JOURNAL OF DYNAMICS AND CONTROL

VOL 8 ISSUE 8

TRANSCRIPTOME BASED IDENTIFICATION OF KEY GENES RESPONSIBLE FOR HEAT STRESS TRANSCRIPTION FACTORS IN AEGILOPS SPELTOIDES AND TRITICUM AESTIVUM

Jaimeet kaur[1], Parveen Chhuneja[2], Gunjan Mukherjee[1]

[1]University Institute of Biotechnology, Chandigarh University, Mohali, Punjab - 140413, India
[2]School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, Punjab - 141004, India

TRANSCRIPTOME BASED IDENTIFICATION OF KEY GENES RESPONSIBLE FOR HEAT STRESS TRANSCRIPTION FACTORS IN AEGILOPS SPELTOIDES AND TRITICUM AESTIVUM

Jaimeet kaur[1], Parveen Chhuneja[2], Gunjan Mukherjee[1]

[1]University Institute of Biotechnology, Chandigarh University, Mohali, Punjab - 140413, India
[2]School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, Punjab - 141004, India

Abstract: Rising global temperatures are harming wheat, one of the world's most significant grain crops. *Aegilops speltoides*, a wheat diploid B genome progenitor, has been identified as a prospective donor for heat tolerance. Thus, the aim of this work was to assess the transcriptome profiles of hexaploid wheat cultivars, *Triticum aestivum* cv. c273, and heat-tolerant *Aegilops speltoides* acc. Pau 3809. The comparative transcriptome analysis was used to identify and validate the key genes that contribute to heat stress resistance, known as heat stress transcription factors (HSFs). In As-Pau3809 and cv. c273, transcriptome analysis revealed a total of 94,805 and 65,085 genes, respectively. The two wheat transcriptomes have a high degree of gene ontology profile homogeneity under molecular, biological, and cellular activities, implying gene function conservation. There are 43 HSFs in the As-Pau3809 transcriptome data, while 41 in the c273 data were identified. The ExpVIP platform investigated the HSF, broadening the abiotic stress study, and their role in the heat stress response was demonstrated. Along with that, a web-based database, the Heat Stress Genes in Wheat Database (HSGDB, http://www.hsgdb-wheat.com/), was developed as the first heat stress comparative study database and contains the transcriptome assembly and functional annotation. HSGDB is an open-access resource.

Keywords: Wheat, heat stress, heat stress genes , gene ontology, database.

1. Introduction

Global agriculture is confronted with two big challenges: rising food demand due to population growth and climate change. In this scenario, crop yields should grow by 60% by 2050 to ensure global food security, but a decline in the worldwide cultivated area is projected as a result of less rainfall and rising temperatures [1–3]. Human-caused greenhouse gas emissions, on the other hand, will likely elevate global surface temperatures by 1 to 5.7 degrees Celsius by 2081–2100. Extreme weather phenomena like heat waves and droughts have been predicted on a global scale, as they have been since the 1950s. [4].

Wheat is one of the most frequently farmed agricultural plants [5] and the second-most significant component of dietary energy and nutrition around the globe, behind rice. Even though India is the second-largest producer of wheat in the world, yields there have been falling for some time [6]. Global climate change, particularly the current anomalous temperature patterns, poses a serious threat to agriculture. Ensuring nutritional and dietary security for a growing population would be extremely difficult due to future climatic uncertainties and resource limitations [7].

Due to the rise in temperatures ever since the beginning of the decade, climate simulations worldwide predict an increase in average temperature varying between 1.8 and 5.8 °C through the conclusion of the millennium. [8].

Heat stress is a major abiotic factor that affects crop and grain output worldwide. Heat stress has a significant negative impact on grain yield and productivity, potentially outperforming other abiotic stresses such as cold and drought. Research conducted in a controlled environment has shown that the yield of wheat can be decreased by 3-5 percent for every degree over 15°C when the average temperature rises [9]. According to field statistics, yield declines may fall in the range of 190 kg/ha for every one degree increase in temperature on average,with temperature having a greater impact on yield loss than accessibility to water in some scenarios [10] [11].

Numerous investigations have shown that stress during pollen development and grain filling has a detrimental effect on a number of physiological and molecular processes within wheat plants, all of which result in a decreased grain yield. Increased temperature conditions have an effect on pollen viability, resulting in a reduction in grain number. Grain fill duration is reduced during grain filling, resulting in lower grain size. Earlier research has predicted that a rise in temperature at intervals of 18–22 degrees celsius during grain filling reduces grain filling length and grain production by 5% and 3–4%, respectively. This calculation estimates yield potential losses of up to 50% when exposed between 32 and 38 °C at a critical phase of grain development [12] [13].

Heat tolerance in plants is a highly complicated procedure comprising multiple molecular pathways. [14][15] The generation of heat-tolerant organisms may greatly benefit from the identification of multigenic regulatory mechanisms for elevated temperature stress responses at the transcript level.

The word "wheat" refers to a member of the family Poaceae, tribe Triticeae, and Triticum genus [16]. *Triticum aestivum* (Bread Wheat or Modern Wheat), a hexaploid species, has a genome size of approximately 16,000 megabytes and three distinct genomes (AABBDD). The three ancient wheat genome species are AA (*Triticum monococcum*), BB (*Aegilops speltoides*), and DD (*Aegilops tauschii*). Triticum aestivum is extremely vital for humanity in the form of food, feed for animals, and supplies of raw materials for industries [17]. The bread wheat and durum wheat B genomes were produced from an exactly similar diploid species of Aegilops speltoides (BB) wild wheat. Even though this only happened once or twice and only involved a small number of progenitors, due to this, the inherent genetic variation of Ae. speltoides has been underestimated in modern wheat and durum wheat germplasm. [18].

Various studies have documented the significance of wild Aegilops species in breeding strategies, where they improve tolerance to heat [19], salinity [20], drought stress [22], a variety of diseases and pests such as rust, and contribute to complex attributes such as yield [22–25]. As a result, in addition to being a modern wheat progenitor, Aegilops speltoides has shown promise as a genetic asset for wheat development. Furthermore, the potential of such a wild progenitor to respond to a wide range of climatic conditions, including drought and heat zones, infection hotspots, and nutrient-deficient environments, has raised its worth as a resource for breeding.

Heat stress transcription factors (HSFs) activate practically all heat shock genes, which are critical components of heat stress genes, thereby defending against heat stress [26]. In response to heat stress, some heat-inducible genes that represent proteins called heat shock proteins, which protect proteins within the cell from denaturation, are up-regulated. [27]. In reaction to heat, transcription is activated by the presence of conserved heat shock elements (HSEs) in the promoter region of heat stress genes.

The plant heat shock factor gene family is diverse and plays a salient function in transcriptional regulation during heat stress [28], and HSFs have been classified into three groups based on the characteristics of their oligomerization domains: classes that are represented by A, B, and C HSFs. Different types of HSFs have different properties; for example, A-type HSFs can transactivate because they contain a C-terminal AHA-type activating domain, whereas B-type HSFs are being put forward to be either co-activators or repressors. Wheat includes 56 duplicates of known HSFs; pepper has 25, Arabidopsis has 21, and tomato has 24 copies. [29] [30] [31][32]

All HSFs possess two highly conserved domains: the oligomerization domain (OD), which possesses the bipartile heptad recite trends that form the hydrophobic associated region (HRA/B), and the DNA-binding domain (DBD), which connects via "heat-shock elements" (5′-nGAAnnTTCn-3′) that occur in the target gene's regulatory sequences [33, 34]. Previous studies have shown that HSFs are important in a range of abiotic stress situations, including heat stress in wheat. [35–36]. TaHSFA6b, previously known as TaHSFA2d in wheat, was successfully cloned for the first time, and its overexpression in Arabidopsis was found to increase thermotolerance [37]. TaHsfA6f is a transcription stimulator that actively affects the wheat genes TaHSP, TaRof1, and TaGAAP, and their genes control the network impacts on the thermotolerance in Triticum aestivum [38]

In the present work, we analyzed transcriptomes from two species, diploid progenitors (BB: Pau accession 3809) and bread wheat (AABBDD: cultiver c273), in order to examine gene expression profiles and monitor heat stress transcription factor genes. Furthermore, transcriptome sequencing of modern wheat and its ancestors provides a potential avenue for research in wheat development projects to improve heat stress tolerance. Intriguingly, despite careful screening, we were incapable of finding any autonomous platform specifically focused on heat stress-responsive genes in wild and modern wheat. This led us to develop an interactive web-based transcriptomic resource for wheat plants that can act as a thorough manual for understanding the comparative analysis of heat stress genes. According to our knowledge, the Heat Stress Genes in Wheat Database (HSGDB, http://www.hsgdb-wheat.com/) is a unique, first-of-its-kind database that was developed and is openly accessible. HSGDB provides information on sequence, homologs, protein families, gene ontology, metabolic pathways, orthologs in both varieties of wheat and Hsps, and visualization of various genes. HSGDB is a reliable source for researchers exploring the sequence and function of HSPs.

Materials and Methods

Plant Material

Wheat seeds derived from two distinct genotypes, i.e., diploids, were acquired from the Rehovot-based Weizmann Institute of Science in Israel. Dr. Parveen Chhuneja classified it as a temperature-tolerant accession,as well as a sample frame submitted to India's National Bureau of Plant Genetic Resources (NBPGR) Herbarium in New Delhi and a hexaploid from Punjab Agricultural University, Ludhiana. The Punjab Agricultural University preserved these two germplasm lines. (Table 1).

Table 1 Wheat accessions used in the study are

Genotypes	Genome	Accession No.
Aegilops speltoides	BB	As-Pau3809
Triticum aestivum	AABBDD	c273

Growth Conditions

Plants of the two wheat accessions were planted in flowerpots comprising a 3:1:1 (v/v/v) mixture of sand, soil, and peat and grown under 16 h of light at 22 °C with 60% relative humidity at day and 8 h of darkness at 18 °C with 80% relative humidity at night in a glasshouse at PAU, Ludhiana. The seedlings from the two wheat lines were 10–12 days old, removed throughout the day, and then snap-frozen alongside liquid nitrogen before being stored at -80 °C for later usage. To create one biological copy, seedlings from the two separate plants were combined.

RNA Isolation, Library development, and Illumina Sequencing

RNA was isolated via two biologically identical copies of frozen seedling tissue samples using an RNA isolation kit (Qiagen), following the directions provided by the manufacturer. The NanodropTM 1000 Spectrophotometer was used to determine the RNA concentration and purity (Thermo Scientific). The RNA-Seq library was built using RIN (RNA integrity numbers) ranging from 7.4 to 10.0, and high-quality samples of RNA at OD 260/280 values of 1.8 to 2.2 were considered. Four libraries, each containing two biological replications of the two genotypes, were produced and sent to a third party for transcriptome sequencing. For the four RNA samples, indexed TruSeq libraries were created, paired-end 100-bp sequencing was performed using HiSeq 2000, and the raw reads were obtained.The generated raw reads were submitted to the NCBI sequence read archives (SRA) bearing accession number PRJNA843543 and PRJNA767375 .

Quality control and filtering of raw reads

The generated paired-end (PE) raw reads from the two accessions (*As-Pau*3809 and cv. c273) were analyzed for quality evaluation and the exclusion of low-quality bases prior to *de novo* assembly of wheat transcripts. Following that, FastQC (v0.11.8) was used to assess the raw data quality [39], and Trimmomatic (version 0.38) using PE mode with "Phred 33" and MINLEN:50 parameters was used to trim low-quality reads and adapter-containing sequences before executing the assembly using raw reads [40]. FastQC v0.11.8 was used to re-examine the altered reads to ensure they were final, clean-quality reads.

De Novo Assembly

The high-quality PE reads of *As-Pau*3809 and cv. c273 obtained after filtering out low-quality reads and adaptor-containing sequences were *de novo* assembled using the Trinity assembler (v2.8.5) [41] with the default *k-mer*,K = 25, and based on the *de Bruijn* graph. Trinity has been reported to be effective in recovering full-length transcripts and spliced isoforms [42]. The assembled transcripts were used for functional annotation. The two main assembly approaches were evaluated using Trinitystat.pl, which provides standard statistics for evaluating *de novo* transcripome assemblies. These comprised the overall number of contigs created, the number of genes and transcripts, the N50 length of each assembly, and the number of contigs longer than 1,000 bps.

Abundance Quantification

The normalization method used for the abundance quantification of transcripts, in which the sequenced pair-end reads of *As-Pau*3809 and cv. c273 were individually aligned using Bowtie v1.1.2 [43] and the assembled transcripts were implemented with a perl script align_and_estimate.pl, was followed by RSEM v1.0 (RNA-Seq by Expectation-Maximization) software for the calculation of expression values in the form of FPKM (fragments per kilobase million reads) [44]. RSEM has the ability to produce both gene and isoform-level expression estimates. The feebly communicated isoforms were evacuated based on their expression values using the filter_low_expr_transcripts.pl parameter "-highest_iso_only in the Trinity script [45].

Protein coding regions and domains identification

Candidate coding regions in two assemblies were identified by TransDecoder v5.5.0 (http://trans decod er. github.io/).Additionally, it is used in the process of selecting the best single open reading frame (ORF) for each transcript. (-single best orf). PfamScan v. 1.6 [46] was employed to identify the protein structure families and protein domains for all the transcripts by seeking the Protein Family database and utilizing the protein sequences identified by TransDecoder. PfamScan was an HMM-based tool.

Evaluation of assembly completeness

To present an exhaustive and quantitative analysis of the degree of completeness attained for two assemblies [47] against deeply conserved single-copy orthologs. This was fulfilled utilizing the BUSCO (Benchmarking Universal Single-Copy Orthologs) v3.1.0 pipeline [48] and the OrthoDB plant database liliopsida_odb10 for class monocot containing 3,278 single-copy genes from 18 species [49]. We determined the number of complete (length given within two standard deviations of the BUSCO mean length), fragmented (partially recovered), duplicated (complete BUSCOs addressed by more than one transcript), and missing (not recovered) BUSCOs in each of the *two de novo* assemblies.

Orthologous genes

Orthofinder v2.4.0 [50] was used to locate genes that are orthologs between the two species. Orthofinder was started with default settings, taking the prediction from the transcoder with the longest orf of both *Ageliops speltoides (As-Pau 3809)* and *Triticum aestivum cv. (c273) as* the input. The alignment of single-copy orthologs was imported. Subsequently, single-copy orthologous genes were extracted,facilitating the identification of stress-relevant putative proteins.

Gene Ontology

The assembled transcripts from every set were compared for sequence homology through a standalone BLASTX using an e-value of 1e-5 with the openly accessible NR database (NCBI non-redundant protein sequences). The Blast2GO v5.2.5 tool (https://www.blast2go.com) was employed for generating Gene Ontology (GO) annotations over the genes in the biological process, molecular function, and cellular component categories devoid for each assembly. [51] Wego v2.0 (https://wego.genomics.cn/) was implemented to display the comparison of GO expressions for biological processes, molecular functions, and cellular components over all of the two assemblies. [52]

Metabolic Pathway Analysis

MapManVer 3.6.0RC1 (http://mapman.gabipd.org/) analysis was performed using our two transcripts as experimental input data. We will assign MapMan "Bins" to nucleotide sequences using the Mercator. In order to visualize the information, the output was placed into a mapping file in MapMan. By scanning several reference databases (BLAST, RPSBLAST, and InterProScan), the Mercator tool generates functional predictions. It then analyzes and compiles the search results for each input gene to suggest a functional Bin. [53] [54].

In silico expression analysis of genes

The heat stress transcription factor genes were identified from the annotation of the two assembled wheat varieties. The HSF transcripts that were chosen were BLASTn from the Ensembl plant database (https://plants.ensembl.org/Triticum_aestivum/Tools/Blast) and got the new transcript ids. Wheat transcriptome data obtained through the Wheat Expression Browser expVIP (http://www.wheat-expression.com/) was used to investigate the expression patterns of selected transcripts, and the expression values were extracted in the form of TPMs [55]. The database contains RNA-Seq datasets obtained from various tissues (root, spike, grain, stem, and leaf) from the common hexaploid wheat varieties Azhurnaya and Chinese Spring at multiple stages of development. The resulting values were subsequently used to create heatmaps at various folds.

Construction and Implementation of the HSGDB Database

The two wheat species used to build the database (**HSGDB**) were *Aegilops speltoides* (As-Pau3809) and the modern wheat cultivar (cv. c273), *Triticum aestivum*. The HSGDB website is free and accessible to all users without the need for a login. PHP, CSS, HTML, and JavaScript were used to create web interfaces. The XAMPP software was utilised to compile all of the data. Apache assembled the data and stored it in the MySQL relational database (v5.7). The HSGDB website is located at http://www.hsgdb-wheat.com/. All database material, including species information, data, heat stress genes, orthologs, and gene ontologies, is compiled sequentially. The database contains all of the information about the graphs, pie charts, and functional annotations of the two varieties

Results

Quality control and filtering of raw reads

We used Illumina sequenced libraries for As-Pau3809 and cv. c273 and obtained raw reads of 27,991,059 for As-Pau3809 and 15,044,857 for cv.c273. The FastQC (v0.11.8) shows adapter contamination and low-quality raw reads. Trimmomatic (version 0.38) was used to remove the adapter content from bases that had poor quality. Sequences with lengths below 50 were discarded, and the remaining sequences were named "clean reads". Thus, generating 21,876,692 clean reads from the As-Pau3809 transcriptomes and 12,684,414 clean reads from the c273 transcriptomes (Table 2) **summarizes** the findings of pre-assembly processes.

Table 2 Basic statistics of the pre-assembly process and adapters trimmed with Trimmomatic.

Feature\Condition	As-Pau3809	c273
Raw Reads	27,991,059	15,044,857
Length (bp)	100	100
Reads after trimming of adapters	21,876,692 (78.16%)	12,684,414 (84.31%)
Left only reads	5,727,409 (20.46%)	2,192,232 (14.57%)
Right only reads	191,361 (0.68%)	89,886 (0.60%)
Improper pairs	195,597 (0.70%)	78,325 (0.52%)
Length (bp)	50-100	50-100

De Novo Assembly

In this study, a total of 27.9 million paired reads were analyzed, which is considerably more than the 7.4 GB required to assemble the diploid transcript of *Aegilops speltoides* accession As-Pau3809, and

15.6 million pair-end reads (4.0GB) were used to assemble the hexaploid transcriptome of *Triticum aestivum cultivar* c273 of wheat. After trimming and removing low-quality reads, As-Pau3809 got 21.8 million (5.8 GB) clean reads, and c273 got 12.6 million (3.3 GB) clean reads. All clean reads of As-Pau3809 and c273 were de novo assembled using Trinity v2.8.5. In As-Pau3809, the total number of transcripts was 135,793, with genes of 94805 bp. The median contig length of this transcript was 419 bp, while the N50 was 1402 bp. While in c273, with a total of 103,339 transcripts, which contain 65085 genes. The median contig length of these reads was 406 bp, with 1226 bp of N50. The de novo assemblies' statistical overview is provided in (Table 3).

Table 3 Trinity assembly statistics

Evaluation parameters	As-Pau3809	c273
The counts of total genes and transcripts		
Total genes	94805	65085
Total transcripts	135793	103339
GC percent	47.59	47.22
The statistics based on all transcript contigs		
N50 Contig	1402	1226
Median length contig	419	406
Average contig	806.55	743.39
Total Assembled bases	109524209	76821598
The statistics based on only the longest isoform per gene (unigenes)		
N50 Contig	891	767
Median length contig	338	322
Average contig	602.83	554.77
Total assembled bases	57151402	36107475

Abundance Quantification Results

The normalization method used for the abundance quantification of transcripts based on alignment by RSEM The expression level varies significantly; with 94805 genes, 135793 isoforms were expressed in As-Pau3809, and genes expressed in cv. c273 were 65085, isoform 103339. We assessed the FPKM value for each transcript and finally determined 1,24,689 and 97169 transcripts that have been consistently expressed (FPKM value >= 0.5) in As-Pau3809 and c273, respectively. Moreover, the RSEM approach was used to detect expression levels, and only isoforms that were highly expressed were kept for future use. Following this filtering stage, between 58 and 66% of the generated transcripts were retained (Table 4).

Table 4 Total number of contigs filtered and FPKM>=0.5

Species Id	No of genes	Isoforms	FPKM>= 0.5	Highly expressed isoforms
As-Pau3809	94,805	1,35,793	1,24,689	90,600 (66.72 %)
c273	65,085	1,03,339	97,169	60,525 (58.57%)

Protein coding regions and domains identification

TransDecoder was used to determine the coding domain sequences (cds) for each species. The single-best open reading frame (ORF) for every transcript was first chosen, and only transcripts longer than 200 bp were kept. For As-Pau3809 and c273, respectively, there were a total of 53,763 and 42,694 protein-coding transcripts among the final non-redundant transcripts. The transcripts identified as ORFs (28.06%) were 5′

partial (9.14%) were 3′ partial, (19.90%) were internal and (42.89%)complete in AS3809 whereas in c273 5′prime partial (42.64%), 3′prime partial (8.59%) internal (23.99%) and complete (24.76%) (Figure 1).

PfamScan classified transcripts into coiled, families, motifs, domains, and repeats, with c273 being the most abundant and As-Pau3809 being the least abundant (Table 5)displays the annotation statistics of both of the assemblies that were retrived.

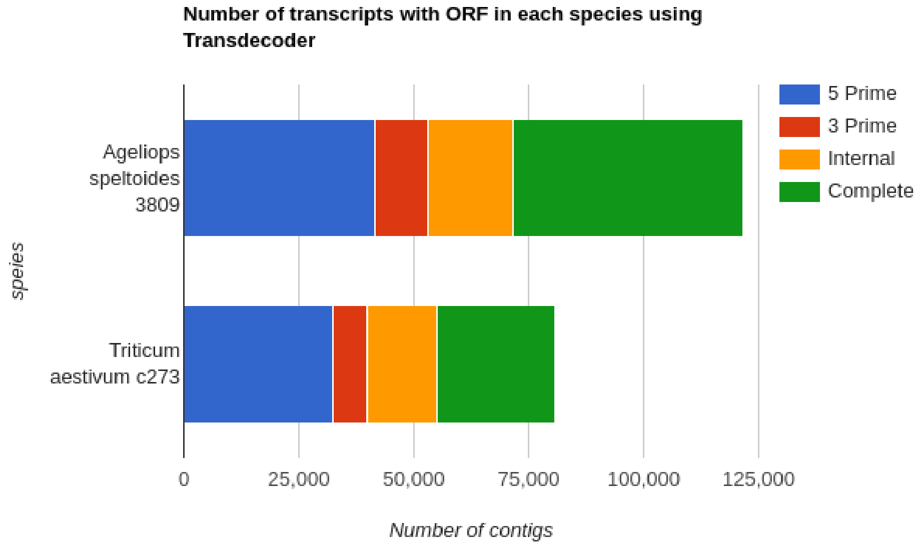


Figure 1. The TransDecoder pipeline was used to count the total number of transcripts in two species that had open reading frames. The term "complete ORF" describes sequences that contain both the beginning codon and the stop codon. Sequences that are "5 prime" have the start codons, while sequences that are "3 primer" lack the start codon. The phrase "internal ORF" describes sequences that are devoid of both the start and stop codons.

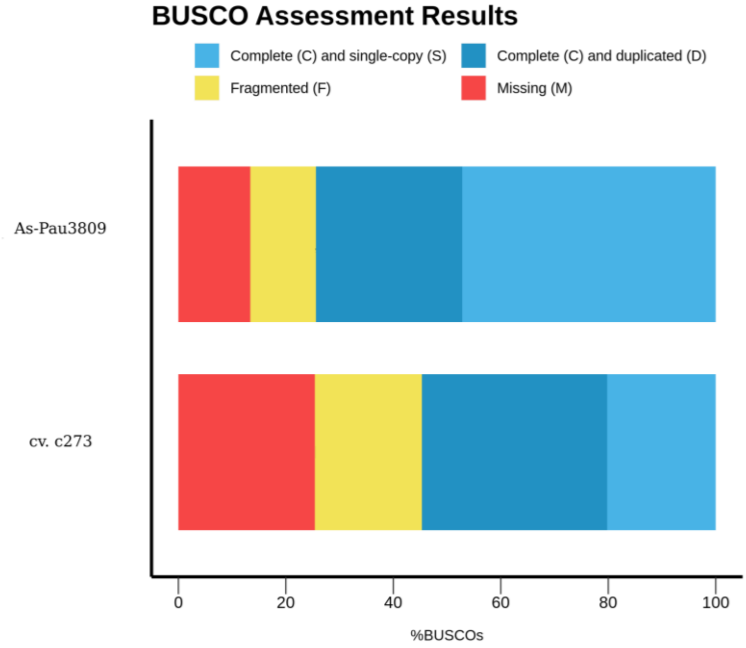
Table 5. Summary of two transcripts with ORF and PfamScan.

Orf	**Ageliops speltoides (As-Pau 3809)**	**Triticum aestivum cv. c273**
3 prime- partial	4916 (9.14%)	3669 (8.59%)
5 prime-partial	15087 (28.06%)	18207 (42.64%)
Complete	23059 (42.89%)	10575 (24.76%)
Internal	10701 (19.90)	10243 (23.99%)
PfamScan		
Total number of domains	11409	22696
Total number of families	8679	16527
Total number of repeats	2085	3950
Total number of coiled	38	96
Total number of motif	208	341

Assessment of assembly completeness

Assemblies were compared to set completeness using the BUSCO, which found that a preponderance of liliopsida core genes had been strongly reclaim in all two transcript assemblies. Notably, of the 3278 single-copy orthologs found, we found 74.4% completely . Only amid 13 and 25.4% of the 3278 single-copy orthologs were restricted as missing from both assemblies, indicating very good coverage and the higher quality of the assembly of the transcriptomes. Out of 3278 single-copy ortholog genes common to the liliopsida lineage there were 47.2 and 20.2% complete single-copy BUSCOs, 27.2 and 34.5% duplicated BUSCOs, 12.2 and 19.9% fragmented BUSCOs, and 13.4 and 25.4% missing BUSCOs respectively for As-Pau*3809* and c*273*.As expected, considering the high quality of the dataset, recuperation for 'complete' was higher for As-Pau3809 with 74.4 % and 'complete and duplicated' , 'Fragmented 'and 'missing' busco was considerably higher for c273, with 34.5% , 19.9% and 25.4% were duplicated, fragmented and missing BUSCOs (Table 6 and Figure 2).

Figure 2



Busco analysis of As-Pau3809 and cv.c273

Table 6 Summary of the Benchmarking Universal Single-Copy Orthologs (BUSCO) search for liliopsida compared to 3278 single-copy orthologs

BUSCO statistic	As-Pau3809	c273
Complete BUSCOs	2437(74.4%)	1793(54.7%)
Complete-single-copy BUSCOs	1547(47.2%)	662(20.2%)
Complete–duplicated BUSCOs	890(27.2%)	1131(34.5%)
Fragmented BUSCOs	401(12.2%)	652(19.9%)
Missing BUSCOs	440(13.4%)	833(25.4%)

Orthologous genes

Longest orf from transdcoder from 2 trancripts were used to create orthogroups using OrthoFinder v2.4.0 This program assigned 162,469 total genes out of 202,309 (80.3% of total) to 44,218 orthogroups . Mean orthogroup size is 3.7 proteins, There were 24,686 orthogroups that included all species, and 7,785 of them only contained single-copy genes. Further study of these unique single-copy genes might give insight into the identification of heat stress realted genes (Table 7).

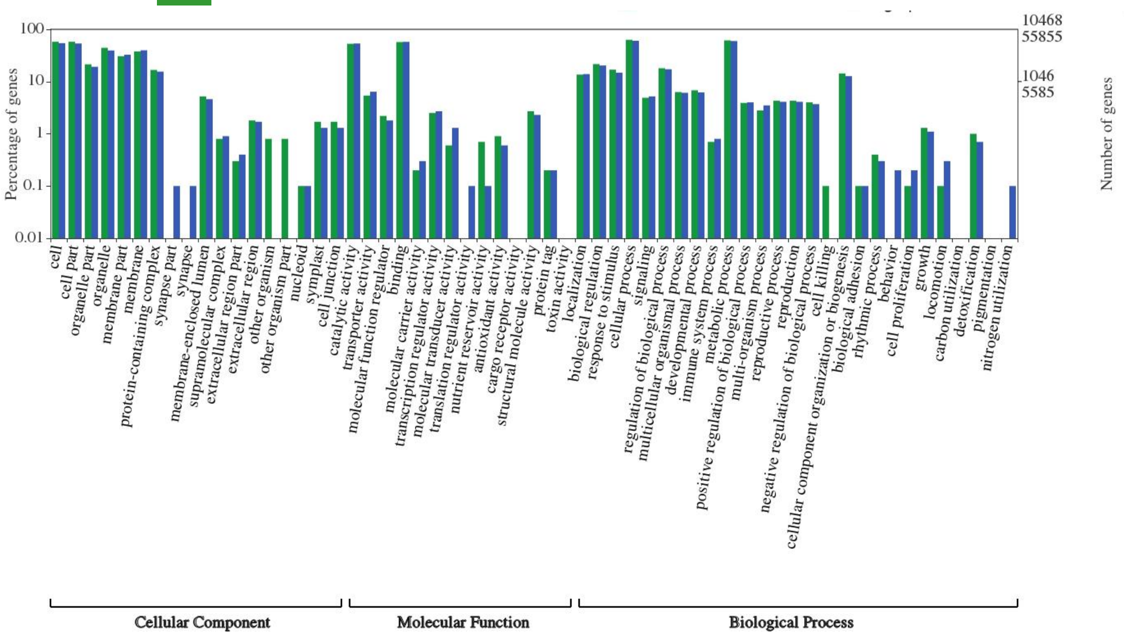
Table7. OrthoFinder statistics for As-Pau*3809* and *c273*

OrthoFinder statistics	**Number**
Total no of genes	202309
Total no of genes in orthogroups	162469
Total no of unassigned genes	39840
% of genes in orthogroups	80.3
% of unassigned genes	19.7
Total No of orthogroups	44218
Total no of species-specific orthogroups	19532
Total no of genes in species-specific orthogroups	64299
% of genes in species-specific orthogroups	31.8
The Mean of orthogroup size	3.7
The Median of orthogroup size	3.0
G50 of assigned genes	4
G50 of all genes	3
O50 of assigned genes	12643
O50 of all genes	17942
Total No of orthogroups with all species present	24686
Total No of single-copy orthogroups	7785

Gene ontology

Based on the BLASTX results, Blast2GO annotation resulted in functional characterization of 55,855 and 48,159 transcripts in As-Pau3809 and c273, respectively.Every assembly displayed a varied collection of Gene Ontology, demonstrating that the biological processes, cellular components, as well as molecular functions were each accurately reflected. The three major categories were subdivided into further 53 GO functional subcategories that were evenly distributed throughout the two transcriptomes. Genes the specify cellular and metabolic functions, as well as reaction to stimuli and biological control, were notably present in all of the assemblies within the realm of biological processes.Despite the fact that percentage of genes **encoding** binding and catalytic **activity**in the molecular function category was much higher. Among the cellular components, proteins associated with cells and cellular sections seemed the most prevalent, subsequent to organelles and cellular components Figure 3.

Figure 3 Gene Ontology of both the variety shows percentage of genes with total number of genes 🟦 As-Pau3809 🟩 cv. c273



Metabolic Pathway Analysis

To identify the pathways impacted by heat stress, we used the MapMan tool to map transcripts to various functional categories. All transcripts were associated with 35 pathways(Figure 3) They were significantly enriched on signalling (3.64%), stress (1.54%), misc (1.59%), cell (1.55%), and deveopment (1.13%) except for RNA (3.63%), protein (6.39%), and transport (2.04%) in AS3809 (Figure 3 A). whereas in c273 transcripts were significantly enriched on signalling (2.63%), stress (1.39%), misc (1.66%), cell (1.72%), and deveopment (1.27%) except for RNA (4.81%), protein (8.13%), and transport (1.93%) (Figure 3B). Therefore, these process genes were **detailed** in Figure 4.

Figure 3 A Enriched Pathways shown by As-Pau3809

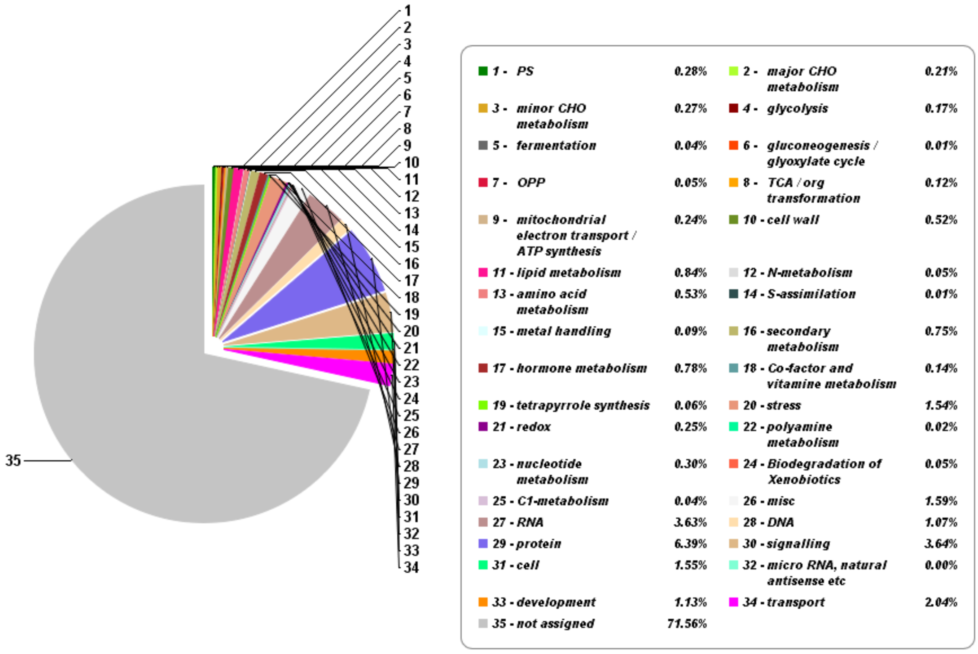
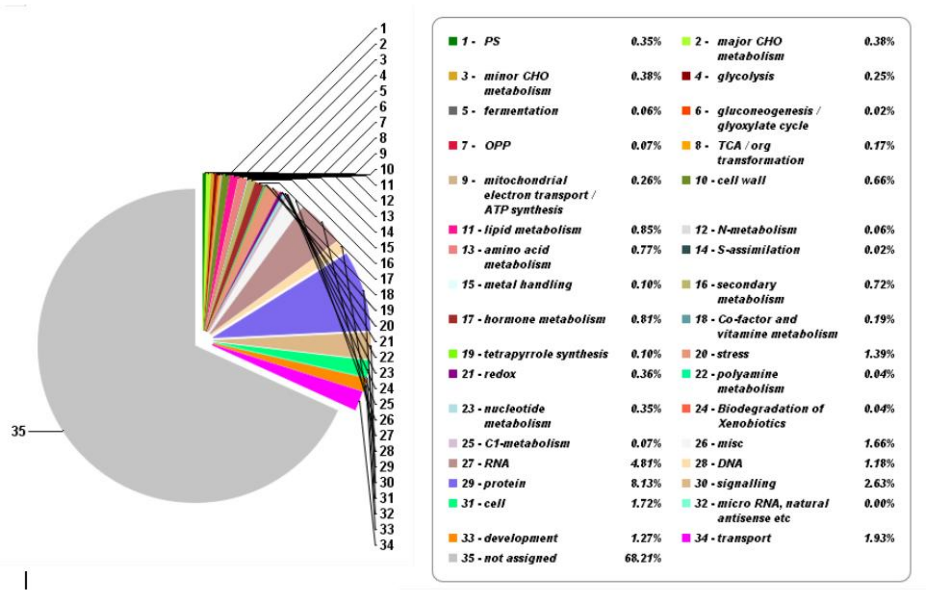


Figure 3 B Enriched Pathways shown by cv. c273



Functional categories of transcripts using MapMan 3.6 with initial preferences, Numbers 1-34 on the pie chart represent different pathways. Comprehensive details on gene numbers and pathway names are also shown in **Figure 3 A** and **B** in both the varieties of wheat.

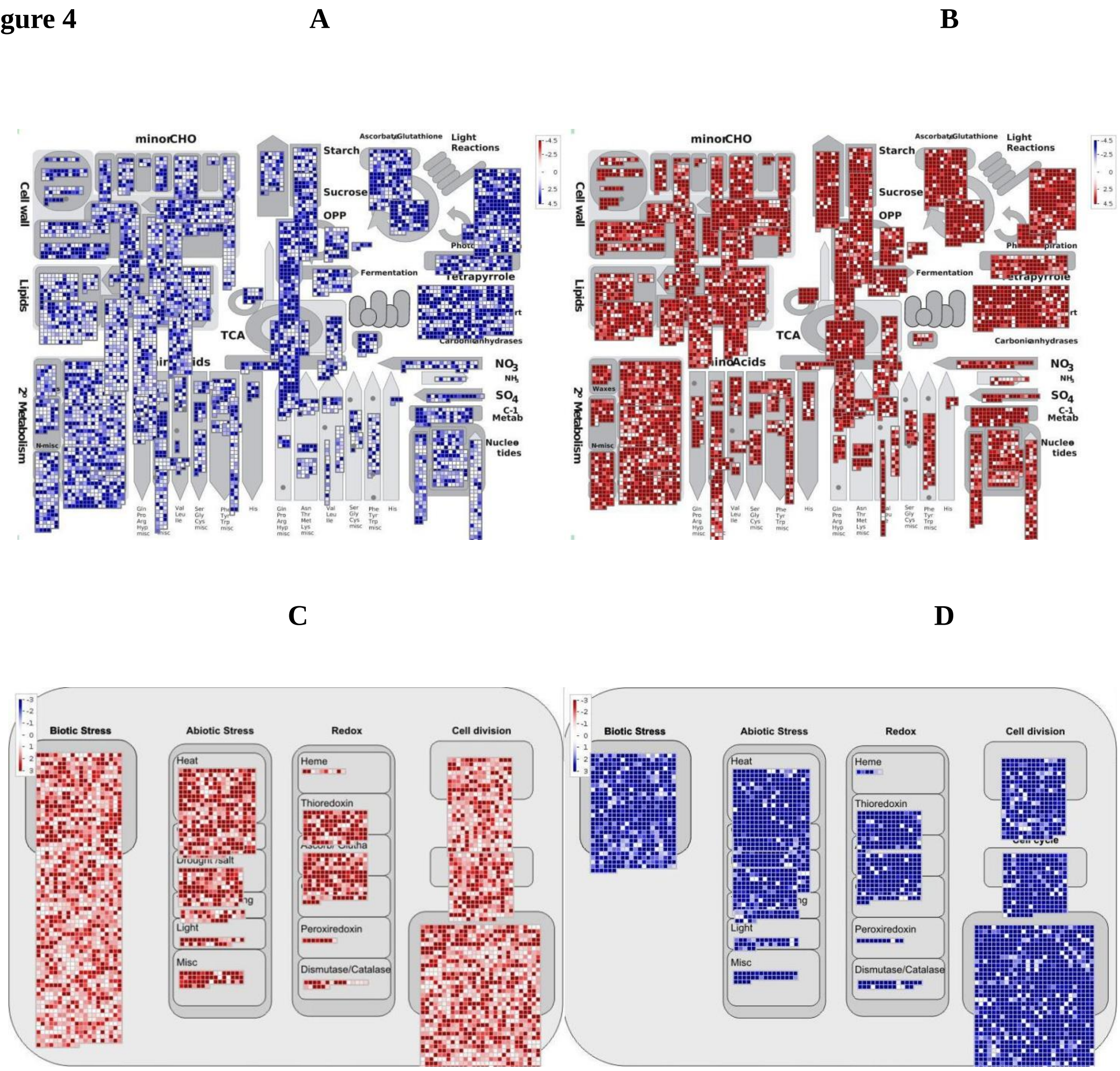
The visual comparison of the transcripts from the two varieties was examined, and MapMan software Ver. 3.6.0RC1 was used to import the acquired levels of expression for all of the present genes .Characteristic expression patterns can be found, as shown by the "metabolism" and "cellular response overview"in Figure 4.

The expression levels of each transcript shows the total number of mapped transcripts were 1,43,088 in As-Pau3809 whereas 1,08,429 mapped transcripts in c273 The metabolic pathways of two comparative groups of transcripts were visualized by MapMan (Figure 4A). In As-Pau3809 the most active genes in terms of gene expression were invloved in the PS.lightreaction (bin:1.1), mitochondrial electron transport / ATP synthesis (bin:9), PS.calvin cycle (bin:1.3) redox.ascorbate and glutathione (bin:21.2) and secondary metabolism.phenylpropanoids (bin:16.2). To a lesser extent amino acid metabolism (bin:13.1), secondary metabolism pathway (bin:16.10)and the minor CHO metabolism (bin: 3.2 were also among the ones with more highly expressed genes.Comparatively, a number of functional categories that are listed in the c273 metabolic pathways highly active genes in (Figure 4B) expression levels are major CHO metabolism.synthesis.starch(bin:2.1.2),TCA/orgtransformation.TCA(bin:8.1),PS.lightreaction (bin:1.1),cell wall.cellulose synthesis (bin:10.2) and C1-metabolism (bin:25)

Many of the genes depicted within the MapMan representation of cellular responses were continually upward or downward regulated throughout the research, indicating their crucial function in the response to heat stress shown in (Figures 4C and 4D).The total number of heat stress gens in abiotic stress in As-Pau3809 371 and 534 in c273.we found both common and specific BINs in As-Pau3809 and c273 were stress.abiotic.heat (bin:20.2.1), redox.ascorbate and glutathione (bin:21.2), and redox.thioredoxin (bin:21.1) and development (bin:33) In contrast less active in As-Pau3809 were cell division (bin:31.2) and stress biotic (bin:20.1) in c273.

A.Metabolic overview of As-Pau3809 in blue color , B shows the metabolic pathways in c273 in red color.
C Cellular response of As-Pau3809 in red color, it shows abiotic stress include heat also D cellular response of c273 in blue color with heat stress in abiotic stress.

Figure 4



In silico expression analysis of heat stress transcription factor genes

The recognition of the three distinct HSF (heat stress transcription factor) categories—A, B, and C—came about as a result of the transcripts. From As-Pau3809, a total of 43 HSF transcripts were obtained, and from the cv.c273 assembly, a total of 41 HSF transcripts were obtained. The chosen HSF genes share a high degree of sequence similarity (96–98.5%) with the two species of wheat, and Pfam analysis revealed that they had a higher proportion of HSF DNA binding domains and a few families with coiled, disoredered predictions. These HSF transcripts have been annotated using the Gene Ontology by Blast2GO, TPM, FPKM, and PfamScan and are given in (Supplementary Table 1). The selected 43 HSF transcripts from As-Pau3809 and 41 HSF transcripts from c273 were used for expression analysis. A BLASTn search of heat stress transcription factor transcripts was performed with the Triticum astivum species from the Ensembl plant database and got the new transcript ids (Supplementary Table 1). In addition to this,the expression values were extracted using expVIP (http://www.wheat-expression.com/) as Transcript per millions (TPM). The levels of expression in various tissues (root,seedling, spike, shoot and leaf) and developmental stages , along with heat stress, were extracted from both accession of wheat and depicted as a heatmap (Figure 5 and 6).

In the heat-resistant accession As-Pau 3809, the highest expression was seen for TraesCS6A02G407700.1, with a fold change from the 7 range. No expression was seen in TraesCS5D02G445100.2, TraesCS7B02G406000.1, TraesCS7D02G498800.1, and TraesCSU02G095900.3 in all the tissues and developmental stages.

In terms of tissue-specific expression response, TraesCS6A02G407700.1 was abounding in grain and predominantly expressed in these tissues, including spikes, leaves, shoots, and roots. The transcript exclusively expressed in grain was TraesCS6A02G407700.1 at the reproductive stage, along with spike at the vegetative and reproductive stages. Transcript TraesCS6B02G237400.1: with high induction for leaves and shoots at the reproductive stage as compared to the seedling stage. The Highest expression occurs at roots in seedling stage induction was seen for TraesCS6B02G237400.1, TraesCS6A02G407700.1.

While no significant changes were observed for abiotic stress in TraesCS7D02G412100.1, TraesCS7B02G272500.1, TraesCS7D02G367700.1, TraesCS7A02G355300.1, and TraesCSU02G127300.1, they were discovered to be significantly suppressed by ~ 4-5 fold, Under heat stress conditions at 6 hours, expression in TraesCS7B02G168300.2 was highest at more than ~6 fold, followed by TraesCS7D02G270600.1, TraesCS6B02G237400.1, TraesCS7D02G270600.2, and TraesCS6A02G407700.1, demonstrating that these transcripts are stimulated under 6 hours of heat stress in plants and also have a regulatory role. After 1 h of heat stress, the expression of *TraesCS7B02G168300.2 and TraesCS7D02G270600.1* was found to be the highest amongst all the heat stress genes, with more than ~7 fold . However, under prolonged heat stress (6 h), the expression values were relatively low in the expression database compared to those seen under brief heat stress (1 h).
While in heat-susceptible cv. c273, at a tissue-specific region abundant expression above 7-fold is shown in TraesCS6A02G303000.2 in roots, leaves, shoots, and spikes. The highest expression was seen for TraesCS7D02G527500.1 in roots with fold change in the 7 range. In addition, no expression overall in five tissues(roots, leaves/shoots, spike, and grain) was observed in TraesCS7D02G270600.2, TraesCS7D02G261700.4,TraesCS7D02G261700.2,TraesCS7D02G261700.3,TraesCS7A02G535100.1,TraesCS5D02G445100.2.

At different developmental stages, the transcript exclusively expressed in roots, leaves, and shoots was TraesCS6A02G303000.2 at the seedling, vegetative, and reproductive stages. Transcripts show TraesCS7D02G527500.1 with high induction for roots at the reproductive stages as compared to vegetative stages. The Highest expression at spike in vegetative stage induction was seen for TraesCS7A02G535100.1 along with TraesCS4A02G027700.1 at leaves/shoot at reproductive stage.
For abiotic stress, no major changes were observed for TraesCS7B02G168300.2,TraesCS7A02G270100.1, TraesCS7D02G270600.2,TraesCS4D02G239600.1were found to be significantly down regulated by ~ 4–6 fold under high-level stress disease.

Under heat stress conditions at 6 hours, expression in TraesCS7B02G168300.2 was highest at more than ~6 fold, followed by TraesCS7D02G270600.1 and TraesCS7A02G270100.1, demonstrating that these transcripts are stimulated in plants under 6 hours of heat stress and also have a regulatory role. After 1h under heat stress conditions, the expression of *TraesCS7A02G360400.1 and TraesCS7B02G267300.1*was found to be the highest amongst all the heat stress genes, with more than ~9 fold. However, the expression values determined under extended heat stress (1 h) were relatively higher in the expression database than those seen under brief heat stress (6 h).

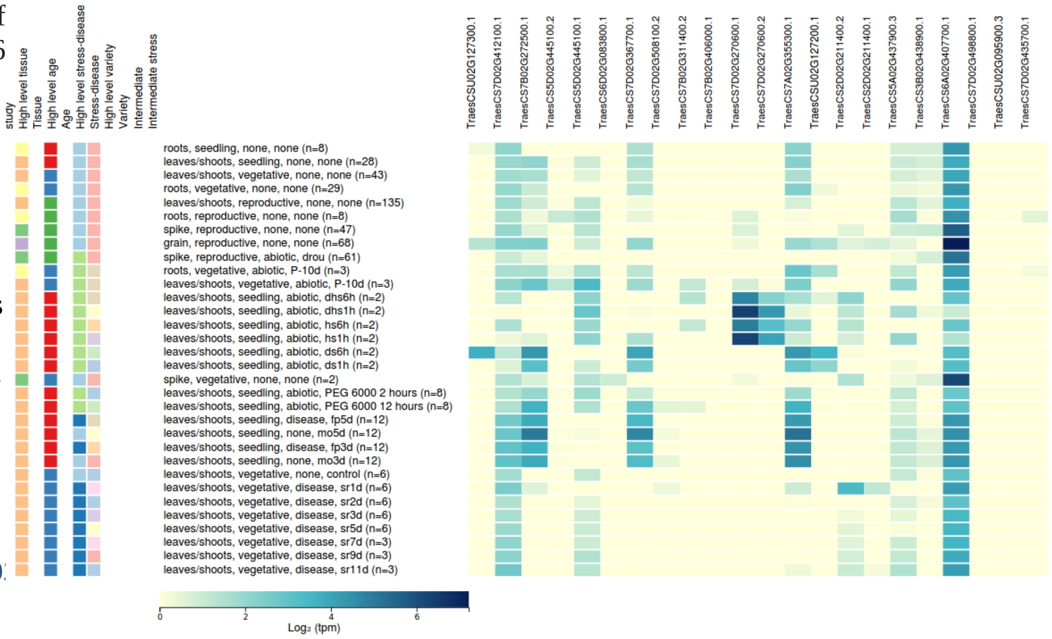


Figure 5 As-Pau3809 heat map to show the expression of transcripts at different development stages

Figure 6 cv. c273 heat map to show the exprssion of transcripts at different development stages

Construction of HSGDB content and web interface

HSGDB offers Nine functional sections: “Species”, “Data”, “Heat Stress Genes”, “Quality check”, “Graph”, “Orthologs”, “Gene Ontology”, “Tools” and “Search” centralized as Overview and one sections termed “About” for a detailed introduction (Figure 7a). These created functional modules can work independently as well as together.

Species

Under the “Species” drop-down menu, HSGDB provides information about the Taxonomic hierarchy, in terms of crop height, genome size, chromosome number and ploidy of two species.

Data

HSGDB provides an easy-to-use web interface in “Data” menu where users have access to two functional sidebar : ‘Transcripts overview’ and ‘Transcripts Assembly’ (Figure 7b). The ‘Transcripts overview’ provides Total no of Reads, Type of Reads, Contig N50 and Total Genes from both the varieties. In ‘Transcripts Assembly ’ Trinity assembler (v2.8.5) was utilized to perform de novo transcriptome assembly of two libraries. Users can access the assembled Transcripts of both the variety by clicking the link. It shows the sequence id with sequence.

Heat Stress Genes

“Heat Stress Genes” mode split in three slide bars, first and second shows the heat stress genes in two variety presented in the form of a paginated table. Heat stress genes records can be viewed by clicking the sequence ID for further analysis. Individual genes records comprise heat shock proteins, functional annotation,protein family detail, Transcript per million(TPM), Fragments per kilobase of transcript per million mapped reads (FPKM), sequence and metabolic pathways details in th eform of mercator results. Last slidebar shows the heatmap of the heat stress genes (Figure 7c).

Quality check

The “Quality check” drop-down menu shows the clean reads of the transcripts using fastQC. The reads shows Perbase sequence quality, Per sequence qualityscores, Perbase N content and Adapters.

Graphs

“Graph” mode split into three sidebars ‘Heatmap’, ‘Busco’ and ‘ Transdecoder’ . Heatmap means the expression of highly and low expreesed genes in differnet tissue developmental stages and heat stress conditions (Figure 7d). BUSCO means Benchmarking Universal Single-Copy Orthologs. It estimates the completeness of assembled transcriptome.. It is based on the notion, orthologs of single-copy which should be extremely conserved across the related species. Transdecoder predicts ORF and coding regions in the transcriptomes of both the variety.

Orthologs

“Orthologs” dropdown menu shows the total number of orthologs in both the variety in the form of list. OrthoFinder v2.4.0 is used to find the total 7,785 orthologs (Figure 7e).

Gene ontology

Gene ontology is divided into various slide bars. GO shows the biological process, cellular components, molecular function bars. These contains the Go id, Term and number of genes, it also present the go graphs along with piecharts of biological process, cellular components and molecular function, bar charts are presented in the last , it contains Annotation Distribution, GO mapping distribution, go level distribution and total count bar chart. (Figure 7f).

Tools

In HSGDB, other than a database, the extesive analysis tools menu was used . “TOOLS” contains FastQC , Trimmomatic ,Trinity,blastx, BUSCO , Transdecoder, PfamScan , Mercator and Blast2Go for the analysis of transcriptomic data.

Search

The “SEARCH” mode provides a easily accesible and user-friendly web interface to query SequenceID gives details of the sequence, Best Blast Homologs,Function and Gene Ontology (Figure 8a). Protein Family searches for sequence id, Accession, Domain/Family (Figure 8b), Gene Ontology searches by Sequene Id, Gene Ontology Function and Gene Ontology Id (Figure 8c) and Metabolic Pathways search with sequence Id are implemented in HSGDB (Figure8d).

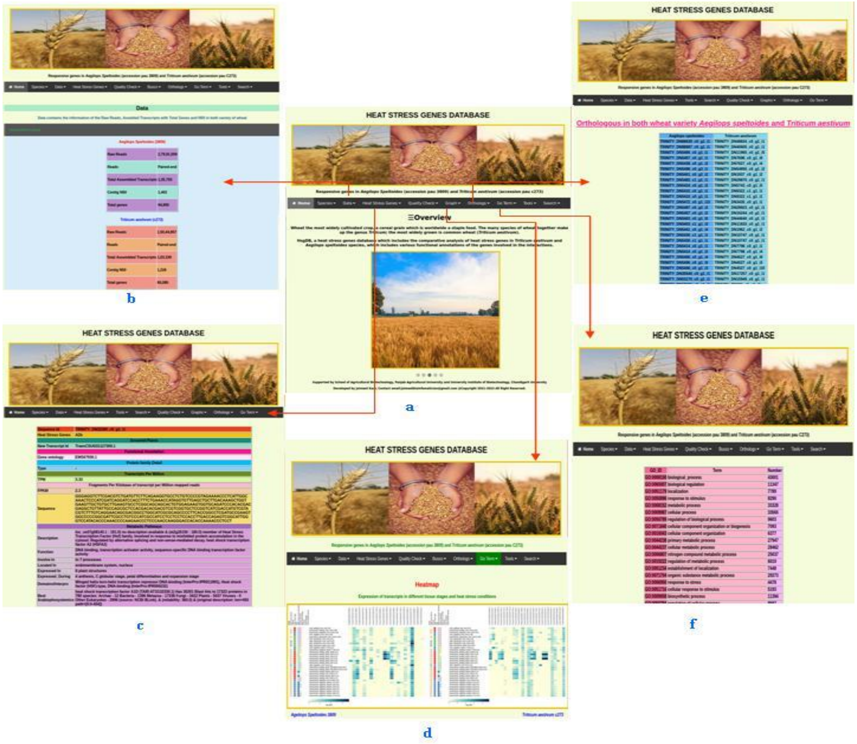
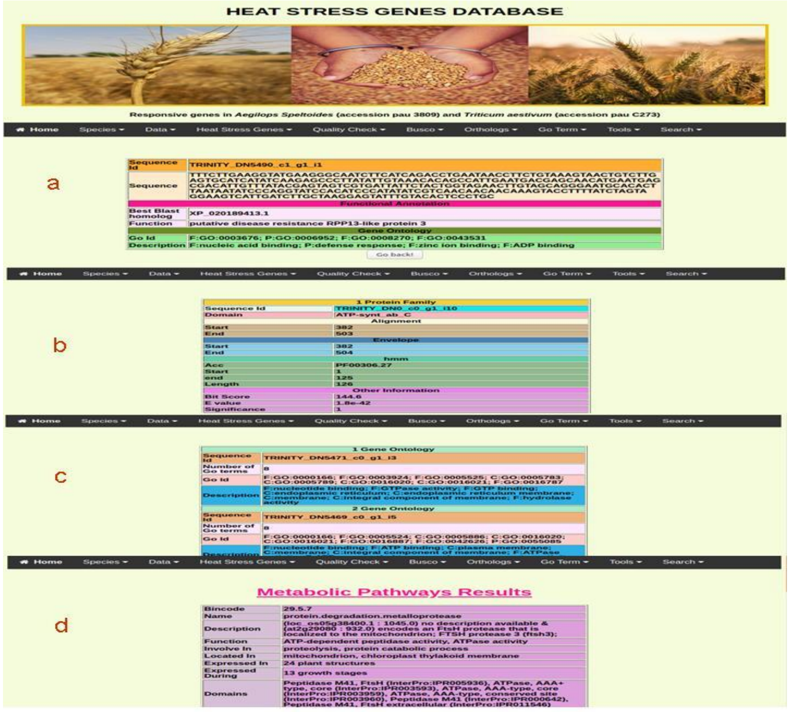


Figure 7 Screen dumps of HSGDB(a) Home page (b)Data information (c) Heat stress genes details (d) Heatmap under graphs (e) Orthologs of both variety (f) Gene ontology



Sequence Id	TRINITY_DN5490_c1_g1_i1
Sequence	TTTCTTGAAGGTATGAAGGGCAATCTTCATCAGACCTGAATAACCTTCTGTAAAGTAACTGTCTTG AGTGCATCATATCAAGAGCCCTTATATTGTAAACACAGCCATTGAATGACGAGCAACATGAATGAG CGACATTGTTTATACGAGTAGTCGTGATTATTCTACTGGTAGAACTTGTAGCAGGGAATGCACACT TAATAATATCCCAGGTATCCACATCCCATATATCGTCAACAACAACAAAGTACCTTTTATCTAGTA GGAAGTCATTGATCTTGCTAAGGAGTATTTGTACACTCCCTGC
Functional Annotation	
Best Blast homolog	XP_020189413.1
Function	putative disease resistance RPP13-like protein 3
Gene Ontology	
Go Id	F:GO:0003676; P:GO:0006952; F:GO:0008270; F:GO:0043531
Description	F:nucleic acid binding; P:defense response; F:zinc ion binding; F:ADP binding

1 Protein Family	
Sequence Id	TRINITY_DN0_c0_g1_i10
Domain	ATP-synt_ab_C
Alignment	
Start	382
End	503
Envelope	
Start	382
End	504
hmm	
Acc	PF00306.27
Start	1
end	125
Length	126
Other Information	
Bit Score	144.6
E value	1.8e-42
Significance	1

1 Gene Ontology	
Sequence Id	TRINITY_DN5471_c0_g1_i3
Number of Go terms	8
Go Id	F:GO:0000166; F:GO:0003924; F:GO:0005525; C:GO:0005783; C:GO:0005789; C:GO:0016020; C:GO:0016021; F:GO:0016787
Description	F:nucleotide binding; F:GTPase activity; F:GTP binding; C:endoplasmic reticulum; C:endoplasmic reticulum membrane; C:membrane; C:integral component of membrane; F:hydrolase activity
2 Gene Ontology	
Sequence Id	TRINITY_DN5469_c0_g1_i5
Number of Go terms	8
Go Id	F:GO:0000166; F:GO:0005524; C:GO:0005886; C:GO:0016020; C:GO:0016021; F:GO:0016887; F:GO:0042626; P:GO:0055085
Description	F:nucleotide binding; F:ATP binding; C:plasma membrane; C:membrane; C:integral component of membrane; F:ATPase

Metabolic Pathways Results

Bincode	29.5.7
Name	protein.degradation.metalloprotease
Description	(loc_os05g38400.1 : 1045.0) no description available & (at2g29080 : 932.0) encodes an FtsH protease that is localized to the mitochondrion; FTSH protease 3 (ftsh3);
Function	ATP-dependent peptidase activity, ATPase activity
Involve In	proteolysis, protein catabolic process
Located In	mitochondrion, chloroplast thylakoid membrane
Expressed In	24 plant structures
Expressed During	13 growth stages
Domains	Peptidase M41, FtsH (InterPro:IPR005936), ATPase, AAA+ type, core (InterPro:IPR003593), ATPase, AAA-type, core (InterPro:IPR003959), ATPase, AAA-type, conserved site (InterPro:IPR003960), Peptidase M41 (InterPro:IPR000642), Peptidase M41, FtsH extracellular (InterPro:IPR011546)

Figure 8 Screen dumps of HSGDB Search Dropdown Menu Details (a) Sequence Id Information (b) Protein details (c) Gene Ontology (d) Metabolic Pathways

Discussion

The productivity and quality of wheat crops around the world have been severely impacted by changes in the climate and other environmental factors, such as heat stress and water scarcity [56] and [57]. Heat stress affects around 57 percent of the wheat planted area in emerging nations, making it the principal abiotic factor endangering global wheat output and food security [58]. During the reproductive (anthesis and post-anthesis) stages as well as the grain filling stage, which are crucial for maintaining total crop output and quality, wheat is particularly susceptible to heat stress. [57]. In order to maintain yield and quality during a heat stress response, numerous biochemical and molecular machines must interact and be under control. [59]. It is crucial to look into how many important genes express themselves and how heat stress affects wheat during the grain filling stage. Next-generation sequencing (NGS) technology utilising RNA-seq offers a very practical and affordable way to precisely identify significant genes, transcript level changes, and regulatory factors. [60]. Wild species have a variety of unusual genes and alleles that have almost completely disappeared in farmed wheat as a result of domestication and breeding, which is why scientists are constantly drawn to them [61] . Ae. speltoides was discovered to be a very heat-tolerant species in earlier research, which can be used to increase wheat's thermotolerance [62]. As a result, our study aimed to both investigate the expression profiling of identified HSFs as well as to detect the distinct transcript from the selected accession As-Pau3809 by comparing it to grown hexaploid wheat cv.c273.

RNA-Seq Analysis of wild and cultivated wheat's

Trinity was implemented to execute a de novo transcriptome assembly on two transcriptomes in order to remove bias brought on by the lack of reference genomes for Ae. speltoides and T. aestivum. Despite the large differences in genome size (5.14–7.9 GB) [63] [64] and gene count (65,000–95,000) [65], the total number of expressed transcripts discovered in wild diploid and hexaploid wheat was 1,35,793 in As-Pau3809, which was slightly higher than the 1,03,339 transcripts obtained in hexaploid wheat c273. The assembled transcripts in tri. aestivum (K7903) ranged in size from 200 to 18,870 bp, had a N50 value of 473 bp, and were larger at 1,696,570 bp. [66], whereas bread wheat cv. (Zhoumai 18) yielded 142,221 transcripts with a N50 value of 1001 bp [67]. The N50 values acquired from both transcriptome assemblies were determined to be adequate for further investigation. [68] [69]. The presence of an elevated number of sequences that repeat in the non-expressing heterochromatin region with greater ploidies compared to diploid ancestors has been linked to the observed asymmetry between the size of the genome and expression of genes in polyploid species. . Furthermore, the observed tendency appears to be directly related to the recessive nature of certain genes and the functioning of the alleles during domestication and selective breeding. [70] [71]

The completeness assessment of the two transcriptomes using BUSCO analysis confirmed that the diploid wild wheat had the greatest recovery for retained single-copy orthologs, followed by hexaploid wheat, confirming As-Pau3809's high coverage and quality when compared to gene loss at higher ploidies,the domestication process and polyploidy events [72]**.** The data for the assemblies was sufficient to identify single-copy orthologs.[73], implying that the presentassemblies of wild and cultivated wheat species will supplement existing wheat resources.

In hexaploid, the observed percentage of mapped reads was around 54%, which is lower than the 78.5 percent achieved in T. aestivum [74] and the 84 percent acquired in T. turgidum [75]. Comprehensive annotation using BLAST2GO and MapMan indicates that the GO illustration of various divisions of molecular functions, biological processes,and cellular componentswas effectively documented and that **the** proportion of mapped GOterms exhibits an exaggerated GOprofile**.**[76].

HSF identification and expression studies

HSFs are crucial for plant responses to abiotic stress and are involved in a variety of molecular processes.Previous research in rice, tomato,maize, sorghum, has found 25 hsf genes Arabidopsis and chickpea discovered 21 and 22 Hsf genes, respectively [77-81] Previous research revealed 82 and 56 members of the wheat HSF family, respectively. [82,83], but we identified 526 in As-Pau3890 and 534 in c273 Hsfs in this study. To evaluate the variations in the number of discovered wheat HSF members, we compared the protein sequences found by Xue et al. [83] and Duan et al. [82] to the most recent wheat reference genome (IWGSC V1.1) [84].In a genome-wide analysis of wheat (Triticum aestivum L.), at least 82 Hsf members were discovered. The aforementioned findings support TaHsfA2-10's regulatory role in controlling the expression of Hsp genes during heat stress and indicate that various Hsf belonging to the identical subclass only affect the regulation of thermotolerance of particular Hsp genes. Additionally, HSFA2 improves resistance to a variety of other abiotic stresses, such as infused high-light (HL) and HS stresses (Nishizawa et al., 2006), oxygen deprivation stress (Banti et al., 2010), and salt as well as osmotic stress. Whereas the extent of the genes involved in stress expression, ABA metabolism, and signalling, as well as other genes that confer drought, heat, and salt resistance in Arabidopsis, was found to have changed as a result of TaHsfA6f upregulation.Using the most recent version gene IDs as standards, all the genes were detected in all two investigations, with the current analysis identifying 426 and 534 unique genes in As-

Pau3809 and c273, respectively, whereas Duan et al. [82] identified six and Xue et al. [83] identified one. (Figure S3, Table S1). In silico expression of selected transcripts were analysed under abiotic stress, tissues and developmental stages were shown in the Fig. 9.Under abiotic stress in As-Pau3809 TraesCS7B02G168300.2 (A-7a) and TraesCS7D02G270600.1 (A-4b) shows high expression at 7.21 tpm along with that TraesCS7D02G270600.1(A-4b),TraesCS6B02G237400.1(A-2c), TraesCS7D02G270600.2 (B2A), and TraesCS6A02G407700.1 (A-2a) having tpm between 6.17 to 6.94. In various developmental stages and diferernt tissues levels TraesCS6A02G407700.1(A-2a), TraesCS6B02G237400.1(A-2c) were highly expresed with tpm values 6.29 to 7.76. Considering that in c273 , TraesCS7A02G360400.1(B2B) and TraesCS7B02G267300.1(A6B) showed high tpm at 9.21 along with TraesCS7B02G168300.2(B-1) show tpm above 6 in abiotic condition. Moreover, during varios developmental stages , tissues and variety TraesCS6A02G303000.2(A6B),TraesCS7D02G527500.1(A5),TraesCS7A02G535100.1(C1b),TraesCS4A02G027700.1 (A-2d)was present with tpm 6.19 tp 7.28 Table 2. It was additionally found that hsfA2b had the greatest number of orthologs in As-Pau3809 and hsfA6b in c273 while hsfB4 ,hsfB1 showed least orthology in as3809 Overall, hsfA6b was discovered to be extremely conserved, with hsfB1 being the least conserved family in the progression of plant evolution.

Utility of database

The HSGDB database provides a comparative study of heat stress-responsive genes in two wheat species, i.e., Aegilops speltoides and Triticum aestivum. This database helps users distinguish and classify the HSP diversification in modern and wild wheat.HSGDB also develops our understanding of the functional annotation of the genes and generates transcription data for the perception of heat stress genes.

Overall, the current study adds to our contemporaneous knowledge of familiar heat response genes in wheat by identifying novel transcripts. Further research into the functional validation of these genes may open the way for the development of enormous varieties of wheat with eminent tolerance to heat and large production potential in accordance with the present global warming scenario.

Conclusion

Climate change has impacted the number and quality of important food crops, posing a severe danger to global food security. The first few days of grain filling are critical for crop net yield.Heat stress during this time cycle significantly affects overall yield. Our de novo transcriptomics-based investigation of the heat stress-tolerant wheat genotype As-Pau3809 suggests that overexpression of critical genes during the early stages of grain filling may play an important role in the heat tolerance mechanism. As-PAu3809 may contain genetic and epigenetic regulatory mechanisms that contribute to avoiding the negative effects of heat stress and hence sustaining yield and grain quality. The noteworthy findings might be functionally confirmed using reverse genetic methods in wheat or other model plants, which could help molecular breeders create appropriate strategies for developing heat-tolerant wheat genotypes. In this study, we identified the conservation of gene expression across the two species as well as the loss of functional alleles and intergenome interactions during domestication and polyploidy as major reasons behind the reduction of expression in higher ploidies in contrast to their diploid relatives. We report the seven important HSFs identified in the comparative RNA-Seq analysis of two wheat species, diploid Ae. speltoides and hexaploid T. aestivum c273, of which six were validated as potential novel candidates for thermotolerance in Ae. speltoides. Transcript profiling of HSFs under basal and heat stress conditions revealed good consistency between the expression levels of the seven genes analysed by expvip and their transcript levels detected using RNA-Seq in all the two wheat species. The purpose of HSGDB is to offer a specialised resource for genes associated with heat stress that have functional annotations. Researchers will be able to conduct comparative analyses and explore previously underappreciated additional functions of heat stress genes in wild and modern wheat varieties thanks to the interactive search features and information compiled in the database. Incorporating information on heat stress genes into additional transcriptomes is the goal for the future, with a focus on wheat species. Furthermore, there are no limitations, like registration, on the availability of the data. We think that the HSGDB offers a useful platform for scientists and breeders interested in the wheat crop. Our results encourage the exploitation of novel alleles from Ae. speltoides and other wild relatives for broadening the narrow genetic base of cultivated wheat. This knowledge can be further utilised in identification, characterization, and breeding strategies to develop heat-stress-tolerant wheat varieties.

Keypoints

- It’s a Comparative study on transcripts of wild and modern variety of wheat.
- The study involves the identification of the key genes responsible for heat stress transcrption factor.
- Database is build to visualize the results of the heat stress genes identified in *Aegilops speltoides* As-Pau3809 and *Triticum Aestivum* cv. c273

References

1.Ray DK, Ramankutty N, Mueller ND, et al. Recent Patterns of Crop Yield Growth and Stagnation.*Nature Communications* 2012*;* 3(1293):1-7.

2.Zhao C, Liu B, Piao S,et al. Temperature IncreaseReduces Global Yields of Major Crops in Four Independent Estimates.*Proceedings of the National Academy of Science* 2017*;* 114**:** 9326–9331 (2017).

3.Fischer RA, Connor DJ. Issues for Cropping and Agricultural Science in the next 20 Years.*Field Crops Research* 2018; 222 :121–142.

4.Masson DV, Zhai P, Pirani A, et al. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press: Cambridge, UK, 2021.

5.Habash DZ, Kehel Z, Nachit M. Genomic approaches for designing durum wheat ready for climate change with a focus on drought. *Journal of Experimantal Botany* 2009; 60 : 2805–2815.

6.Akter N, Islam MR. Heat stress effects and management in wheat. A review. *Agronomy for Sustainable Development* 2017; 37: 1-17.

7.Arora S, Cheema J, Poland J, et al. Genome-wide association mapping of grain micronutrients concentration in Aegilops tauschii. *Frontiers in Plant Science* 2019;10: 1-14.

8.Climate change 2007: The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change;Cambridge University Press :2007.

9.Gibson LR, Paulsen GM. Yield components of wheat grown under high temperature stress during reproductive growth. *Crop Science* 1999; 39:1841-1846.

10.Bennett D, Izanloo A, Reynolds M, et al. Genetic dissection of grain yield and physical grain quality in bread wheat (*Triticum aestium L.*) under water-limited environments. *Theoretical and Applied Genetics* 2012;125:55-271.

11. Kuchel H, Williams K, Langridge P, et al. Genetic dissection of grain yield in bread wheat. II. QTL-by-environment interaction. *Theoretical and Applied Genetics.* 2007; 115:1015-1027.

12. Gupta PK, Balyan HS, Gahlaut V. et al. Phenotyping, genetic dissection, and breeding for drought and heat tolerance in common wheat: Status and prospects. in Plant Breeding Reviews (ed. Janick, J.) 85–168 (Wiley, 2012).

13. Awlachew Z T, Singh R, Kaur S, et al. Transfer and mapping of the heat tolerance component traits of Aegilops speltoides in tetraploid wheat Triticum durum. *Molecular Breeding* 2016;36(78):1-15.

14. Kotak S, Larkindale J, Lee U, et al. Complexity of the heat stress response in plants. *Current Opinion in Plant Biology* 2007; 10:310- 316.

15. Suzuki N, Rizhsky L, Liang H, et al. Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. *Plant Physiology* 2005; 139:1313–1322.

16.Kucek LK, Veenstra LD, Amnuaycheewa P, et al. A grounded guide to gluten: How modern genotypes and processing impact wheat sensitivity. Comprehensive Reviews in Food Science and Food Safety 2015; 1414: 285–302.

17. Wan Y, Poole RL,Huttly AK, et al. Transcriptome analysis of grain development in hexaploid wheat. BMC Genomics 2008; 99(121): 1-16.

18. Li J, Wan H, Yang, W. Synthetic hexaploid wheat enhances variation and adaptive evolution of bread wheat in breeding processes. *Journal of Sytematics and Evolution* 2014; 52: 735–742 .

19. Waines J. High temperature stress in wild wheats and spring wheats. Funct. Plant Biology 1994; 21: 705–715.

20. Jafarzadeh J, Bonnett D, Jannink JL, et al. Breeding value of primary synthetic wheat genotypes for grain yield. PLoS ONE 2016; 11(9):1-24.

21.Colmer TD, Flowers TJ, Munns R. Use of wild relatives to improve salt tolerance in wheat. Journal of Experimental Botany 2006; 57: 1059–1078.

22. Farooq, S. Co-existence of salt and drought tolerance in Triticeae. *Hereditas* 2004;135: 205–210.

23.Vikas, VK, Sivasamy M, Kumar J, et al. Stem and leaf rust resistance in wild relatives of wheat with D genome (Aegilops spp.). Genetic Resources and Crop Evolution 2014; 61:861–874.

24. Huang M, Mheni N, Guedira BG, et al. Genetic analysis of heading date in winter and spring wheat. *Euphytica* 2018; 214, (128): 1-18.

25. Olivera PD, Rouse MN, Jin Y. Identification of new sources of resistance to wheat stem rust in Aegilops spp. in the tertiary genepool of wheat. *Frontiers in Plant Science* 2018; 9(1719):1-7.

26.Von Koskull-Döring, P., Scharf, K.-D. , Nover, L. The diversity of plant heat stress transcription factors. *Trends in Plant Science* 2007; 12:452–457.

27. Baniwal SK, Bharti K, Chan YK, et al. Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. Journal of Biosciences 2004; 29: 471–487.

28. Nover L, Bharti K, Scharf KD. Arabidopsis and the heat stress transcription factor world: How many heat stress transcription factors do we need?. *Cell Stress Chaperones* 2001;6(3): 177–189.

29. Fragkostefanakis S, Röth S, Schleiff E, et al. Prospects of engineering thermotolerance in crops through modulation of heat stress transcription factor and heat shock protein networks: HSFs and HSPs for improvement of crop thermotolerance.Plant Cell Environment 2015;38: 1881–1895.

30. Guo M, Lu JP, Zhai YF, et al. Genome-wide analysis, expression profile of heat shock factor gene family (CaHSFs) and characterisation of CaHsfA2 in pepper (*Capsicum annuum L.*). BMC Plant Biology 2015;15(151): 1-20.

31. Scharf KD, Berberich T, Ebersberger I, et al. The plant heat stress transcription factor (Hsf) family: Structure, function
and evolution. Biochimica et Biophysica Acta BBA Gene Regul. Mechanisms 2012; 1819: 104–119 (2012).

32. Xue GP, Sadat S, Drenth J, et al.The heat shock factor family from Triticum aestivum in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. Journal of Experimental Botanym 2014; 65: 539–557.

33.Nover L, Bharti, K, Döring, P, et al. Arabidopsis and the heat stress transcription factor world: How many heat stress transcription factors do we need? *Cell Stress Chaperones* 2001; 6(3):177-189.

34.von KDP, Scharf KD, Nover L. The diversity of plant heat stress transcription factors. *Trends Plant Sci*. 2007; 12: 452–457.

35. Xue GP, Sadat S, Drenth J, et al. The heat shock factor family from Triticum aestivum in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. Journal of Experimental Botany 2014; 65: 539–557.

36. Agarwal P, Khurana P Functional characterization of HSFs from wheat in response to heat and other abiotic stress conditions. Functional and Integrative Genomics 2019;19: 497–513.

37. Chauhan H, Khurana N, Agarwal P, et al. A seed preferential heat shock transcription factor from wheat provides abiotic stress tolerance and yield enhancement in transgenic Arabidopsis under heat stress environment. PLoS ONE 2013; 8(11): 1-13.

38. Xue GP, Drenth J, McIntyre CL. TaHsfA6f is a transcriptional activator that regulates a suite of heat stress protection genes in wheat (*Triticum aestivum* L.) including previously unknown Hsf targets. *Journal of Experimental Botany* 2015; 66 (3): 1025–1039.

39. Andrews S. FastQC: a quality control tool for high through put sequence data. 2010; https://www.bioinformatics.babra ham.ac.uk/projects/fastqc/.

40. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics 2014;*30: 2114–2120.

41 Haas BJ, Papanicolaou A, Yassour M, et al. (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature Protocol 2013; 8:1494–1512.

42.Grabherr MG, Haas BJ, Yassour M, et al. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology 2013;* 29: 644–652.

43. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012;* 9: 357–359.

44. Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 2013; 8: 1494–1512.

45.Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics 2011;* 12 (323):1-16.

46. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic Acids Residues* 2014; 42:222–230 .

47. Martin JA, Wang Z. Next-generation transcriptome assembly. *Nature Reviews Genetics* 2011;12:671–82.

48. Simão FA, Waterhouse RM, Ioannidis P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–3212.

49.Kriventseva,EV, Kuznetsov F,Tegenfeldt M, et al. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Residue 2019;**47**:807–811.*

50. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology 2015;16:157.

51 Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics 2008; 2008:1–12.

52. Ye, J, Fang L, Zheng H, et al. WEGO: A web tool for plotting GO annotations. Nucleic Acids Research 2006; 34: 293–297.

53. Thimm, O, Blasing O, Giben Y, et al. Mapman: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant Journal 2004; 37: 914–939.

54.Lohse M, Nagel A, Herter T, et al. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environment 2014;37: 1250–1258.

55. Borrill P, Gonzalez RR, Uauy C. ExpVIP: a customisable RNA-seq data analysis and visualization platform opens up gene expression analysis. Plant Physiology 2016;170(4): 2172–86.

56. Pareek A, Dhankher OP, and Foyer CH. Mitigating the impact of climate change on plant productivity and ecosystem sustainability. Journal of Experimental Botany 2020; 71: 451–456.

57. Rangan P, Furtado A, Henry R. (2020). Transcriptome profiling of wheat genotypes under heat stress during grain-filling. Journal of Cereal Science 2020;91:1-8.

58. Kosina P, Reynolds M, Dixon J, et al. Stakeholder perception of wheat production constraints, capacity building needs, and research partnerships in developing countries. Euphytica 2007; 157:475–483.

59.Mwadzingeni L, Shimelis H, Dube E, et al. (2016). Breeding wheat for drought tolerance: progress and technologies. Journal of Integrative Agriculture 2016; 15: 935–943.

60.Chu C, Wang S, Paetzold L, et al. RNA-seq analysis reveals different drought tolerance mechanisms in two broadly adapted wheat cultivars 'TAM 111′ and 'TAM 112′. Scientific Report 2021; 11(4301):1-16.

61. Olivera PD, Rouse MN, Jin Y. Identification of new sources of resistance to wheat stem rust in *Aegilops* spp. in the tertiary genepool of wheat. *Front. Plant Science 2018;* 9(1719):1-7.

62. Pradhan GP, Prasad PVV, Fritz AK., et al. High temperature tolerance in *Aegilops* species and its potential transfer to wheat. *Crop Science 2012;* 52: 292–304.

63. Salse J, Chague V, Bolot S, *et al.* New insights into the origin of the B genome of hexaploid wheat: Evolutionary relationships at the SPA genomic region with the S genome of the diploid relative *Aegilops speltoides*. *BMC Genomics* 2008; **9(**555): 1-12.

64 Ruban AS, Badaeva ED. Evolution of the S-genomes in *Triticum-Aegilops* alliance: Evidences from chromosome analysis. *Frontiers in Plant Science 2018;* 9(1756): 1-25.

65.Baidouri EM, Murat F, Veyssiere M, *et al.* Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytologist 2017;* **213:**1477–1486.

66. Paul, S, Duhan JS, Jaiswal S, *et al.* RNA-Seq Analysis of Developing Grains of Wheat to Intrigue Into the Complex Molecular Mechanism of the Heat Stress Response.*Frontiers in Plant Science 2022; 13:1-16.*

67 . Chu Z, Chen J, Sun J, *et al. (2017)De novo assembly and comparative analysis of the transcriptome of embryogenic callus formation in bread wheat (Triticum aestivum L.)BMC Plant Biology 2017;17(244):1-22.*

68. Liu W, Zhihui W, Zhang Y, et al. Transcriptome analysis of wheat grain using RNA-Seq. Frontiers of Agricultural Science and Engineering 2014; 1(3): 214–222.

69. Yadav IS. Sharma A, Kaur S, et al. Comparative temporal transcriptome profiling of wheat near isogenic line carrying Lr57 under compatible and incompatible interactions. Frontiers in Plant Science 2016; 7(1943): 1-18.

70. Appels , Eversole K, Stein N, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 2018;361(6403):1-62.

71. Miki, Y, Yoshida K, Mizuno N, et al. Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of Aegilops. DNA Residue 2019; 26: 171–182.

72. Chen ZJ, Ni Z. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. BioEssays 2006; 28: 240–252.

73Kushwaha KS, Vetukuri RR, Odilbekov F, et al. Differential gene expression analysis of wheat breeding lines reveal molecular insights in yellow rust resistance under field conditions. Agronomy 2020; 10(1888):1-16.

74. Kumar KS, Vetukuri RR, Odilbekov F,et al. Differential Gene Expression Analysis of Wheat Breeding Lines Reveal Molecular Insights in Yellow Rust Resistance under Field Conditions. Agronomy 2020;10(1888):1-17.

75.Vendramin V, Ormanbekova D, Scalabrin S, et al. Genomic tools for durum wheat breeding: de novo assemblyof Svevo transcriptome and SNP discovery in elitegermplasm.BMC Genomics 2019;20(1): 1-12.

76.Kumar RR, Goswami S, Sharma SK, et al. Harnessing next generation sequencing in climate change: RNA-Seq analysis of heat stress-responsive genes in wheat (Triticum aestivum L.). *OMICS* 2015;19: 632–647.

77 Guo J, Wu J, Ji Q, et al. Genome-wide analysis of heat shock transcription factor families in rice and Arabidopsis. Journal of Genetics and Genomics 2008; 35: 105–118.

78. Lin YX, Jiang HY, Chu ZX, et al. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. BMC Genomics 2011; 12 (76):1-14.

79.Li PS, Yu TF, He GH,et al. . Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. BMC Genomics 2014; 15(1009):1-16.

80. Chidambaranathan P, Jagannadham PTK, Satheesh V, et al. Genome-wide analysis identifies chickpea (Cicer arietinum) heat stress transcription factors (Hsfs) responsive to heat stress at the pod development stage. Journal of Plant Research 2018; 131:525–542.

81 Chen YB, Wang Y, Zhang HL. Identification and expression analysis of heat shock factor (Hsf) gene family in tomato (*Solanum lycopersicum*). Journal of Agricultural Biotechnology 2015; 23: 492–501.

82 Duan S, Liu B, Zhang Y, et al. Genome-wide identification and abiotic stress-responsive patternof heat shock transcription factor family in Triticum aestivum L. BMC Genomics 2019; 20: 257.

83 Xue G, Sadat S. Drenth J, et al.The heat shock factor family from Triticum aestivum in responseto heat and other major abiotic stresses and their role in regulation of heat shock protein genes. Journal of Experimental Botany 2014; 65: 539–557.

84 Eckardt NA. Everything in its place: Conservation of gene order among distantly related plant species.Plant Cell 2001; 13: 723–725.