–320.

Fox SE, Geniza M, Hanumappa M, et al. (2014). De novo transcriptome assembly and analyses of gene expression during photomorphogenesis in diploid wheat *Triticum monococcum*. PLoS One 9, e96855.

Gong F, Yang L, Tai F, Hu X, and Wang W. (2014). ''Omics'' of maize stress response for sustainable food production: opportunities and challenges. OMICS 18, 714–732.

Goswami S, Kumar RR, Sharma SK, et al. (2014). Calcium triggers protein kinases-induced signal transduction for augmenting the thermotolerance of developing wheat (*Triticum aestivum*) grain under the heat stress. J Plant Biochem Biotechnol 1-12, doi:10.1007/s13562-014-0295-1.

Grabherr MG, Haas BJ, Yassour M, et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644–652.

Guttman M, Garber M, Levin JZ, et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28, 503–510.

Guy C. (1999). Molecular responses of plants to cold shock and cold acclimation. J Mol Microbiol Biotechnol 1, 231–242.

Hao QN, Zhou XA, Sha AH, Wang C, Zhou R, and Chen SL. (2011). Identification of genes associated with nitrogen use efficiency by genome-wide transcriptional analysis of two soybean genotypes. BMC Genomics 12, 525–540.

Hays D, Mason E, Do JH, Menz M, and Reynolds M. (2007). Expression quantitative trait loci mapping heat tolerance during reproductive development in wheat (*Triticum aestivum*). Wheat Prod Stress Environ 12, 373–382.

Huang Y, Li MY, Wang F, et al. (2014). Heat shock factors in carrot: Genome-wide identification, classification, and expression profiles response to abiotic stress. Mol Biol Rep 42, 893–905.

Issac B, and Raghava GPS. (2004). EGPred: Prediction of eukaryotic genes using ab initio methods after combining with sequence similarity approaches. Genome Res 14, 1756–1766.

Kugler KG, Siegwart G, Nussbaumer T, et al. (2013). Quantitative trait loci-dependent analysis of a gene co-expression network associated with Fusarium head blight resistance in bread wheat (*Triticum aestivum* L.). BMC Genomics 14, 728.

Kumar RR, and Rai RD. (2014a). Can wheat beat the heat: Understanding the mechanism of thermotolerance in wheat (*Triticum aestivum* L). Cereal Res Commun 42, 1–18.

Kumar RR, Goswami S, Sharma SK, et al. (2012). Differential expression of heat shock protein and alteration in osmolyte accumulation under heat stress in wheat. J Plant Biochem Biotechnol 22, 16–26.

Kumar RR, Pathak H, Sharma SK, et al. (2014b). Novel and conserved heat-responsive microRNAs in wheat (*Triticum aestivum* L.). Funct Integr Genomics 15, 323–348.

Kumar RR, Sharma SK, Goswami S, et al. (2013). Characterization of differentially expressed stress associated proteins in starch granule development under heat stress in wheat (*Triticum aestivum* L.). Ind J Biochem Biophys 50, 126–138.

Langmead B, Trapnell C, Pop M, and Salzberg SL. (2014). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25.

Large EC. (1954). Growth stages in cereals illustration of the Feekes scale. Plant Pathol 3, 128–129.

Li M-Y, Tan H-W, Wang F, et al. (2014). De novo transcriptome sequence assembly and identification of AP2/ERF transcription factor related to abiotic stress in parsley (*Petroselinum crispum*). PLoS One 9, e108977.

Li R, Zhu H, Ruan J, et al. (2010). De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20, 265–272.

Li YF, Wang Y, Tang Y, Kakani VG, and Mahalingam R. (2013). Transcriptome analysis of heat stress response in switchgrass (*Panicum virgatum* L.). BMC Plant Biol 13, 153.

Liu S, Li W, Wu Y, Chen C, and Lei J. (2013). De novo transcriptome assembly in Chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of Capsaicinoids. PLoS One 8, e48156.

Maestri E, Klueva N, Perrota C, Gulli M, Nguyen HT, and Marmiroli N. (2002). Molecular genetics of heat tolerance and heat shock proteins in cereals. Plant Mol Biol 48, 667–681.

Maher CA, Palanisamy N, Brenner JC, et al. (2009). Chimeric transcript discovery by paired-end transcriptome sequencing. Proc Natl Acad Sci USA 106, 12353–12358.

Marcussen T, Sandve SR, Heier L, et al. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Sci 345, 1250092–1250092.

Martin JA, and Wang Z. (2011). Next-generation transcriptome assembly. Nat Rev Genet 12, 671–682.

Misson J, Raghothama KG, Jain A, et al. (2005). A genome-wide transcriptional analysis responses to phosphate deprivation. Proc Natl Acad Sci USA 102, 11934–11939.

Mizuno H, Kawahara Y, Sakai H, et al. (2010). Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L.). BMC Genomics 11, 683–696.

Mizuno H, Kawahigashi H, Kawahara Y, et al. (2012). Global transcriptome analysis reveals distinct expression among duplicated genes during sorghum-interaction. BMC Plant Biol 12, 121.

Mochida K, Kawaura K, Shimosaka E, et al. (2006). Tissue expression map of a large number of expressed sequence tags and its application to *in silico* screening of stress response genes in common wheat. Mol Genet Genomics 276, 304–312.

Mochida K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, and Shinozaki K. (2011) Global landscape of a co-expressed gene network in barley and its application to gene discovery in *Triticeae* crops. Plant Cell Physiol 52, 785–803.

Molina C, Zaman-Allah M, Khan F, et al. (2011). The salt-responsive transcriptome of chickpea roots and nodules *via* deepSuperSAGE. BMC Plant Biol 11, 31–57.

Mortazavi A, Williams BA, McCue K, Schaeffer L, and Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5, 621–628.

Mousavi S, Alisoltani A, Shiran B, Fallahi H, Ebrahimie E, Imani A, and Houshmand S. (2014). De-novo transcriptome assembly and comparative analysis of differentially expressed

genes in *Prunus dulcis* Mill. in response to freezing stress. PLoS One 9, e104541.

Nagalakshmi U, Waern K, and Snyder M. (2010). RNA-seq: A method for comprehensive transcriptome analysis. Curr Protoc Mol Biol 4–11, 1–13.

Oono Y, Kobayashi F, Kawahara Y, et al. (2013). Characterisation of the wheat (*Triticum aestivum* L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in P_i-stressed wheat. BMC Genomics 14, 77.

Pertea G, Huang X, Liang F, et al. (2003). TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. Bioinformatics 19, 651–652.

Pfaffl MW. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29, e45.

Plaut Z, Butow BJ, Blumenthal CS, and Wrigley CW. (2004). Transport of dry matter into developing wheat kernels and its contribution to grain yield under post-anthesis water deficit and elevated temperature. Field Crops Res 86, 185–198.

Qian X, Ba Y, Zhuang Q, and Zhong G. (2014). RNA-Seq technology and its application in fish transcriptomics. OMICS 18, 98–110.

Qin D, Wu H, Peng H, et al. (2008). Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array. BMC Genomics 9, 432.

Rabara RC, Tripathi P, and Rushton PJ. (2014). The potential of transcription factor-based genetic engineering in improving crop tolerance to drought. OMICS 18, 601–614.

Rensink WA, and Buell CR. (2005). Microarray expression profiling resources for plant genomics. Trends Plant Sci 10, 603–609.

Robinson MD, McCarthy DJ, and Smyth GK. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Sarkar NK, Kim YK, and Grover A. (2014). Coexpression network analysis associated with call of rice seedlings for encountering heat stress. Plant Mol Biol 84, 125–143.

Song Y, Chen Q, Ci D, Shao X, and Zhang D. (2014) Effects of high temperature on photosynthesis and related gene expression in poplar. BMC Plant Biol 14, 111.

Stanke M, Diekhans M, Baertsch R, and Haussler D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24, 637–644.

Sun C, Li Y, Wu Q, et al. (2010). De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. BMC Genomics 11, 262.

Thimm O, Blasing O, Gibon Y, et al. (2004). MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37, 914–939.

Trapnell C, Williams BA, Pertea G, -et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511–515.

Vega-Arreguin JC, Ibarra-Laclette E, Jimenez-Moraila B, et al. (2009). Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing. BMC Genomics 10, 299.

Wall PK, Leebens-Mack J, Chanderbali AS, et al. (2009). Comparison of next generation sequencing technologies for transcriptome characterization. BMC Genomics 10, 347.

Wan Y, Poole RL, Huttly AK, et al. (2008). Transcriptome analysis of grain development in hexaploid wheat. BMC Genomics 9, 121.

Wang Z, Gerstein M, and Snyder M. (2009) RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63.

Wardlaw IF, and Wrigley CW. (1994). Heat tolerance in temperature cereals: An overview. Aust J Plant Physiol 21, 695–703.

Wu AR, Neff NF, Kalisky T, et al. (2014). Quantitative assessment of single-cell RNA-sequencing methods. Nat Methods 11, 41–46.

Xiao J, Jin X, Jia X, et al. (2013). Transcriptome-based discovery of pathways and genes related to resistance against Fusarium head blight in wheat landrace Wangshuibai. BMC Genomics 14, 197.

Xin M, Wang Y, Yao Y, et al. (2011). Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. BMC Plant Biol 11, 61.

Zerbino DR, and Birney E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18, 821–829.

Zhang T, Zhao X, Wang W, Pan Y, Huang L, Liu X, et al. (2012). Comparative transcriptome profiling of chilling stress responsiveness in two contrasting rice genotypes. PLoS One 7, e43274.

Zhang X, Rerksiri W, Liu A, et al. (2013). Transcriptome profile reveals heat response mechanism at molecular and metabolic levels in rice flag leaf. Gene 530, 185–192.

Zhang X, Xiong H, Liu A, Zhou X, Peng Y, Li Z, et al. (2014). Microarray data uncover the genome-wide gene expression patterns in response to heat stress in rice post-meiosis panicle. J Plant Biol 57, 327–336.

Address correspondence to:
Dr. Ranjeet Ranjan Kumar
Lab No. 311, Division of Biochemistry
Indian Agricultural Research Institute (IARI)
Pusa Campus, New Delhi
110012, India

E-mail: ranjeetranjaniari@gmail.com

Abbreviations Used

DEGs = differentially expressed genes
DETs = differentially expressed
DGE = digital gene expression
GO = gene ontology
HS = heat stress
NGS = next-generation sequencing
qRT-PCR = quantitative real-time PCR
SAGs = stress-associated genes