Transcriptome analysis of response to heat stress in heat tolerance and heat susceptible wheat (Triticum aestivum L.) genotypes

Nandha AK, Mehta DR, Tulsani NJ, Umretiya N, Delvadiya N and Kachhadiya HJ

Abstract
Common bread wheat is one of the most crucial cereal crops for human nutrition. Due to its complex structure, the study of molecular biology of this plant become important for further improvement which can provide better quality and quantity of final yield. Wheat is highly sensitive toward the temperature variation as little change in temperature lead to high loss. High quality transcriptome analysis can be helpful to understand this complexity and allow us to do desirable changes. For study two genotypes have been used, one was heat susceptible (WH 147) and second was heat tolerant (GW 451). The transcriptome analysis of wheat was done using IonS5 which gives 287,080 and 263,340 numbers of transcripts in GW 451 and WH 147, respectively. These data were used for preparation of further unigen assembly. Determination of various proteins was done using four different databases and retrieved proteins were matched with different species to get top hits. Gene ontology of stressed and control plant was categorised in three categories and 49 subcategories which showed versatility in each category. The functional analysis also gives functional categories between two genotypes during control and stressed condition. Differential gene expression analysis indicated the notable variation in various genes under stressed condition in both genotypes, which were identified by their up regulated or down regulate level. Identification of pathways which gives the buffering ability to plant under heat stress condition were identified using Kyoto Encyclopedia of Genes and Genomes pathway analysis. Different types of heat shock proteins were identified in this study among which high molecular weight HSP and moderate molecular weight HSP were abundance then low molecular weight. This data can be useful for the further characterisation of molecular analysis of wheat which can helpful to generated heat tolerance variety.

Keywords: heat stress, wheat, transcriptome, heat tolerance, heat susceptible

Introduction
Heat stress is a key abiotic stress affecting crop and cereal production in all regions of the world. Heat stress can have high adverse effect on grain yield and productivity, and potentially greater than other abiotic stress such as drought and frost. Controlled environment studies have established that a 3-5% reduction in grain yield of wheat can occur for every one degree increase in average temperature above 15°C (Gibson and Paulsen, 1999) [4]. Field data suggests that yield losses can be in the order of 190 kg/ha for every one degree rise in average temperature and in some situations having a more serve effect on yield loss than water availability (Kuchel et al. 2007; Bennett et al. 2012) [9, 2]. Many studies have shown that stress prior (boot formation stage) and during pollen development as well as during grain filling adversely affects a number of physiological and molecular processes within cereal plants, all resulting in a reduced grain yield. During pollen formation, from as early as Zadoks growth stage 45 (booting), viability is affected under increased temperature conditions, resulting in reduced grain number. During grain filling, grain fill duration is reduced, leading to reduced grain size and increased screenings levels. Compounding this is an increased leaf area senescence rate.

In plant, ability to tolerate heat is highly complex process which involved various molecular mechanisms. (Suzuki et al., 2005; Kotak et al., 2007) [18, 8]. Even though there were many researches have been done on heat stress response in wheat, the complete anatomy of response is still need to be done. Identification of multigenic regulatory mechanism for high temperature stress response at transcript level can be more crucial and beneficial for the development of the heat tolerance species.
Materials and Methods

Plant material, growth conditions and stress treatment

Common bread wheat (Triticum aestivum) cv GW 451 (Heat tolerance) and WH 147 (Heat sensitive), which is a heat tolerance type of genotype, was used in this study and plants were grown greenhouse during crop season in potted soil at 22°C:16°C day:night temperature in a 16:8 h photoperiod. The experiment was conducted at booting stage. All plants from GW 451 genotype were divided into two groups with similar replication (Group I, II). Group I of wheat plant was kept at 22/16°C day/night in the controlled-environment of greenhouse. Group II was subjected to heat treatments at 40°C in the controlled-environment growth chambers. Except from the temperature, other conditions should be remaining same as control conditions. After heat stress treatment, control and stressed flag leaf and root from boot stage were sampled, flash frozen in liquid nitrogen and stored at -80°C until RNA isolation. Total four samples were collected from treated as well as control wheat genotype.

Isolation of total RNA, mRNA and construction of subtracted cDNA libraries

Total RNA from flag leaf and root tissues, from boot stage, were isolated by Trizole Reagent (Invitrogen) as per the manufacturer protocol. Total RNA from developing wheat seeds were isolated (stages as mentioned above) as per the method described by Singh et al. (2003) [17]. The concentration and quality of the total RNA was determined running a gel electrophoresis in 1.5% (w/v) agarose gels. The poly (A) mRNA was isolated from purified total RNA using biotin-Oligo (dT) magnetic beads and fragmented into small pieces using an RNA fragmentation kit. First-strand cDNA was generated from the cleaved RNA fragments using reverse transcriptase and random primers. Second-strand cDNA was synthesized using RNase H and DNA Polymerase I. Following adaptor ligation, the unsuitable fragments were removed with AM Pure XP beads, 200-bp cDNA fragments were enriched by 18 cycles of PCR. The product were sequence in Ion S5 from

Processing and mapping of reads from sequencing

The RNA-Seq raw reads, which were obtained from sequencing via Ion S5 were, processed to obtain high quality reads by removing the adapter sequences and low quality bases at the 3’ end, trimming low-quality bases (Q < 20) from the 5’ and 3’ ends of the remaining reads. Reads filtering out reads containing ‘N’ and greater than 25 bp were considered for analysis. Clean reads were assembled into contigs, transcripts and unigenes with Trinity software (http://trinityrnaseq.sf.net). RPKM was used to normalize the abundances of transcripts. More than a 2-fold change was used to identify the significance of different gene expression between different treatment lines.

Identification of differentially expressed unigenes

All unigenes were blastx searched against wheat (Triticum spp.), in KEGG and KOG protein databases (E-value < 10^{-5}) and functionally annotated by Blast2GO Gene Ontology Functional Annotatation Suit (E-value < 10^{-5}) (http://www.blast2go.org/). Transcription factor families were identified from identified unigenes. Metabolic pathways were predicted by KEGG mapping. Putative transcription factors were identified by searching Arabidopsis Gene Regulatory Information Server (AGRIS) Database.

Result and Discussion

Winter crop wheat, belonged to poeases family, is highly sensitive toward the heat stress. Heat stress induced injury in plant at various parts which ultimately lead to the loss in total yield. Leaves and root part were collected from control and treated plant. As per earlier studies and our experiment for RNA-Seq analysis, the expression of genes responsible for stress responses can be measured. The results of this study described the gene level transcription in response to heat stress condition in both part of plant. Ion S5 system was used for sequencing which measured total of more than 112 million reads with raw reads in GW 451 and, more than 89 millions of reads with high quality reads. Out of which total number of transcript obtained was 287,080 with length of 170,148,287 bp. Maximum length of this transcript was 539 bp while N50 of 653 bp. While 090 million reads in WH 147 with total 263,340 numbers of reads which contains 137,264,522 bp of total length. The maximum length of these read was 521 bp with same number of N50. By using Trinity software, various numbers of unigenes were identified, as total number of reads was 141,700 in GW 451 and 135,122 in WH 147. In case of total length, GW 451 showed higher number of bp in GW 451 (83,306,330 bp) as compared to WH 147 (70,621,143 bp). The maximum length was 588 bp in GW 451 and 523 bp in WH 147. While mean length of these reads was 751 bp for GW 451 and 629 bp in WH 147. There was also studied on CDS as total 75,407 CDS in GW 451 and 62,915 in WH 147. There was also studied on CDS as total 75,407 CDS in GW 451 and 62,915 CDS in WH 147 were obtained by using BLASTP software using four different ways. Total 75,407 proteins were identified in GW 451 while 62,915

<table>
<thead>
<tr>
<th>Statistics of Assembly</th>
<th>Number of sequences (bp)</th>
<th>Length of sequence (bp)</th>
<th>Max length (bp)</th>
<th>N50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WH 147: 135,122</td>
<td>WH 147: 70,621,143</td>
<td>WH 147: 523</td>
<td>WH 147: 751</td>
</tr>
<tr>
<td>CDS</td>
<td>GW 451: 75,407</td>
<td>GW 451: 42,928,770</td>
<td>GW 451: 4,899</td>
<td>GW 451: 569</td>
</tr>
</tbody>
</table>

Similar findings were also reported by Hu et al. (2014) [6] in turfgrass. They reported that PI578718 (heat sensitive) genotype, obtained transcript of 484,390 numbers with maximum length of 15,993 bp and average length of 1,218 bp and for PI234881 (heat tolerant) genotype, obtained transcript of 483,175 numbers with maximum length of 17,029 bp with average length of 1,192 bp. While Kumar et al. (2015) [10] identified total of 567,485,727 transcripts read accounting for description of various genes were used for the further analyses. HD2329, a variety of wheat was assembled and generated total 63.8 million raw reads from which 52.1 millions high quality reads with N50 values of 678 bp for control sample and 640 bp for treated sample.

Top hit distribution of GW 451 and WH 147 was measured by using BLASTP software using four different ways. Total 75,407 proteins were identified in GW 451 while 62,915


proteins were detected in WH 147, out of which non-redundant database showed higher numbers of hits other then rest (Table 2). GW 451 showed distributions revealed that majority of the hits were found to be against the species *Daucus carota* subsp *sativus* (24,385 proteins) followed by others (10,119 proteins), *Aegilops tauschii* (8,718 proteins) and *Triticum urartu* (5767 proteins) on the other hand for WH147 samples *Aegilops tauschii* (10,009 proteins) and *Triticum urartu* (6,845 proteins), which ranked second and third spaced in top hit distribution. The present investigation work was compared with *Triticum aestivum* hits which showed that 2,686 proteins in GW 451 and 2,964 proteins in WH 147 were match. The greater numbers of hits of *Triticum aestivum* were observed in WH 147 as high as 278 as compared to GW 451 (Fig. 2). The Venn diagram showed the comparison of different database for the detection of unique and common proteins in all four databases.

**Table 2:** Blast statistics against NR, Uniprot, KOG and Pfam database

<table>
<thead>
<tr>
<th>Blast Database</th>
<th>Hits in GW 451</th>
<th>Hits in WH 147</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total protein</td>
<td>75,407</td>
<td>62,915</td>
</tr>
<tr>
<td>NR</td>
<td>67,212</td>
<td>54,030</td>
</tr>
<tr>
<td>UniProt</td>
<td>50,190</td>
<td>37,799</td>
</tr>
<tr>
<td>KOG</td>
<td>33,010</td>
<td>25,470</td>
</tr>
<tr>
<td>Pfam</td>
<td>29,230</td>
<td>21,740</td>
</tr>
</tbody>
</table>

![Venn diagram for comparison of four databases](image_url)

**Fig 1:** Venn diagram for comparison of four databases

A.
GO distribution in the wheat for heat transcriptome

On the basis of Gene Ontology (GO) analysis, WH 147 (1,15,857 genes) showed higher number of genes as compared to GW 451 (92,741 genes). All identified genes were categorized into three main categories 1) biological process (37,961; 40.93%); 2) cellular components (31,931; 34.43%) and 3) molecular functions (22,849; 24.63%) for GW 451. While, in case of WH 147, these data was 47,224 (40.76%); 40,216 (34.71%) and 28,417 (24.52%), respectively (Table 3).

Table 3: Distribution of GO in GW 451 and WH 147

<table>
<thead>
<tr>
<th>Description</th>
<th>GW 451</th>
<th>WH 147</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological Processes</td>
<td>37,961</td>
<td>47,224</td>
</tr>
<tr>
<td>Cellular Components</td>
<td>31,931</td>
<td>40,216</td>
</tr>
<tr>
<td>Molecular Functions</td>
<td>22,849</td>
<td>28,417</td>
</tr>
<tr>
<td>Total</td>
<td>92,741</td>
<td>1,15,857</td>
</tr>
</tbody>
</table>

These three main categories were divided into 49 GO functional sub-categories (Fig. 3). Genes encoding metabolic process and cellular process were top notched process in biological process in both genotypes as 11,560 and 9,944 in GW 451, respectively. While, 14,202 and 12,371 genes respectively, in WH 147. Within the cellular components, protein related to cell (6,391 in GW 451; 8,096 in WH 147) and cellular part (6,295 in GW 451; 7,971 in WH 147) was enriched in both followed by cellular membrane (5,867 in GW 451 7,343 in WH 147) and other mechanisms. The scenario for molecular mechanism was bit different as in catalytic activity (10,698) was higher than other in GW 451 followed by binding (10,608). On the other hand, reverse was the scenario in WH 147 as, 13,281 proteins in binding (13,291) were higher than catalytic activity (13,117).
Andjelkovic and Micic (2011) recorded high numbers of transcripts in maize which were treated under water and heat stress conditions. Out of all transcripts, nine were assigned to biological processes including photosynthesis (three), transport (six), stress (19), cytoskeleton (three), metabolism (11), cell cycle (three), translation (two) and protein formation (one). Li et al. (2013) analysed GO distribution in switchgrass which clearly showed that biological processes associated with metabolism, cellular homeostasis, cell death, regulation of transcription and transporters were significantly over represented among the genes repressed by heat stress in switchgrass. Similarly, Kumar et al. (2015) studied gene ontology and gave various pathways which were categorized in biological, cellular and molecular functions. Most identified unigenes were for metabolic process, cellular, response to stimulus and other biological regulations.

**Functional annotation of KOG and its classification in wheat transcriptome data**

Eukaryotic Orthologous Groups of protein (KOG) was analysed for the further evaluation of function for assembled unigenes. This analysis was classified in 25 different KOG categories. (Fig. 4). In heat tolerant genotype proteins for “Signal transduction showed highest protein counts (3822 counts) followed by “Posttranslational modification, protein turnover, chaperones (O)" (3708 counts), “General function prediction only (R)" (3591 counts) and “Carbohydrate transport and metabolism (G)" (2210 counts) and “Translation, ribosomal structure and biogenesis (O)" (2076 counts), while in heat susceptible genotype “Posttranslational modification, protein turnover, chaperons” showed highest protein counts (3056 counts) followed by "General function prediction only (R)" (2760 counts), "Signal transduction mechanisms (T)" (2726 counts), and "Translation, ribosomal
structure and biogenesis (J)” (1952 counts), “Carbohydrate transport and metabolism (G)” (1507 counts). Lower amount of protein counts were observed in “Cell motility” (seven in GW 451 and 12 in WH 147).

Proteins for Extracellular structure were similar in both genotypes (17 counts). “Function unknown” category indicated not matched proteins with nearby families. In this category GW 451 showed 1656 counts while WH 147 showed 1261 counts. These variation in two genotypes in same KOG categories were exhibited during treatment at various temperature stress. Hu et al. (2014) [6] indicated under high-temperature stress, higher enrichment of the tall fescue unigenes in the most of COG categories for heat-sensitive genotypes compared with heat-tolerant ones suggests that thermotolerance was related with the changed expression of genes involved in transcriptional regulation as well as metabolic pathways. Jung et al. (2012) showed 12 gene ontology terms enrichment for prolonged heat stress relating genes in which 244 early heat stress responsive genes and 238 prolonged heat stress responsive genes were reported in rice.

Differential expression transcript under heat stress
For the identification of relationship of two different genotypes on the basis of their genome wide expression along with their treatment was done using BWA (0.7.12-r1039) and DESeq software. The genes which were detected in two genotypes indicated the number of genes which regulated buffering ability toward the heat stress. In addition, the heat map of identified genes also indicated that the profusion amount of genes under heat treatment (Fig. 5). Out of various identified genes in both genotypes, top 50 up regulated and down regulated genes were indicated in given hit map along with sequence name. The result suggested increase in number of stress responsible genes along with increased temperature. Similar results were seen in tall fescue by Hu et al. (2014) [6]. As well as the decreased number of genes indicates that cool season crop plant, wheat, showed more negative regulated genes in sensitive genotype as compared to positive regulation in tolerance genotype. Qin et al. (2006) reported that heat-
tolerant ‘TAM107’ had more differential expression genes relative to heat-sensitive ‘CS’ (1,011/1,100) at 1 HAT, but less differential expression genes at 24 HAT in wheat using microarray analysis. It also indicated that the greater the tolerance in genotype, higher the complexity in plant.

Frey et al. (2015) indicated the involvement of 53 types of different biological functions in the overall heat responsive genes during stress condition in maize which includes total 14 heat shock proteins. They also identified 607 overall heat responsive genes in maize, of which 460 were up regulated and 147 down regulated, when considering increasing heat levels. In the long-term heat stress imposed on plants, the observed transcript changes reflect the acclimatize response; whereas in the short duration heat stress in the other studies mentioned above, the transcriptional changes reflect the more active defence response. Among uniquely differentially expressed genes, 1,462 were down regulated and 886 genes were up regulated (Li et al., 2013).

Analysis of effect of heat stress on cell by KEGG
The data from expression analysis was further annotated by using Kyoto encyclopedia of genes and genome (KEGG) pathway. Fig. 6 indicated the distribution of various genes in five differential cellular functional categories. More numbers of genes were included metabolism followed by cellular processes and genes for genetic information. In GW 451, genes involved in Translation process were highest with 919 genes while in WH 147 these numbers were 745 genes. On the other hand, Translation involved 384 genes in GW 451 and 332 genes in WH 147. Folding and repair mechanism category, which include various enzymes for reactive oxygen species, contains 226 genes in GW 451 and 180 genes in WH147. Genes which involved in “Signal transduction” are 864 in GW 451 and 672 in WH 147. Genes which involves in membrane transport are 41 in GW 451 and 30 in WH147, which was least number of genes involved in total cellular mechanisms.

Genes involved in metabolism
For carbohydrate metabolism 682 genes were involved in GW 451 while 562 genes were involved in WH 147 followed by energy related metabolism in GW 451 (488) as well as overview of all metabolic pathways (407) in WH 147. Least number of pathways was showed in xenobiotics biodegradation and metabolism in GW 451 (110) and WH 147 (79). Second last was glycan biosynthesis and metabolism, 156 in GW 451 and 123 in WH 147. Apart from these genes for Biosynthesis of various secondary metabolites (218 in GW 451 and 149 in WH 147), Metabolism of cofactors and vitamins (314 in GW 451 and 266 in WH 147), Amino acid metabolism (488 in GW 451 and 382 in WH 147) were also identified using these data. Similar findings were reported by Wei et al. (2014) who reported KEGG pathway annotation against the KEGG database using KAAS. Among the 61393 unigenes, 58724 (95.7%) were returned as matches in at least one database. Out of all pathways, biosynthesis of secondary metabolites (500 DEGs), protein processing in endoplasmic reticulum (223 DEGs), ribosome (185 DEGs), and starch and sucrose metabolism (161 DEGs) were the dominant pathways.

Some other genes for example, genes for “Flag leaf protein inhibitor”, “nitrogen metabolism”, “sulfer metabolism”, “Oxidative phosphorylation” were also identified in these analysis. Out of more than 250 different pathways in each genotype, which were identified in this study, there was two path ways identified only in heat resistance genotype (Fig.7).

Heat shock proteins affected by high temperature stress in wheat
All plant have tendency to produce various set of heat shock proteins (HSPs) to acquire ability to tolerate heat stress. According to Mian et al. (2008) there are three types of
families of heat shock proteins were identified: high molecular weight, moderate molecular weight and low molecular weight. Various researches also indicated that increase the profoundness of various transcript leads to the complexity of the stress response mechanism as molecular chaperone was also involved with heat shock proteins (Wang et al. 2003; Tian et al. 2009; Glorno et al. 2010; Li et al. 2011) [20, 19, 5, 12]. In this study, we identified 12 different most abundant heat shock proteins. Out of which five were high molecular weight HSPs, six were moderate (70 kDa) molecular weight proteins and one was lower molecular weight protein. Which indicated that the high as well as moderate molecular weight containing HSP were play important role to induce the heat stress ability in plants.

Fig 6: Graphical representation of Pathway distribution of GW 451 and WH 147
Conclusion
In summary, this study provided large number of dataset of wheat transcriptome for control and heat stressed plants. Millions of high quality reads were acquired which were used for the further assembly which gives 141,122 unigenes in GW 451 and 135,122 unigenes in WH 147. These data use for the assembly of CDS. Assembled unigenes were use for further top hit distribution, functional annotation and identification of various pathway distributions. After comparison of differential expression analysis, different heat shock proteins from various families were identified along with other ROS enzymes and other secondary molecules, which could play crucial role in order to get heat resistance in wheat.

References
6. Hu T, Sun X, Zhang X, Nevo E, Fu J. An RNA sequencing transcriptome analysis of the high-


