

A COMPARATIVE ANALYSIS OF DE NOVO TRANSCRIPTOME ASSEMBLY TO UNDERSTAND THE ABIOTIC STRESS ADAPTATION OF DESERT PLANTS IN SAUDI ARABIA

BAESHEN, M. N.^{1*} – AHMED, F.^{2*} – MOUSSA, T. A. A.^{1,3} – ABULFARAJ, A. A.⁴ – JALAL, R. S.¹ – NOOR, S. O.⁵ – BAESHEN, N. A.⁵ – HUELSENBECK, J. P.^{1,6}

¹*Department of Biology, College of Science, University of Jeddah, Jeddah, Saudi Arabia*

²*Department of Biochemistry, College of Science, University of Jeddah, Jeddah, Saudi Arabia*

³*Botany and Microbiology Department, Faculty of Science, Cairo University, Giza 12613, Egypt*

⁴*Department of Biological Sciences, College of Sciences and Arts - Rabigh Campus, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

⁵*Department of Biology, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia*

⁶*Department of Integrative Biology, University of California, Berkeley, USA*

**Corresponding authors*

e-mail: mnbaeshen@uj.edu.sa, fahmed1@uj.edu.sa

(Received 24th Nov 2020; accepted 20th Mar 2021)

Abstract. *Rhazya stricta*, *Senna italica*, and *Zygophyllum simplex* are important desert plants of Saudi Arabia with great economic and medicinal value. However, their tolerance and survival mechanisms under combined abiotic stresses such as high temperature, high salinity, and drought are not well understood. In order to investigate the potential molecule mechanism of abiotic stress tolerance in these plants, we used de novo transcriptome assembly and their comparative analysis. This study used leaf tissues to construct three cDNA libraries of these plants and then generated RNA-seq data by the Illumina HiSeq2000 platform. Sequencing reads were de novo assembled to generate: (a) 71,116 unigenes in *R. stricta*; (b) 59,274 unigenes in *S. italica*; (c) 70,300 unigenes in *Z. simplex*. Furthermore, the unigenes from these plants were annotated and analyzed with different databases. A comparative analysis of KEGG pathways identified several common pathways induced in these plants, including “Plant-pathogen interaction”, “Plant hormone signal transduction”, “Spliceosome”, “RNA transport”, and “Protein processing in endoplasmic reticulum”, which may play an important role in combined abiotic drought, heat, and salinity stress. Finally, a comparative analysis of transcriptional regulators identified C2H2, C3H, CCHC(Zn), MYB-HB-like, PHD, WD40-like, and bHLH as common Transcription Factors responsible for abiotic stress tolerance in these plants. Our study revealed key factors involved in abiotic stress tolerance, which could be applied to develop high-yield transgenic crops capable of growing under combined abiotic stresses in the field.

Keywords: plant transcriptome, GO, KEGG, simple sequence repeats, heat stress, drought stress, salinity stress, transcription factors

Introduction

Abiotic (non-biological) stresses, including drought, hot climate, salinity, and nutrient are the leading limiting factors in agricultural productivity in the desert (Fahad et al., 2017). These factors negatively impact plant growth and development, cause a significant decrease in biomass, crop yield, and quality resulting in extensive losses of agriculture production. The combined stress of drought and heat are the most important abiotic stress widely encountered by crops in the field of Saudi Arabia. The Kingdom of

Saudi Arabia covers a land area of approximately 2,150,000 square kilometers (<http://www.mofa.gov.sa/>), with an estimated population of 32.55 million by 2017 (<https://www.stats.gov.sa/>). However, due to continuous drought, heat, and salinity stresses in the field, a big land of Saudi Arabia is deserted and not suitable for agriculture farming. Therefore, Saudi Arabia is unable to use its vast land for agriculture to meet domestic food demands and depends on importing them. For example: In 2013, 660,145 tons of wheat has been locally produced under 102,613 hectares cultivation, while 2,117,052 tons of wheat were imported (Fiaz et al., 2016). In the same year, 11,267 tons of barley has been locally produced in 1502 hectares of land, while 10,446,332 tons of barley were imported (Fiaz et al., 2016).

Researchers showed that a variety of plants have evolved their strategy to respond to different abiotic stresses, including drought and heat encounter in their extreme natural environment. These plants adapted effectively by activating the expression of various stress-responsive genes producing different metabolites, inducing signaling and biochemical pathways to mitigate stress-induced damages (An et al., 2013). These stress induced gene expression consequences on the: (a) morphological modifications in plants, including stomatal conductance, size of leaf and roots; (b) physiological and molecular modifications, such as the production of antioxidant compounds and cellular osmotic adjustment to develop abiotic stress tolerance. Studies found that each stress's molecular response, such as drought or heat, is unique (Rizhsky et al., 2002). Furthermore, combinations of various stresses show different molecular responses and cannot be extrapolated from their individual response (Rizhsky et al., 2002). This is the reason why transgenic plants developed for drought stress under experimental conditions failed tolerance when tested in the field under a naturally occurring environmental condition such as a combination of drought and heat stresses. During drought stress, stomata are close to avoiding water loss, while during heat stress, the stomata are open to reducing leaf temperature. In this way, drought and heat stresses have opposing physiological changes (Rizhsky et al., 2004). However, during combined drought and heat stresses, stomata are close (Rizhsky et al., 2002; Prasch and Sonnewald, 2013). In response to drought stress, proline synthesis and its degradation pathways are activated and repressed, respectively (Krasensky and Jonak, 2012). This results accumulation of proline in the leaf to protect plants against osmotic stress due to drought (An et al., 2013). Heat stress results in the accumulation of unfolded protein in the cytoplasm, which induces the expression of heat shock proteins (HSPs). HSPs bind and stabilize misfolded proteins and inhibit protein aggregation (Al-Whaibi, 2011). Thus, HSPs act as molecular chaperones. Interestingly, HSPs expression is controlled by various heat shock factors (HSFs). Overexpression of HSF2 in transgenic *Arabidopsis thaliana* shows greater tolerance of combined heat and high light stress than control (Nishizawa et al., 2006). Expression of Ethylene Response Factor1 (ERF1) (AT3G23240) in *A. thaliana* is greatly induced by salt and drought stress, and its overexpression enhanced the drought, heat, and salt tolerance in *A. thaliana* (Cheng et al., 2013). In response to various abiotic stress, ERF1 up-regulates a different set of stress-related genes by binding to GCC box and DRE/CRT elements in their promoter region (Cheng et al., 2013). Transcription of the wheat ERF1 (TaERF1) gene was also induced by different biotic and abiotic stresses (Xu et al., 2007). The study found that the BT2 acts as a central stress regulator in *A. thaliana* and mediate different biotic and abiotic stress the responses. Furthermore, diverse microbial communities were

identified in the soils, which help abiotic stress tolerance in plants through association with plants' roots (Baeshen et al., 2020).

Interestingly, several wild plants remarkably adapted to grow in the desert area under combined stress of heat, drought and high salt. However, the molecular mechanism, and pathways underlying stress response and tolerance has not been well understood in these plants. The advancement in next generation sequencing (NGS) technology gives ample opportunities to analyze the transcriptomic and genomics data to understand the abiotic stress tolerance in non-model plants without a reference genome (Grabherr et al., 2011; Li et al., 2017; Park et al., 2014; Sabir et al., 2016). For instance: de novo genome assembly of *Thellungiella parvula* that adapted the saline and resource poor environment (Dassanayake et al., 2011); Transcriptome analysis of *R. stricta* that grows in arid zone (Yates et al., 2014); transcriptomes analysis of desiccation-tolerant *Craterostigma plantagineum* (Rodriguez et al., 2010); and abiotic stress tolerance *Rhizophora mangle* and *Heritiera littoralis* (Dassanayake et al., 2009).

In order to investigate the molecule mechanism to adapt combined abiotic stress tolerance, this study used de novo transcriptome sequencing and assembly of three different plants *Rhazya stricta*, *Senna italica*, and *Zygophyllum simplex*. These plants grow in their natural environment, having combined abiotic stresses of high temperature, drought, and high salt conditions in Saudi Arabia. The assembled unigenes were annotated with NR, SwissProt, NT GO, KEGG, and GOC databases to elucidate the expressed genes and pathways involved in the abiotic stress tolerance. Finally, the comparative analysis of unigenes among three plants identified the common genes and TFs involved in abiotic stress tolerance. The finding of this study would be applied to develop high-yield transgenic crops with improved abiotic stress resistance. The present study would also provide transcriptome resources for further study of the molecular mechanism of abiotic stress tolerance in plants.

Materials and methods

Plant sample collection

The leaf shows significant physiological and biochemical changes than root in drought-tolerant wheat, which supports that the leaf is a more abiotic stress-sensitive part of the plant (Kang et al., 2019). Therefore, in our study, we consider leaves from three different plants. The plants' leaves were collected during the morning as the temperature was 38 °C from a public site at Hadda, Mecca-Jeddah Road, Saudi Arabia (N21° 45' 04.03", E39° 53' 88.92") (Table 1). The leaf samples of three plants *R. stricta*, *S. italica*, and *Z. simplex* were taken in plastic bags and stored at -80 °C for RNA extraction.

Table 1. Location of plant samples with their GPS coordinates

Plants	Family	Photosynthesis	Sample codes	GPS coordinates	
				N	E
<i>Rhazya stricta</i>	Apocynaceae	C3	1A	21° 45' 04.03"	39° 53' 88.92"
<i>Senna italica</i>	Fabaceae	C4	4A	21° 45' 04.03"	39° 53' 88.92"
<i>Zygophyllum simplex</i>	Zygophyllaceae	C4	5A	21° 45' 04.03"	39° 53' 88.92"

RNA extraction, cDNA library construction and Illumina sequencing

The total RNA was extracted from 100 mg of leaf tissue by using TRIzol reagent (Life Technologies, Grand Island, NY, USA) according to manufacturer's instructions. Briefly, mRNAs were isolated using magnetic beads with Oligo (dT) then sliced into short fragments by mixing them in fragmentation buffer. The mRNA fragments were used as templates to synthesized cDNA using random hexamer primers. Short fragments were purified and resolved with EB buffer for end reparation and single nucleotide A (adenine) addition, followed by connecting them with adapters. The suitable fragments size was selected for the PCR amplification as templates. Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System were used to quantify and qualify the sample library. Finally, three libraries were prepared from three plants *R. stricta*, *S. italica*, and *Z. simplex*, each library coming from a single plant. Subsequently, these libraries were sequenced using Illumina HiSeq™ 2000 platform. The sequencing machine gives the raw sequence image data, which was transformed by base calling into sequence data and stored in the FASTQ format. The paired-end raw reads contain adapters and low-quality bases (Q-value ≤ 20) were removed to get a clean read (Mbandi et al., 2014).

De novo transcriptome assembly

The high-quality reads were used for de novo transcriptome assembly using Trinity sequence assembler, which consists of three independent software modules: Inchworm, Chrysalis, and Butterfly (Grabherr et al., 2011). Clean sequence reads were assembled into larger contigs by Inchworm; Contigs were clustered to create a de Bruijn graph for each locus (component) by Chrysalis; the Butterfly then processes each de Bruijn graph and reports unique transcripts (unigenes).

Gene annotation

In order to identify the potential function of unigenes, sequence alignment was performed against a series of protein databases: NR, SwissProt, KEGG (Kyoto Encyclopedia of Genes and Genomes), and COG (Clusters of Orthologous Groups of proteins) with BLASTx (E-value $< 10^{-5}$). In addition, sequence alignment of unigenes was performed against nucleotide databases NT with BLASTn (E-value $< 10^{-5}$). With NR annotation, the Blast2GO program (v 2.5.0) was used to get GO (Gene Ontology) annotation of unigenes (Conesa et al., 2005). Finally, GO functional annotation were classified and plotted using WEGO software (Ye et al., 2006). The Venn diagrams were plotted using (<http://bioinformatics.psb.ugent.be/webtools/Venn/>)

SSR and SNP analysis

SSR (simple sequence repeat) detection is carried out with software MicroSATellite (MISA) using unigenes as reference (Beier et al., 2017), while SNP (single nucleotide polymorphism) analysis was carried out with SOAPSnp (Li et al., 2009).

Protein-coding region prediction (CDS)

Unigenes was firstly aligned by the BLASTx (E-value $< 10^{-5}$) to protein databases in the priority order of NR, Swiss-Prot, KEGG, and COG. Unigenes aligned to a higher priority database will not be aligned to a lower priority database. The alignments end

when all alignments are finished. Proteins with the highest ranks in blast results are taken to decide the coding region sequences of Unigenes. Then the coding region sequences are translated into amino sequences with the standard codon table. So, both the nucleotide sequences (5'→3') and amino sequences of the Unigene coding region are acquired. Unigenes that could not align with any of the above-mentioned databases are scanned by ESTScan, producing nucleotide sequence (5'→3') direction and then translated into the amino acid sequence of the predicted coding region.

Unigene expression analysis

To identify genes associated with abiotic stresses, the clean reads were mapped to Unigenes using Bowtie2, and then the gene expression levels were measured with the number of Fragments Per Kilobase of a given transcript per million mapped reads (FPKM) according to the given formula (Eq. 1). FPKM normalized the reads counts based upon library sizes and the length of the transcripts.

$$FPKM = \frac{10^6 C}{NL/10^3} \quad (\text{Eq.1})$$

where C is the number of reads uniquely aligned to a unigene; N is the total number of reads that are uniquely aligned to all unigenes; L is the base number in the CDS of one unigene.

Unigene transcription factor and transcriptional regulator prediction

In order to predict the transcription factor and transcriptional regulator, Unigenes of *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A) were analyzed with online tool PlantTFcat (<http://plantgrn.noble.org/PlantTFcat/>) (Dai et al., 2013). Then the output results were analyzed and the most abundant TF were selected using an in-house program.

Results and discussions

RNA sequencing

In order to understand the comprehensive transcriptome of *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A), RNAs were extracted from leaf samples, and three cDNA libraries were constructed. Paired-end sequencing of these libraries was performed on an Illumina HiSeq 2000 platform, which generated a total of 56,965,920; 54,266,528; and 54,062,860 raw reads from *R. stricta* (1A); *S. italica* (4A); and *Z. simplex* (5A), respectively. After pre-processing of reads (trimming of the adaptor and filtering for the low-quality region), 54,201,896; 51,152,532; and 51,449,822 high-quality reads were obtained from *R. stricta* (1A); *S. italica* (4A); and *Z. simplex* (5A), respectively (Table 2).

De novo transcriptome assembly

In absence of whole genome sequencing data, de novo transcriptome assembly is employed to study the gene expression profiling, to identify novel transcripts and genetic markers, and to understand the signaling pathways in different physiological conditions (Deng et al., 2019). The clean reads from each library were used for de novo assembly

with Trinity programs that generated: (a) 98,278 contigs and 71,116 unigenes for *R. stricta* (1A); 107,175 contigs and 59,274 unigenes for *S. italica* (4A), and 116,444 contigs and 70,300 unigenes for *Z. simplex* (5A). The assembly statistics, the length distribution of contigs and unigenes of each plant are provided in *Table 3* and *Figure 1*. The sequence file of contigs and unigenes are provided into the supplementary dataset [*Tables S1* and *S2* for *R. stricta* (1A); *Tables S3* and *S4* for *S. italica* (4A); and *Tables S5* and *S6* for *Z. simplex* (5A)]. A previous study used the leaves transcriptome of *R. stricta* and, using de novo assembly generated only 28018 unique transcript sequences (mean length = 1643 bp; N50 = 2199 bp) (Yates et al., 2014).

Table 2. Statistical description of sequencing data from *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A)

Samples	<i>R. stricta</i> (1A)	<i>S. italica</i> (4A)	<i>Z. simplex</i> (5A)
Total raw reads	56,965,920	54,266,528	54,062,860
Total clean reads	54,201,896	51,152,532	51,449,822
Total clean nucleotides (nt)	4,878,170,640	4,603,727,880	4,630,483,980
Q20 percentage	97.36%	97.42%	98.01%
N percentage	0.12%	0.09%	0.00%
GC percentage	43.94%	44.67%	43.01%

Q20 percentage is proportion of nucleotides with quality value larger than 20; N percentage is proportion of unknown nucleotides in clean reads; GC percentage is proportion of guanidine and cytosine nucleotides among total nucleotides

Table 3. Statistical of assembly quality of high-quality reads from *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A)

Samples	<i>R. stricta</i> (1A)		<i>S. italica</i> (4A)		<i>Z. simplex</i> (5A)	
	Contig	Unigene	Contig	Unigene	Contig	Unigene
Total number	98,278	71,116	107,175	59,274	116,444	70,300
Total length (nt)	47,232,016	101,448,447	45,217,296	60,317,437	49,262,775	81,584,604
Mean length (nt)	481	1427	422	1018	423	1161
N50	1175	2276	903	1671	945	1830
Total consensus sequences	-	71,116	-	59,274	-	70,300
Distinct clusters	-	34,830	-	23,405	-	33,142
Distinct singletons	-	36,286	-	35,869	-	37,158

Total consensus sequences represents all assembled unigenes; distinct clusters represents the cluster unigenes. The same cluster contains some high similar (more than 70%) unigenes, and these unigenes may come from same gene or homologous gene; distinct singletons represents this unigene come from a single gene

Function annotation and classification of unigenes

Functional annotation and classification of all expressed unigenes were carried out to understand their protein functions (Kerr et al., 2019). Initially, unigenes were searched against protein databases including NR, SwissProt, GO, KEGG, and COG by BLASTx, and then retrieved proteins with the highest sequence similarity (E-value < 10⁻⁵) with unigenes. In addition, the expressed unigenes were searched against the NT nucleotide database by BLASTn and retrieved the highest sequence similarity (E-value < 10⁻⁵). The number of unigenes annotated with each database is provided in *Table 4*.

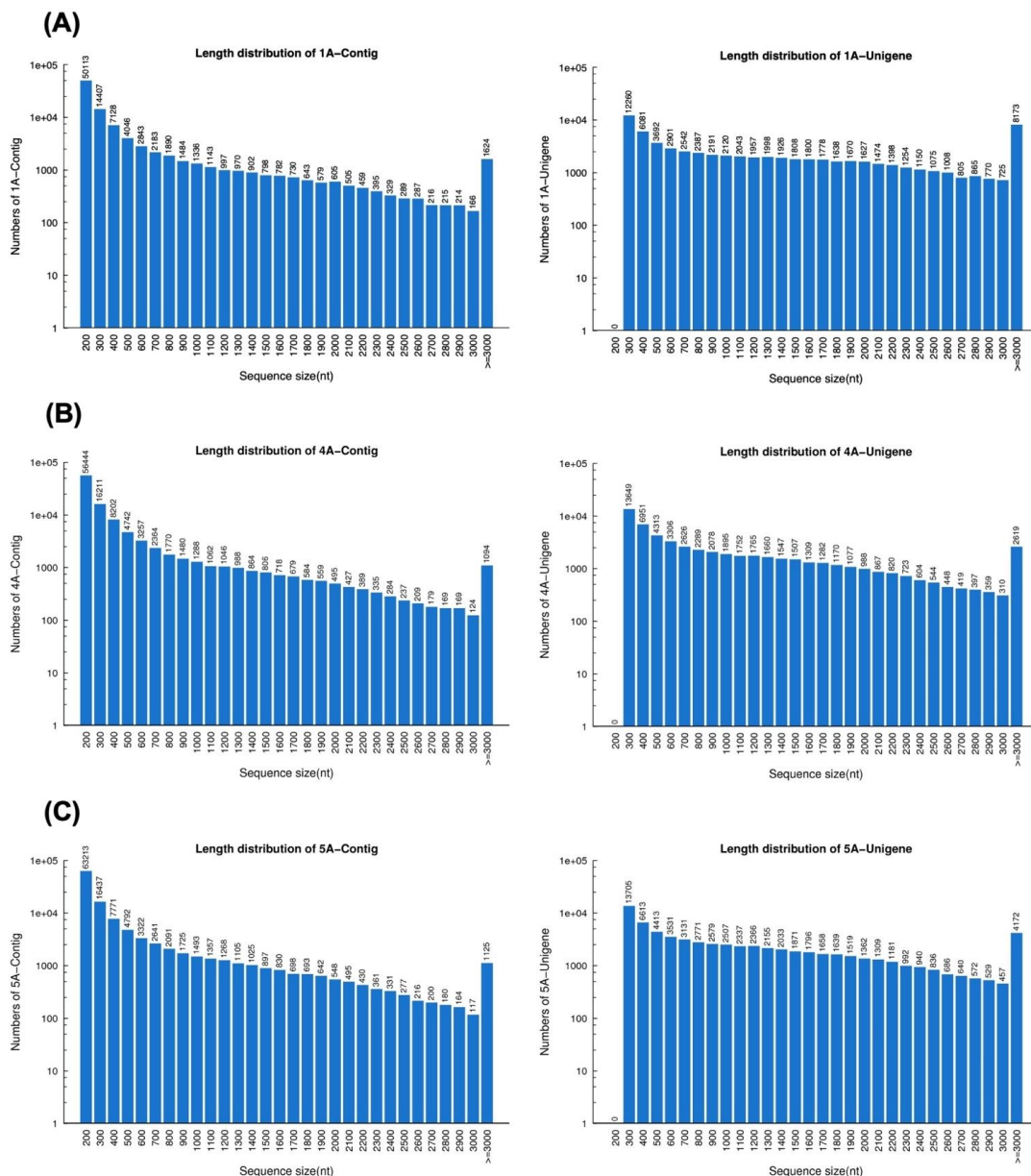


Figure 1. Distribution of contig and unigene length of Trinity transcriptome assembly. (A) *R. stricta* (1A); (B) *S. italica* (4A); and (C) *Z. simplex* (5A)

Table 4. Summary of unigenes annotation of *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A)

Samples	<i>R. stricta</i> (1A)	<i>S. italica</i> (4A)	<i>Z. simplex</i> (5A)
NR	47,940	40,957	51,456
NT	43,789	40,313	45,272
Swiss-Prot	32,126	27,431	35,073
GO	33,559	29,698	36,911
KEGG	29,503	23,739	30,619
COG	20,847	15,261	20,931
All	49,208	43,007	52,588

NR annotation (similarity analysis)

The unigenes annotation of NR database of *R. stricta* (1A) showed: (a) The E-value distribution analysis found that 63.44% of sequences showed strong homology (E-value < 1e-45), while 36.56% of sequences showed homology between 1e-5 and 1e-45 (Fig. 2A); (b) The similarity distribution analysis showed that 69.54% of the unigenes have similarity higher than 60%, while 23.21% unigenes have similarity between 40% to 60% (Fig. 2B); (c) The species distribution revealed that 21.66% of unigenes were matched with sequences of *Solanum tuberosum* followed by *Vitis vinifera* (19.46%), *Solanum lycopersicum* (11.97%), and *Erythranthe guttata* (9.25%) (Fig. 2C).

The unigenes annotation of NR database of *S. italica* (4A) showed: (a) The E-value distribution analysis found that 63.13% of sequences showed strong homology (E-value < 1e-45), while 36.87% of sequences showed homology between 1e-5 and 1e-45 (Fig. 3A); (b) The similarity distribution analysis showed that 81.82% of the unigenes have similarity higher than 60%, while 13.63% unigenes have similarity between 40% to 60% (Fig. 3B); (c) The species distribution revealed that 39.83% of unigenes were matched with sequences of *Glycine max* followed by *Phaseolus vulgaris* (13.44%), and *Cicer arietinum* (12.93%) (Fig. 3C).

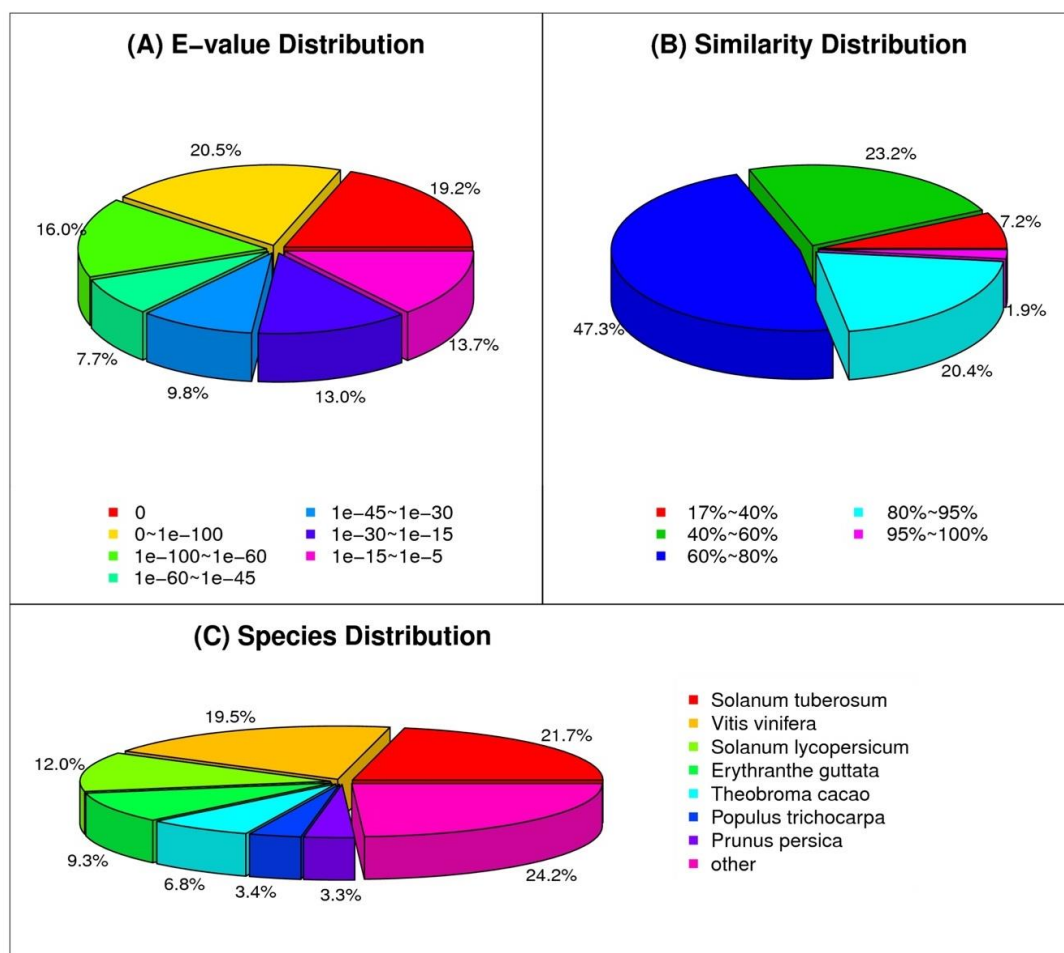


Figure 2. Annotation results of homology search of *R. stricta* (1A) unigenes against NR database. (A) The E-value distribution of the result of NR annotation. (B) The similarity distribution of the result of NR annotation. (C) The species distribution of the result of NR annotation

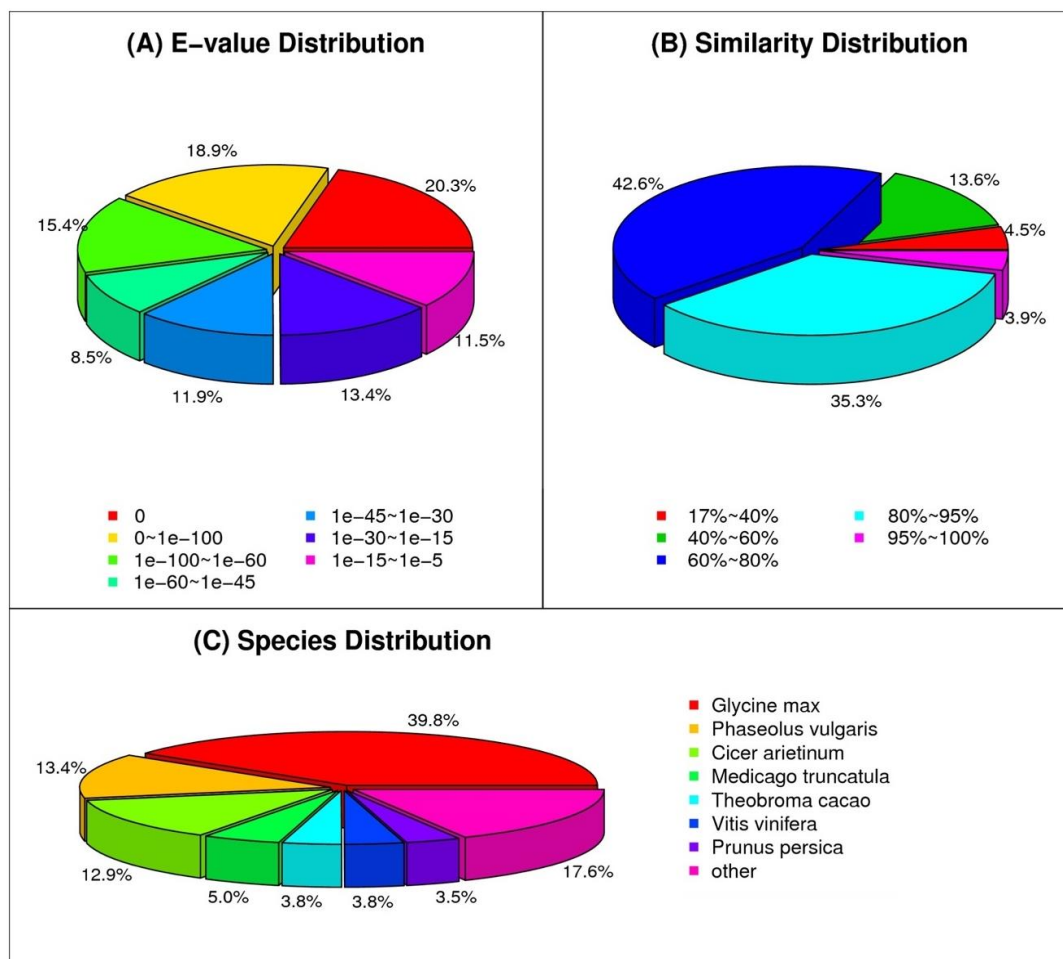


Figure 3. Annotation results of homology search of *S. italica* (4A) unigenes against NR database. (A) The E-value distribution of the result of NR annotation. (B) The similarity distribution of the result of NR annotation. (C) The species distribution of the result of NR annotation

The unigenes annotation of NR database of *Z. simplex* (5A) showed: (a) The E-value distribution analysis found that 62.11% of sequences showed strong homology (E-value < 1e-45), while 37.89% of sequences showed homology between 1e-5 and 1e-45 (Fig. 4A); (b) The similarity distribution analysis showed that 66.67% of the unigenes have similarity higher than 60%, while 25.85% unigenes have similarity between 40% to 60% (Fig. 4B); (c) The species distribution revealed that 17.05% of unigenes were matched with sequences of *Theobroma cacao* followed by *Vitis vinifera* (13.97%), and *Prunus persica* (8.53%) (Fig. 4C).

GO annotation

With NR annotation, the Blast2GO program was used to get GO annotation, including molecular function (MF), biological process (BP), and cellular component (CC) of unigenes of each plant. After getting GO annotation, WEGO software was used for functional classification for all unigenes to understand gene functions distribution.

(a) In *R. stricta* (1A), there were 33,559 unigenes assigned with at least one GO term in BP, CC, or MF (Fig. 5A). The BP was further classified as 22 functional groups, among them “metabolic process” is the highest number of unigenes (21,297) followed by “cellular process” (18,918 unigenes), and “single-organism process” (16,448 unigenes), “response to stimulus” (7,487 unigenes) and “biological regulation” (5,719 unigenes). The CC were further classified as 17 functional groups, among them “cell”, and “cell part” are the highest number of unigenes (18,711) followed by “organelle” (14,733 unigenes), and “membrane” (9,081 unigenes), “organelle part” (6,036 unigenes) and “membrane part” (4,431 unigenes). The MF was further classified as 17 functional groups, among them “catalytic activity” is the highest number of unigenes (18,293) followed by “binding” (14,695 unigenes), and “transporter activity” (2,715 unigenes), “structural molecule activity” (504 unigenes) and “molecular transducer activity” (476 unigenes).

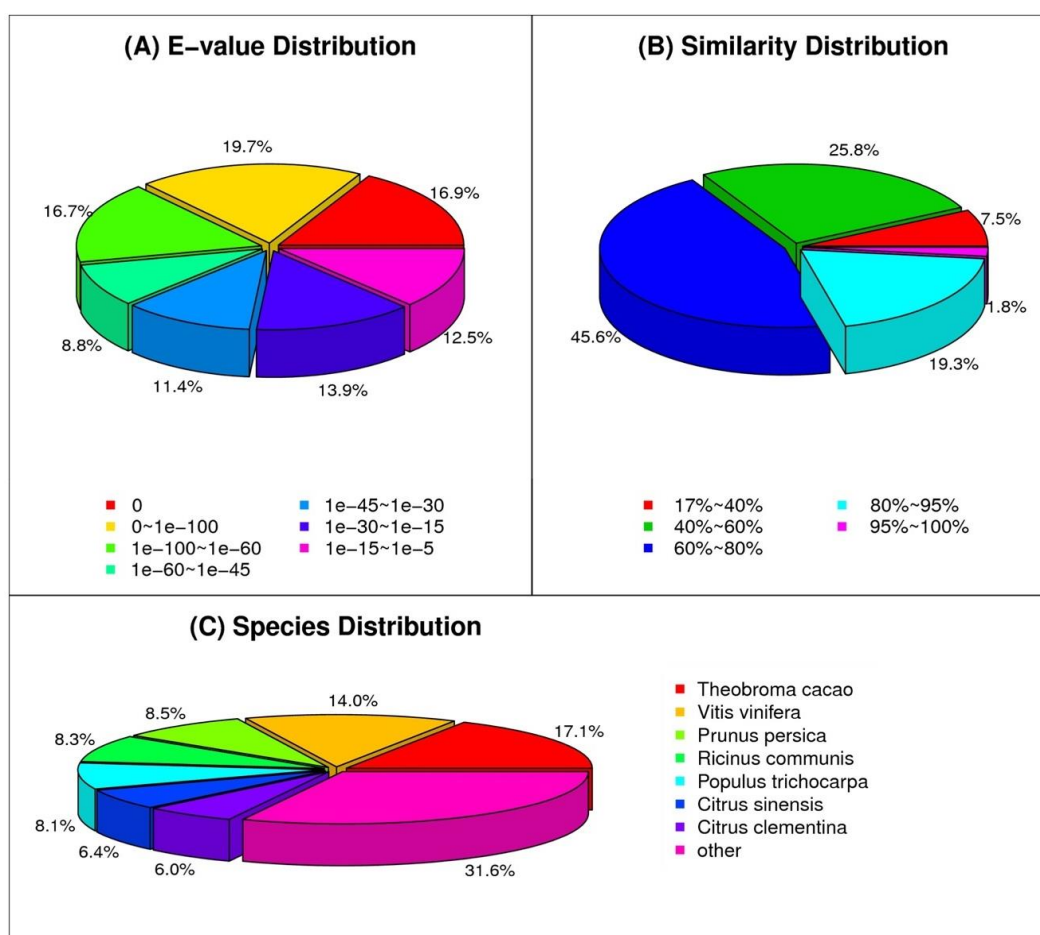


Figure 4. Annotation results of homology search of *Z. simplex* (5A) unigenes against NR database. (A) The E-value distribution of the result of NR annotation. (B) The similarity distribution of the result of NR annotation. (C) The species distribution of the result of NR annotation

(b) In *S. italica* (4A), there were 29,698 unigenes assigned with at least one GO term in BP, CC, or MF (Fig. 5B). The BP was further classified as 23 functional groups, among them “metabolic process” is the highest number of unigenes (19,400) followed by “cellular process” (17,471 unigenes), and “single-organism process” (14,105

unigenes), “response to stimulus” (7,242 unigenes) and “biological regulation” (5,820 unigenes). The CC was further classified as 17 functional groups, among them “cell”, and “cell part” are the highest number of unigenes (16,559) followed by organelle (12,572 unigenes), and “membrane” (7,933 unigenes), “organelle part” (5,237 unigenes) and “membrane part” (3,812 unigenes). The MF was further classified as 17 functional groups, among them “catalytic activity” is the highest number of unigenes (15,515 unigenes) followed by “binding” (14,781 unigenes), and “transporter activity” (2,004 unigenes), “structural molecule activity” (580 unigenes) and “nucleic acid binding transcription factor activity” (565 unigenes).

(c) In *Z. simplex* (5A), there were 36,911 unigenes assigned with at least one GO term in BP, CC, or MF (Fig. 5C). The BP was further classified as 23 functional groups, among them “metabolic process” is the highest number of unigenes (24,134) followed by “cellular process” (22,005 unigenes), “single-organism process” (18,983 unigenes), “response to stimulus” (9,433 unigenes) and “biological regulation” (7,841 unigenes). The CC was further classified as 17 functional groups, among them “cell”, and “cell part” are the highest number of unigenes (22,496) followed by “organelle” (17,197 unigenes), “membrane” (10,107 unigenes), “organelle part” (6,766 unigenes) and “membrane part” (4,527 unigenes). The MF was further classified as 17 functional groups, among them “catalytic activity” is the highest number of unigenes (19,140) followed by “binding” (17,208 unigenes), “transporter activity” (2,553 unigenes), “structural molecule activity” (758 unigenes) and “nucleic acid binding transcription factor activity” (642 unigenes). The analysis of GO annotation indicated that a large number of unigenes were associated with overlapping biological functions among *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A).

In response to the abiotic stress, the plant adjusts the metabolic process to repair the cell damage and restore cell homeostasis. We observed that a majority of unigenes from *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A) are annotated as “metabolic process”. Previous studies showed a strong association of metabolic process in abiotic stress resistance in plants; thus, our study supports the previous findings (Pan et al., 2016; Puranik et al., 2011).

COG annotation

The unigenes were further annotated against COG database for functional prediction and classification. COG is used to annotate the newly sequenced genome into 25 functional categories.

(a) *R. stricta* (1A): 20,847 unigenes were categorized into 25 groups with the largest category was “General function prediction only” (6,669 unigenes) followed by “Replication, recombination and repair” (3,368 unigenes), “Transcription” (3,099 unigenes), “Posttranslational modification, protein turnover, chaperones” (2,523 unigenes), “Signal transduction mechanisms” (2,449 unigenes). However, only 9 and 2 unigenes are assigned to “Extracellular structures” and “Nuclear structure”, respectively (Fig. 6A).

(b) *S. italica* (4A): 15,261 unigenes were categorized into 25 groups with the largest category was “General function prediction only” (5,318 unigenes) followed by “Transcription” (2,612 unigenes), “Replication, recombination and repair” (2,525 unigenes), “Signal transduction mechanisms” (2,311 unigenes), “Posttranslational modification, protein turnover, chaperones” (1,951 unigenes). However, only 6 and 2 unigenes are assigned to “Extracellular structures” and “Nuclear structure”, respectively (Fig. 6B).

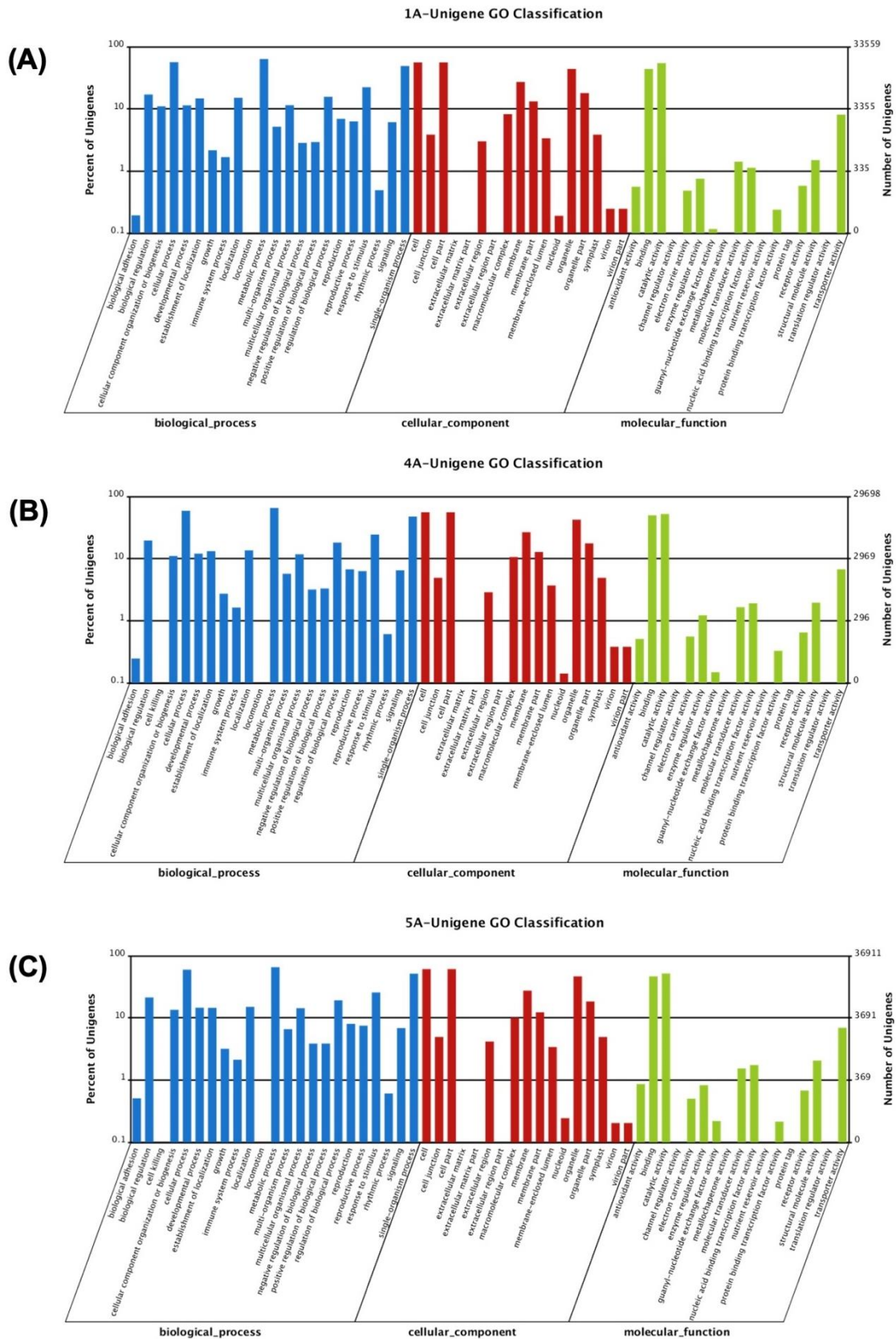


Figure 5. Histogram of GO annotation of assembled unigenes. (A) *R. stricta* (1A); (B) *S. italica* (4A); and (C) *Z. simplex* (5A). GO categories are: biological process (BP), cellular component (CC), and molecular function (MF)

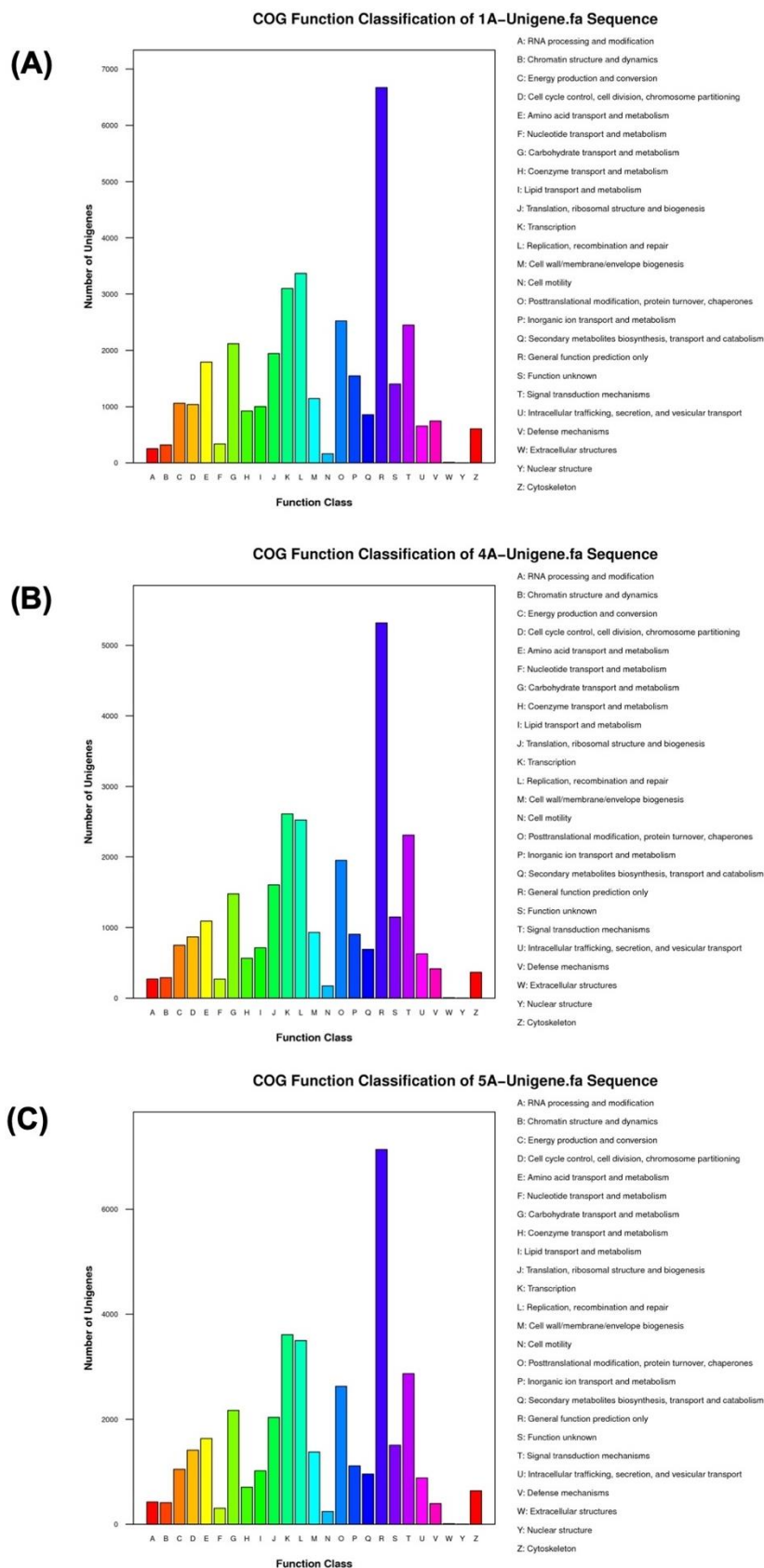


Figure 6. Histogram of COG annotation of assembled unigenes. (A) *R. stricta* (1A); (B) *S. italica* (4A); and (C) *Z. simplex* (5A)

(c) *Z. simplex* (5A): 20,931 unigenes were categorized into 25 groups with the largest category was “General function prediction only” (7,136 unigenes) followed by “Transcription” (3,611 unigenes), “Replication, recombination and repair” (3,498 unigenes), “Signal transduction mechanisms” (2,870 unigenes), “Posttranslational modification, protein turnover, chaperones” (2629 unigenes). However, only 10 and 3 unigenes are assigned to “Extracellular structures” and “Nuclear structure”, respectively (Fig. 6C).

In our study, a large number of unigenes from *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A) were assigned to “Transcription”, “Replication, recombination and repair”, “Signal transduction mechanisms”, and “Posttranslational modification, protein turnover, chaperones”. Activation of these processes is essential in modulating the gene expression and signal transduction, to repair the damage of DNA and proteins, and modification of proteins to preventing the cellular damage due to abiotic stresses and thus support previous findings (Arisha et al., 2020; Wang et al., 2004; Sarwar et al., 2019).

Functional classification by KEGG

In order to understand the biological functions, unigenes were searched against the KEGG database.

(a) *R. stricta* (1A): We found 29503 unigenes were associated with 128 KEGG pathways (Table S7). Among them, top 10 enriched pathways were “Metabolic pathways” (ko01100) with 6077 (20.6%) unigenes, “Biosynthesis of secondary metabolites” (ko01110) with 2970 (10.07%) unigenes, “Plant-pathogen interaction” (ko04626) with 1316 (4.46%) unigenes, “Plant hormone signal transduction” (ko04075) with 1112 (3.77%) unigenes, Spliceosome (ko03040) with 1095 (3.71%) unigenes, “RNA transport” (ko03013) with 849 (2.88%) unigenes, “Protein processing in endoplasmic reticulum” (ko04141) with 780 (2.64%) unigenes, “RNA degradation” (ko03018) with 595 (2.02%) unigenes, “Purine metabolism” (ko00230) with 560 (1.9%) unigenes, “Starch and sucrose metabolism” (ko00500) with 551 (1.87%) unigenes (Fig. 7).

(b) *S. italica* (4A): We found 23,739 unigenes were associated with 128 KEGG pathways (Table S7). Among them, top 10 enriched pathways were “Metabolic pathways” (ko01100) with 4,976 (20.96%) unigenes, “Biosynthesis of secondary metabolites” (ko01110) with 2,399 (10.11%) unigenes, “Plant hormone signal transduction” (ko04075) with 1,305 (5.5%) unigenes, “Plant-pathogen interaction” (ko04626) with 1,296 (5.46%) unigenes, Spliceosome (ko03040) with 972 (4.09%) unigenes, “RNA transport” (ko03013) with 872 (3.67%) unigenes, “mRNA surveillance pathways” (ko03015) with 618 (2.6%) unigenes, “Protein processing in endoplasmic reticulum” (ko04141) with 608 (2.56%) unigenes, “Purine metabolism” (ko00230) with 559 (2.35%) unigenes, “Starch and sucrose metabolism” (ko00500) with 541 (2.28%) unigenes (Fig. 8).

(c) *Z. simplex* (5A): We found 30,619 unigenes were associated with 128 KEGG pathways (Table S7). Among them, top 10 enriched pathways were “Metabolic pathways” (ko01100) with 6,657 (21.74%) unigenes, “Biosynthesis of secondary metabolites” (ko01110) with 2,940 (9.6%) unigenes, “Plant hormone signal transduction” (ko04075) with 1,619 (5.29%) unigenes, “Plant-pathogen interaction” (ko04626) with 1,342 (4.38%) unigenes, “Spliceosome” (ko03040) with 1079 (3.52%) unigenes, “RNA transport” (ko03013) with 1,003 (3.28%) unigenes, “Glycerophospholipid metabolism” (ko00564) with 831 (2.71%) unigenes, “Endocytosis” (ko04144) with 823 (2.69%) unigenes, “Protein processing in endoplasmic reticulum” (ko04141) with 799 (2.61%) unigenes, “Starch and sucrose metabolism” (ko00500) with 747 (2.44%) unigenes (Fig. 9).

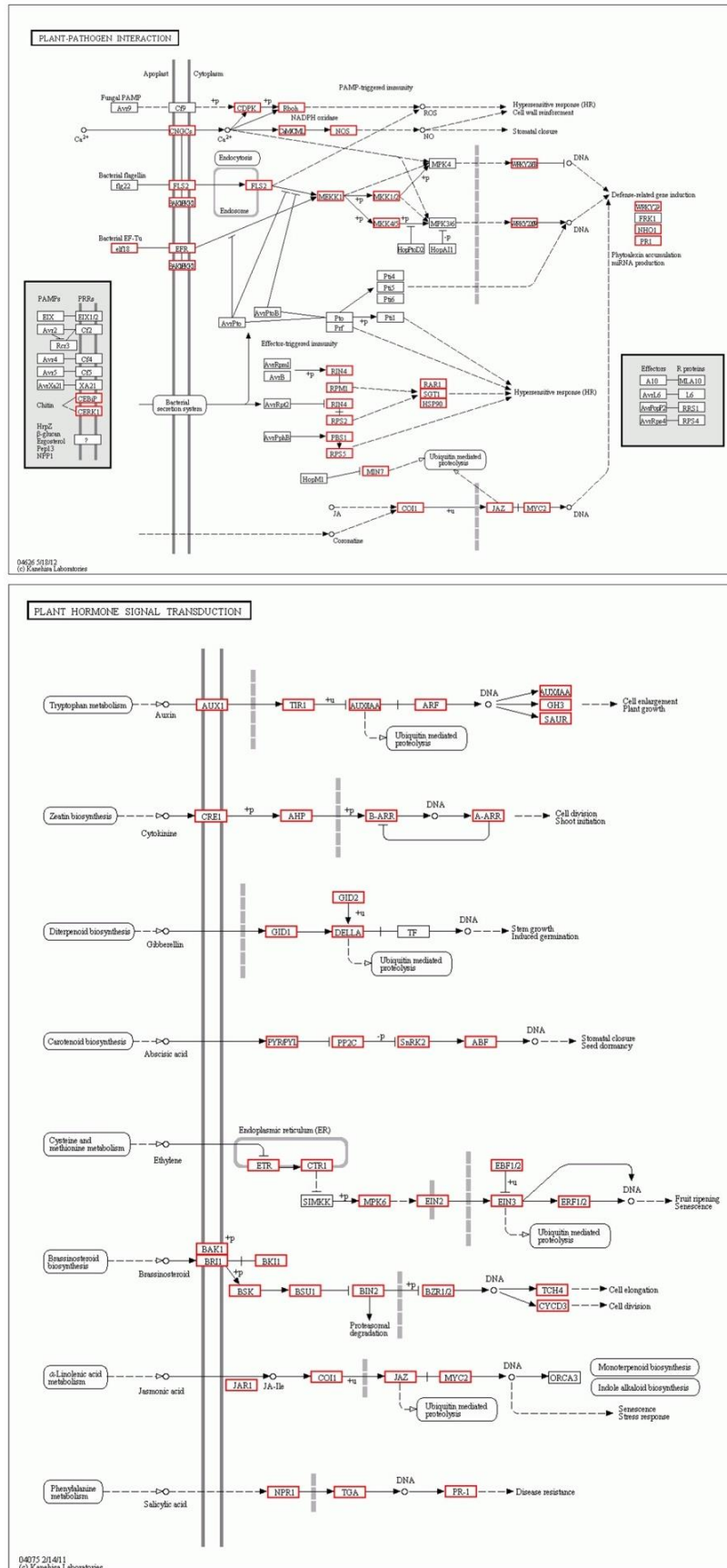


Figure 7. Annotation of *R. stricta* (1A) unigenes to KEGG pathways

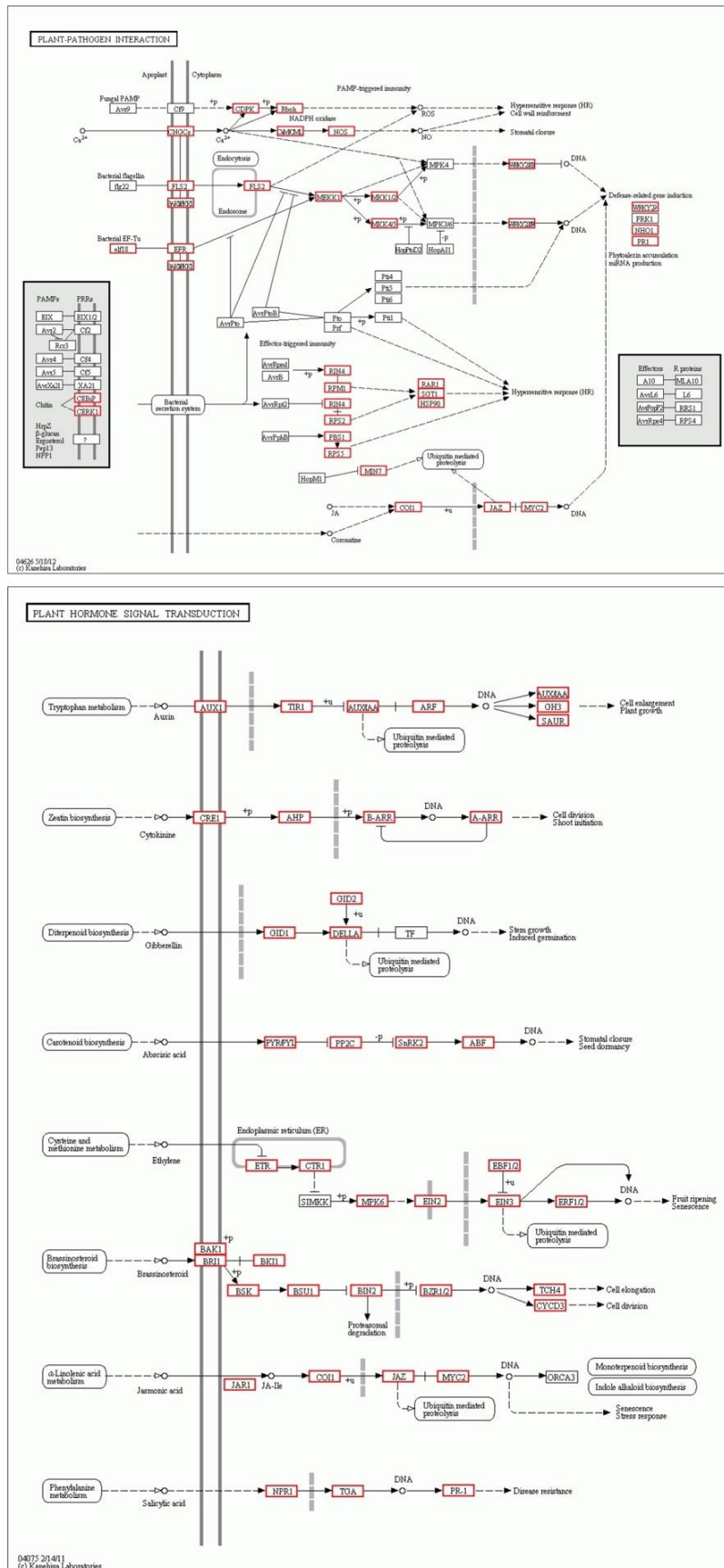


Figure 8. Annotation of *S. italica* (4A) unigenes to KEGG pathways

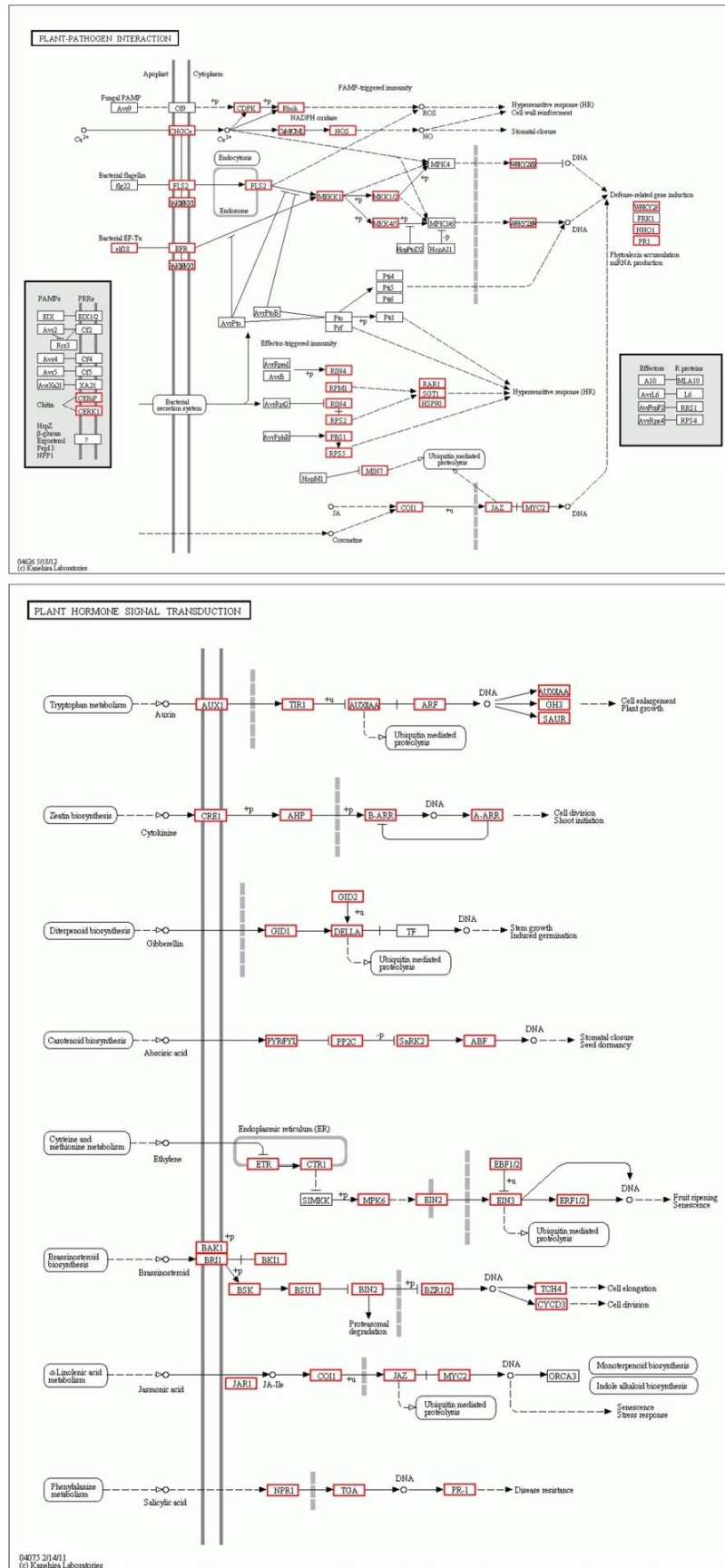


Figure 9. Annotation of *Z. simplex* (5A) unigenes to KEGG pathways

Comparison of KEGG pathways

KEGG is a bioinformatics resource for biological interpretation of gene sequence and linking genomic information in terms of molecular networks (Kanehisa et al., 2012). Based upon sequence similarity, unigenes sequences were annotated in terms of KO (KEGG Orthology). Furthermore, the set of KOs used to construct KEGG pathways and modules for interpreting the functions (Kanehisa et al., 2012). In order to identify the most common pathways and functions activated during the abiotic stress, the distribution of unigenes and KOs assigned to 128 KEGG pathways among *R. stricta*, *S. italica*, and *Z. simplex* were analyzed. We found that the top six common pathways annotated with large number of unigenes are “Metabolic pathways” (ko01100), “Biosynthesis of secondary metabolites” (ko01110), “Plant-pathogen interaction” (ko04626), “Plant hormone signal transduction” (ko04075), “Spliceosome” (ko03040), and “RNA transport” (ko03013) (Tables 5 and S7). In addition, we observed that the top six common pathways annotated with a large number of KOs are “Metabolic pathways” (ko01100), “Biosynthesis of secondary metabolites” (ko01110), “Ribosome” (ko03010), “Spliceosome” (ko03040), and “RNA transport” (ko03013) and “Purine metabolism” (ko00230) (Tables 5 and S7). In order to identify the common KOs present in the KEGG pathways of three plants, we analyzed the KOs data and observed that most of the KOs are common among *R. stricta*, *S. italica* and *Z. simplex* (Fig. 10).

Table 5. Comparative analysis of number of unigenes and KOs present in different pathways

#Pathway	KEGG pathway ID	Unigenes count			KOs count		
		<i>R. stricta</i>	<i>S. italica</i>	<i>Z. simplex</i>	<i>R. stricta</i>	<i>S. italica</i>	<i>Z. simplex</i>
Metabolic pathways	ko01100	6077	4976	6657	795	796	797
Biosynthesis of secondary metabolites	ko01110	2970	2399	2940	365	365	368
Plant-pathogen interaction	ko04626	1316	1296	1342	37	37	37
Plant hormone signal transduction	ko04075	1112	1305	1619	41	41	41
Spliceosome	ko03040	1095	972	1079	101	101	101
RNA transport	ko03013	849	872	1003	97	97	97
Protein processing in endoplasmic reticulum	ko04141	780	608	799	76	76	76
RNA degradation	ko03018	595	492	692	50	49	49
Purine metabolism	ko00230	560	559	595	87	88	88
Ribosome	ko03010	405	471	539	127	124	124

A previous study used ChIP-Seq and RNA-Seq techniques and examined the histone methylation and gene expression pattern under the drought condition in rice (Zong et al., 2013). The study identified the top ten pathways up-regulated in rice under drought stress conditions (Zong et al., 2013). We compared the top ten pathways activated in *R. stricta*, *S. italica* and *Z. simplex* (annotated with a large number of unigenes) with the previous study. We found that the “Metabolic pathways” (ko01100), “Biosynthesis of secondary metabolites” (ko01110), “Plant-pathogen interaction” (ko04626), and “Plant hormone signal transduction” (ko04075) are commonly activated pathways during the abiotic stress condition (Table 5) (Zong et al., 2013). Another study revealed that in response to drought stress in grapevine, differentially expressed genes are associated with “Metabolic pathways” (ko01100), “Biosynthesis of secondary metabolites” (ko01110), and “Plant-pathogen interaction” (ko04626) (Haider et al., 2017).

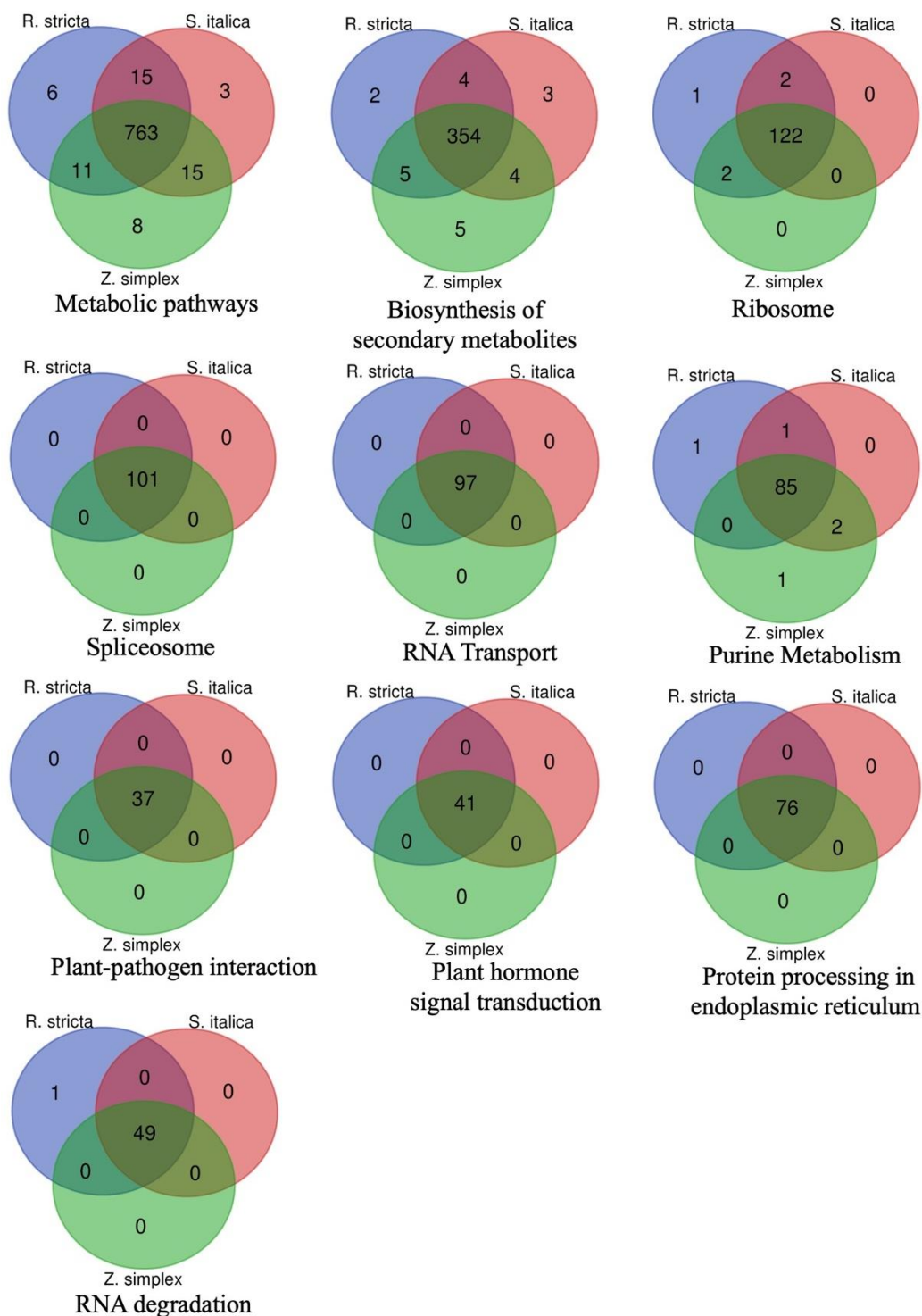


Figure 10. Venn diagrams showing common KOs annotated in different KEGG pathways

Furthermore, our finding also supports the previous study, which showed that under salinity stress, top ten KEGG pathways: “Metabolic pathways” (ko01100), “Biosynthesis of secondary metabolites” (ko01110), “Plant-pathogen interaction” (ko04626), “Plant hormone signal transduction” (ko04075), “Spliceosome” (ko03040), and “RNA transport” (ko03013), “Ribosome” ko03010, “Protein processing in

endoplasmic reticulum” (ko04141), “Endocytosis” ko04144, and “Purine metabolism” (ko00230) are activated in *Chrysanthemum crassum* (Guan et al., 2017).

Our findings are aligned with previous studies and observed that simultaneously abiotic stress of drought, heat, and salinity significantly induce several common pathways in *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A). Thus, suggesting that these pathways play an essential role in maintaining the cellular homeostasis, plants’ growth, and development under harsh stressed conditions (Figs. 7, 8, 9, 10; Table S7).

SSR discovery

SSRs are the most important molecular markers to study the evolutionary, genetics, and breeding in plants (da Costa et al., 2017). SSR markers has been widely used for the genetic linkage mapping (Hong et al., 2010), to understand the genetic diversity (Ren et al., 2014; Feng et al., 2016), and to identify of species (Shirasawa et al., 2013). Furthermore, a gene contains several functional *cis*-elements which influence the transcription and translation of mRNA and regulations of gene expression (Ahmed et al., 2011). Therefore, SSRs motifs and their variations can affect gene expression, translation, mRNA splicing, and mRNA export to the cytoplasm. A strong selective pressure would have been forced to select the SSRs on the coding region compared to the rest of the genomic region.

The unigenes were investigated for the SSR profiles and found the following results: (a) *R. stricta* (1A): There were 10,706 SSRs detected in 9,506 unigenes out of 71,116 examined unigenes (Fig. 11A). (b) *S. italica* (4A): There was 10,277 SSRs detected in 8,465 unigenes out of 59,274 examined unigenes (Fig. 11B); (c) *Z. simplex* (5A): There was 12,374 SSRs detected in 10422 unigenes out of 70,300 examined unigenes (Fig. 11C).

SNP discovery

In plants, SNP is the most usual type of DNA variation. SNPs could alter the structure and stability of proteins in both plants and animals (Milenkovic et al., 2018; Alzahrani et al., 2020; Bhardwaj et al., 2016). Detection of SNPs helps to understand plant species’ genetic diversity, identify the complex traits including biotic and abiotic stress tolerance, and improve crop yields (Rafalski, 2002; Villordo-Pineda et al., 2015; Parida et al., 2012). Our analysis of SNPs profile found the following results:

(a) *R. stricta* (1A): There are 37,754 high-quality SNPs were detected from the unigenes including 22,845 transitions and 14,909 transversions (Fig. 12A). (b) *S. italica* (4A): There are 16,530 high-quality SNPs were detected from the unigenes including 10,406 transitions and 6,124 transversions (Fig. 12B). (c) *Z. simplex* (5A): There are 64,996 high-quality SNPs were detected from the unigenes including 39,014 transitions and 25,982 transversions. (Fig. 12C).

There are different classes of small non-coding RNAs, including miRNAs, siRNA, and tasiRNAs, are produced in plants for regulating gene expression at transcriptional or post-transcriptional levels (Khraiwesh et al., 2012). Under the normal physiology, abiotic, and biotic stress condition, plants produced various small non-coding RNA to modulate the gene expression for plant adaptation (Khraiwesh et al., 2012; Ahmed et al., 2014). Therefore, siRNA could be a design and expressed against specific genes or SNPs for gene silencing for better abiotic-stress tolerant plants (Ahmed et al., 2015; Ahmed and Raghava, 2011; Ahmed et al., 2020).

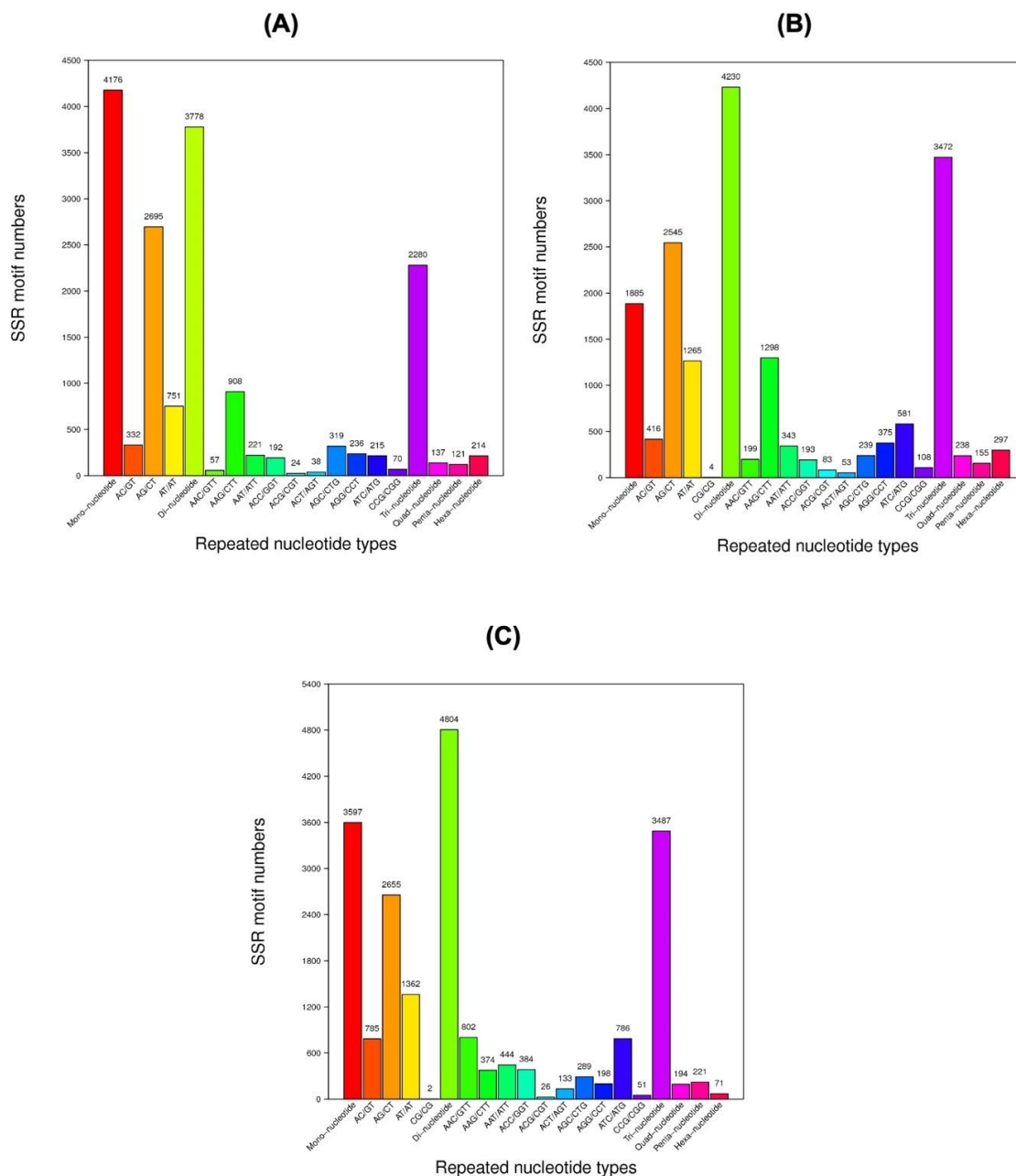


Figure 11. SSR distribution of unigenes. (A) *R. stricta* (1A); (B) *S. italica* (4A); and (C) *Z. simplex* (5A)

Protein-coding sequence prediction (CDS) and expression level analysis of Unigenes

For the analyses of CDS and predicted proteins. The length-frequency distributions of these Unigene CDSs by BLASTx and ESTscan were provided in *Figure 13* for *R. stricta*; *Figure 14* for *S. italica*; and *Figure 15* for *Z. simplex*. The complete list of CDS predicted sequence are provided in Supplementary data: (a) For *R. stricta*: *Tables S8, S9, S10, and S11*; (b) For *S. italica*: *Tables S12, S13, S14, and S15*; For *Z. simplex*: *Tables S16, S17, S18, and S19*. The expression level of Unigene from *R. stricta*, *S. italica* and *Z. simplex* is provided in supplementary data *Tables S20, S21, and S22*, respectively.

Detailed annotation of unigenes and their statistics of *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A) are provided in the “Annotation_Folder” in supplementary data.

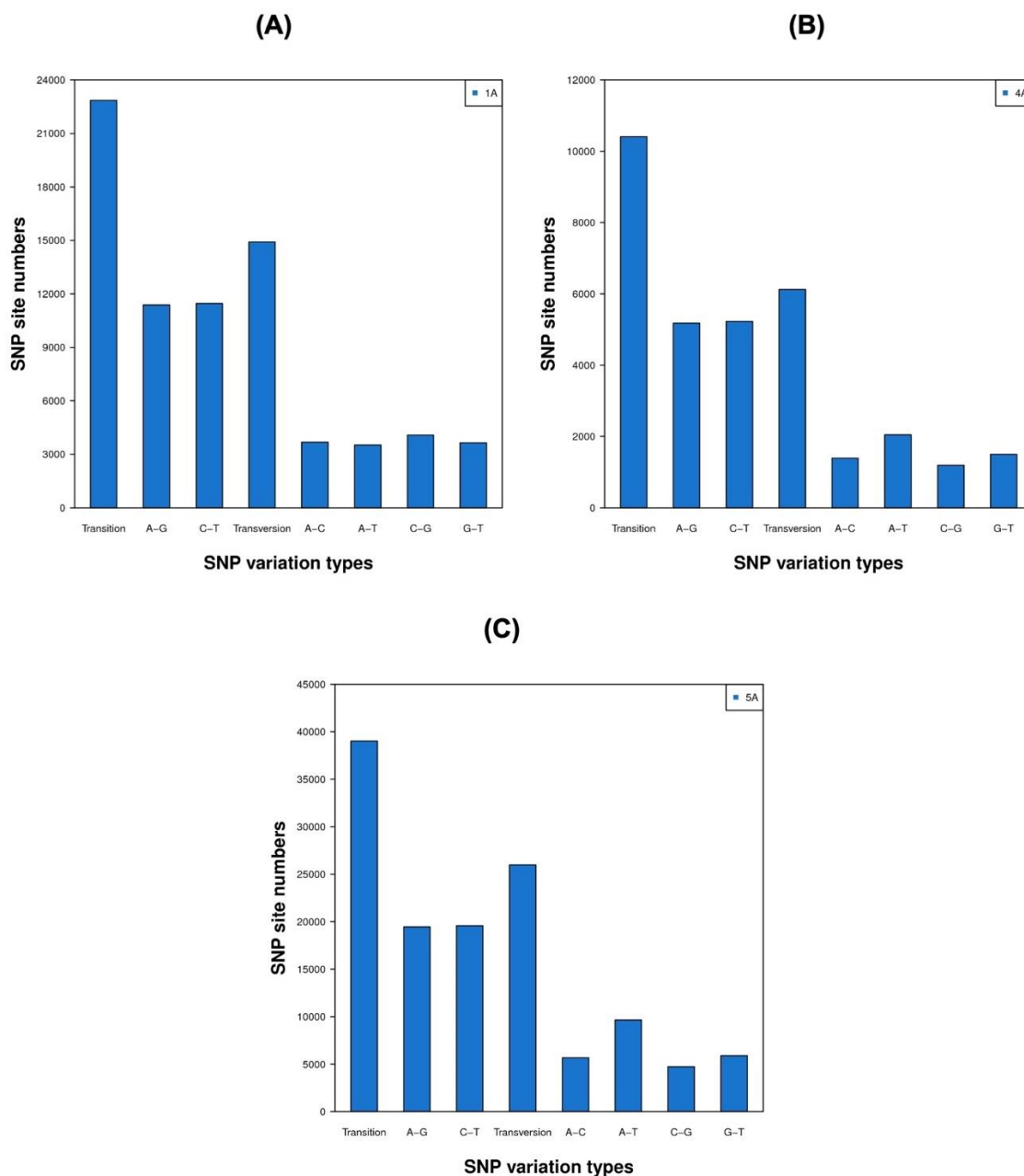


Figure 12. SNP distribution of unigenes. (A) *R. stricta* (1A); (B) *S. italica* (4A); and (C) *Z. simplex* (5A)

Transcription factors responding to combined abiotic stresses

Transcription factors (TFs) are important regulators of gene expression through binding to *cis*-acting elements of target genes. Alteration in the expression of TFs is one of the important regulatory mechanisms in plants to tolerate different abiotic stresses (REF). We analyzed all Unigenes sequences of *R. stricta* (1A), *S. italica* (4A), and *Z. simplex* (5A) for TFs prediction using PlantTFcat and found that 4426, 3895, and 5887

Unigene sequences from *R. stricta* (Table S23), *S. italica* (Table S24) and *Z. simplex* (Table S25) were predicted as TFs, respectively. A comparative analysis of the TFs-family expressed in these three plants was given in Table S26. Furthermore, we identified the most abundant TF-families in unigenes across these plants. For this, we searched the TF-families associated with more than 100 unigenes across these plants and identified seven TFs: C2H2, C3H, CCHC(Zn), MYB-HB-like, PHD, WD40-like, and bHLH (Table 6).

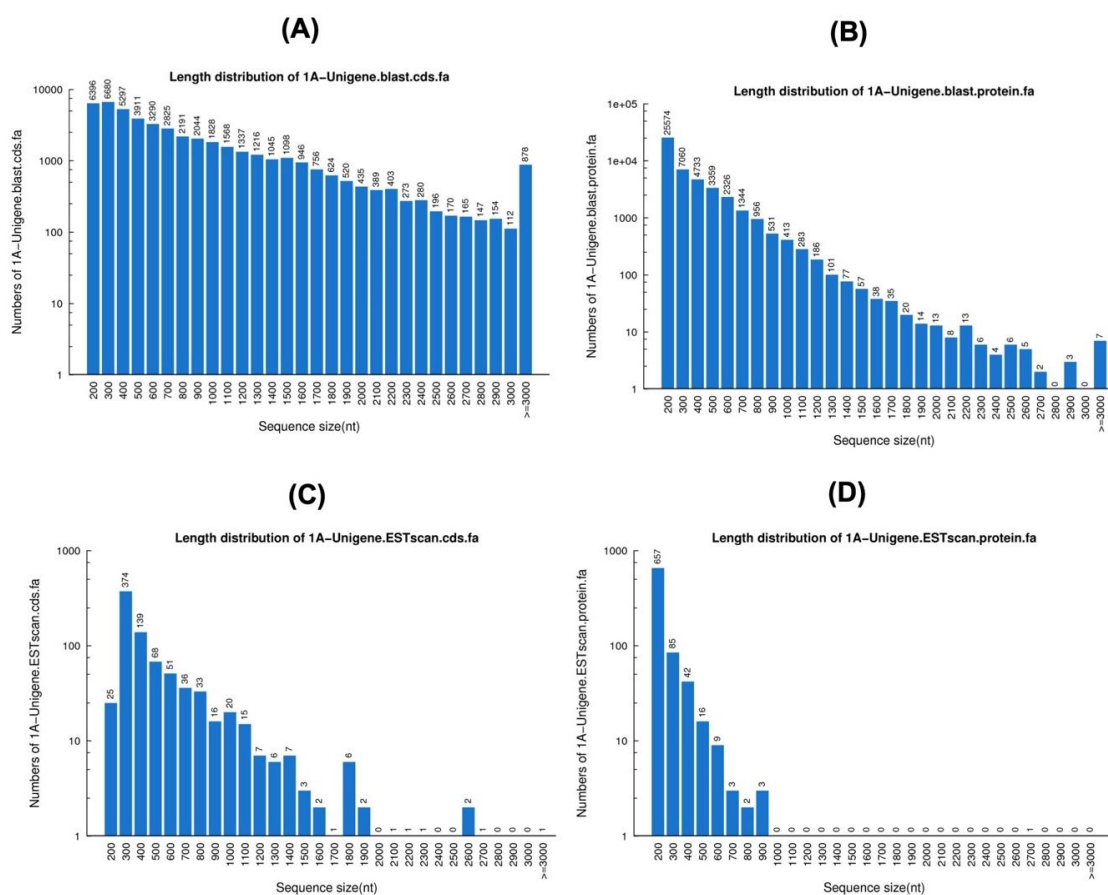


Figure 13. Protein coding sequence (CDS) region prediction of *R. stricta*: (A) The length distribution of CDS nucleotide produced by searching unigenes sequences against various protein databases using BLASTx. (B) The length distribution of proteins predicted from CDS sequence using BLASTx. (C) The length distribution of CDS using ESTscan. (D) The length distribution of protein using ESTscan. The horizontal coordinates are CDS length, and the vertical coordinates are numbers of CDS

Several studies found that the zinc finger TFs families, including C2H2, C3H, and CCHC(Zn), are playing a very important role in plant growth, development, and abiotic stress such as high salt, drought, and high-temperature responses (Kielbowicz-Matuk, 2012; Han et al., 2020; Jiang et al., 2014; Lee and Kang, 2016). In addition, PHD family proteins act as epigenome readers and recruit various chromatin regulators and transcription factors that control gene expression which responds to environmental stresses (Sun et al., 2017; Sanchez and Zhou, 2011). Previous studies showed that overexpression of HbMYB1 TF in tobacco leads to enhanced resistance to UV-B stress

and suppresses stress induces cell death (Li et al., 2015; Peng et al., 2011). A study found in wheat that WD40-like TFs are involved in abiotic stress response (Kong et al., 2015). bHLH TF family response to plant abiotic stresses such as drought, salt, and cold stress (Sun et al., 2018).

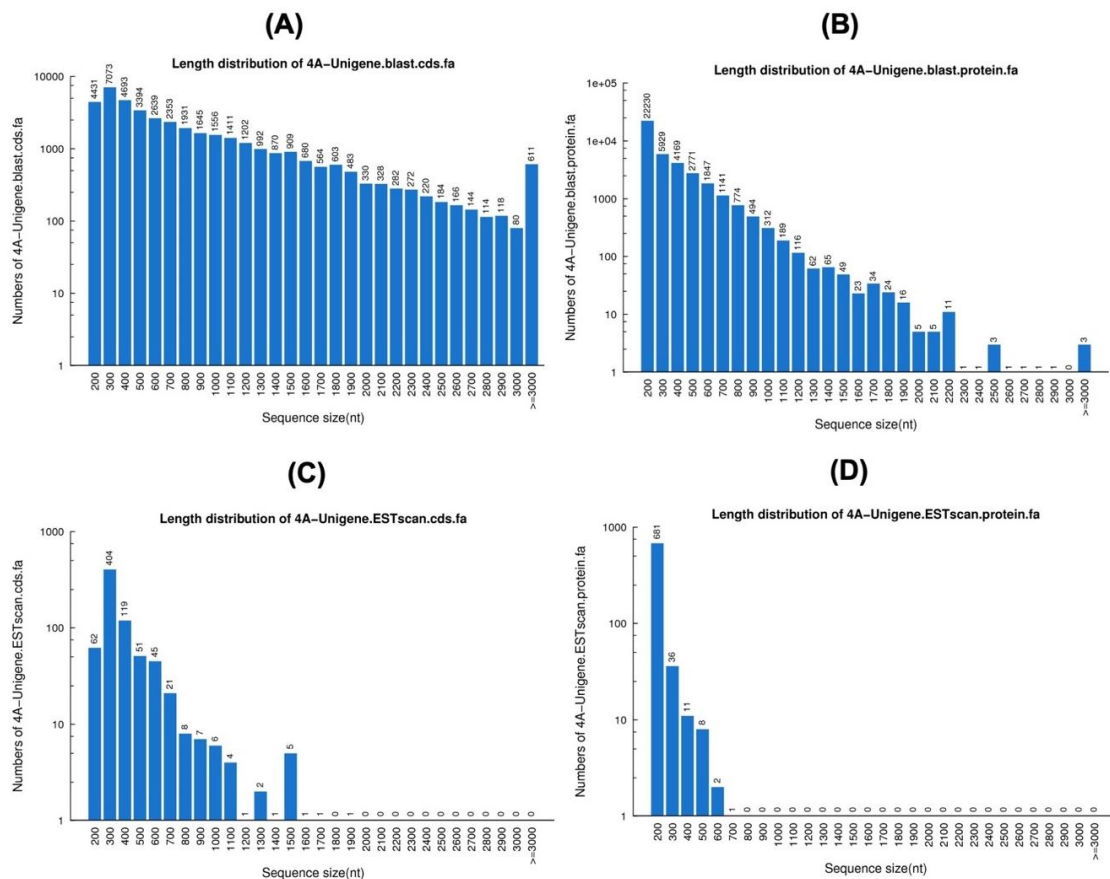


Figure 14. Protein coding sequence (CDS) region prediction of *S. italica*: (A) The length distribution of CDS nucleotide produced by searching unigenes sequences against various protein databases using BLASTx. (B) The length distribution of proteins predicted from CDS sequence using BLASTx. (C) The length distribution of CDS using ESTscan. (D) The length distribution of protein using ESTscan. The horizontal coordinates are CDS length, and the vertical coordinates are numbers of CDS

Table 6. Comparative analysis of number of unigenes belongs to different TFs

TFs-family	Family_type	Unigenes count		
		<i>R. stricta</i>	<i>S. italica</i>	<i>Z. simplex</i>
C2H2	Transcription factor	759	685	923
C3H	Transcription factor	156	152	201
CCHC(Zn)	Transcription factor interactor and regulator	213	264	353
MYB-HB-like	Transcription factor	248	225	358
PHD	Chromatin regulator	213	166	378
WD40-like	Transcription factor	868	654	890
bHLH	Transcription factor	129	131	177

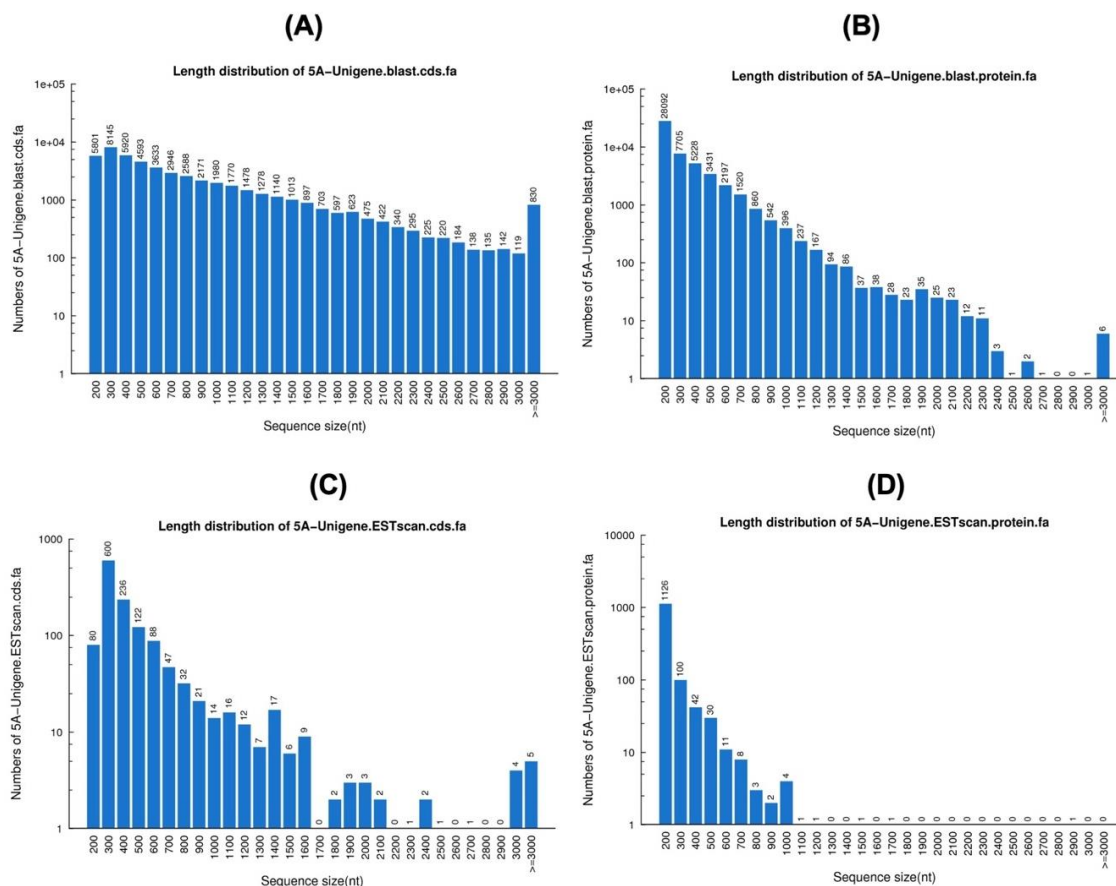


Figure 15. Protein coding sequence (CDS) region prediction of *Z. simplex*: (A) The length distribution of CDS nucleotide produced by searching unigenes sequences against various protein databases using BLASTx. (B) The length distribution of proteins predicted from CDS sequence using BLASTx. (C) The length distribution of CDS using ESTscan. (D) The length distribution of protein using ESTscan. The horizontal coordinates are CDS length, and the vertical coordinates are numbers of CDS

Abiotic stress triggers the osmotic and oxidative stresses signaling in a cell cause damage to the structure of proteins and cell membrane. However, stress tolerance plants induce the signaling which activates the transcription factors that express the stress-responsive genes to repair the damaged proteins and maintain cellular homeostasis (Wang et al., 2003). Our study could be used to understand gene regulatory networks during different stresses and their response mechanisms. Furthermore, finding TFs might be used to develop stress-tolerant crops that could increase the crop yield in the desert climate.

Conclusions

This study reported the first-time de novo transcriptome sequencing of leaf tissues of three abiotic stress plants: (a) *R. stricta*, (b) *S. italica*, and (c) *Z. simplex*. The illumine platform generated sequencing data assembled into 71,116; 59,274; 70,300 unigenes for *R. stricta*; *S. italica*; and *Z. simplex*, respectively. Annotation and analysis of unigenes identified several candidate genes involved in abiotic stress, growth, development, and

signal transduction pathways. Furthermore, our study identified the SSRs and SNPs in the datasets. Comparative analysis of Unigenes among all three plants identified seven transcription factor and transcriptional regulator responses to plants' abiotic stresses.

Abiotic stress triggers the osmotic and oxidative stresses signaling in a cell cause damage to the structure of proteins and cell membrane. However, stress tolerance plants induce the signaling which activates the transcription factors that express the stress-responsive genes to repair the damaged proteins and maintain cellular homeostasis (Wang et al., 2003). Our study could be used to understand how gene regulatory networks behave during different stresses and how networks induce the response mechanisms. Therefore, this study and datasets would be very useful for understanding the mechanism of abiotic stress management in plant studies in the future, and will also help improve crop productions in stressed geographical locations.

Acknowledgments. This work was funded by the Deanship of Scientific Research (DSR), University of Jeddah, Jeddah, under grant No. (UJ-01-18-ICP). The authors, therefore, acknowledge with thanks DSR technical and financial support.

Author contributions. Dr. Mohammed Baeshen is the PI of the project, and along with Prof. Nabih Baeshen they designed the project along with Prof. John Heulsenbeck, Dr. Baeshen and Prof Baeshen determinate the study location, collected the samples of the study, supervised the nucleic acid extraction and following up the sequencing results for further transcriptomic analysis. Dr. Firoz along with team wrote the project proposal and analyzed along with Prof. Tarek Musa the transcriptomic data. All other part of the team revised the manuscript and prepare it for publication.

Declaration of competing interests. The authors declare that they have no conflict of interests.

REFERENCES

- [1] Ahmed, F., Raghava, G. P. (2011): Designing of highly effective complementary and mismatch siRNAs for silencing a gene. – *PLoS One* 6(8): e23443.
- [2] Ahmed, F., Benedito, V. A., Zhao, P. X. (2011): Mining functional elements in messenger rnas: overview, challenges, and perspectives. – *Front Plant Sci* 2: 84.
- [3] Ahmed, F., Senthil-Kumar, M., Lee, S., Dai, X., Mysore, K. S., Zhao, P. X. (2014): Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in *Arabidopsis thaliana*. – *RNA Biol* 11(11): 1414-29.
- [4] Ahmed, F., Dai, X., Zhao, P. X. (2015): Bioinformatics tools for achieving better gene silencing in plants. – *Methods Mol Biol* 1287: 43-60.
- [5] Ahmed, F., Senthil-Kumar, M., Dai, X., Ramu, V. S., Lee, S., Mysore, K. S., Zhao, P. X. (2020): pssRNAit-a web server for designing effective and specific plant siRNAs with genome-wide off-target assessment. – *Plant Physiol*.
- [6] Al-Whaibi, M. H. (2011): Plant heat-shock proteins: a mini review. – *Journal of King Saud University - Science* 23(2): 139-150.
- [7] Alzahrani, F. A., Ahmed, F., Sharma, M., Rehan, M., Mahfuz, M., Baeshen, M. N., Hawsawi, Y., Almatrafi, A., Alsagaby, S. A., Kamal, M. A., Warsi, M. K., Choudhry, H., Jamal, M. S. (2020): Investigating the pathogenic SNPs in BLM helicase and their biological consequences by computational approach. – *Scientific Reports* 10(1): 12377.
- [8] An, Y., Zhang, M., Liu, G., Han, R., Liang, Z. (2013): Proline accumulation in leaves of *Periploca sepium* via both biosynthesis up-regulation and transport during recovery from severe drought. – *PLoS One* 8(7): e69942.

- [9] Arisha, M. H., Ahmad, M. Q., Tang, W., Liu, Y., Yan, H., Kou, M., Wang, X., Zhang, Y., Li, Q. (2020): RNA-sequencing analysis revealed genes associated drought stress responses of different durations in hexaploid sweet potato. – *Sci Rep* 10(1): 12573.
- [10] Baeshen, M. N., Moussa, T. A. A., Ahmed, F., Abulfaraj, A. A., Jalal, R. S., Majeed, M. A., Baeshen, N. A., Huelsenbeck, J. P. (2020): Diversity profiling of associated bacteria from the soils of stress tolerant plants from seacoast of Jeddah, Saudi Arabia. – *Applied Ecology and Environmental Research* 18(6): 8217-8231.
- [11] Beier, S., Thiel, T., Munch, T., Scholz, U., Mascher, M. (2017): MISA-web: a web server for microsatellite prediction. – *Bioinformatics* 33(16): 2583-2585.
- [12] Bhardwaj, A., Dhar, Y. V., Asif, M. H., Bag, S. K. (2016): In silico identification of SNP diversity in cultivated and wild tomato species: insight from molecular simulations. – *Sci Rep* 6: 38715.
- [13] Cheng, M. C., Liao, P. M., Kuo, W. W., Lin, T. P. (2013): The Arabidopsis ETHYLENE RESPONSE FACTOR1 regulates abiotic stress-responsive gene expression by binding to different cis-acting elements in response to different stress signals. – *Plant Physiol* 162(3): 1566-82.
- [14] Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., Robles, M. (2005): Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. – *Bioinformatics* 21(18): 3674-6.
- [15] da Costa, Z. P., Munhoz, C. F., Vieira, M. L. C. (2017): Report on the development of putative functional SSR and SNP markers in passion fruits. – *BMC Res Notes* 10(1): 445.
- [16] Dai, X., Sinharoy, S., Udvardi, M., Zhao, P. X. (2013): PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. – *BMC Bioinformatics* 14: 321.
- [17] Dassanayake, M., Haas, J. S., Bohnert, H. J., Cheeseman, J. M. (2009): Shedding light on an extremophile lifestyle through transcriptomics. – *New Phytol* 183(3): 764-75.
- [18] Dassanayake, M., Oh, D. H., Haas, J. S., Hernandez, A., Hong, H., Ali, S., Yun, D. J., Bressan, R. A., Zhu, J. K., Bohnert, H. J., Cheeseman, J. M. (2011): The genome of the extremophile crucifer *Thellungiella parvula*. – *Nat Genet* 43(9): 913-8.
- [19] Deng, S., Ma, J., Zhang, L., Chen, F., Sang, Z., Jia, Z., Ma, L. (2019): De novo transcriptome sequencing and gene expression profiling of *Magnolia wufengensis* in response to cold stress. – *BMC Plant Biol* 19(1): 321.
- [20] Fahad, S., Bajwa, A. A., Nazir, U., Anjum, S. A., Farooq, A., Zohaib, A., Sadia, S., Nasim, W., Adkins, S., Saud, S., Ihsan, M. Z., Alharby, H., Wu, C., Wang, D., Huang, J. (2017): Crop production under drought and heat stress: plant responses and management options. – *Front Plant Sci* 8: 1147.
- [21] Feng, S., He, R., Lu, J., Jiang, M., Shen, X., Jiang, Y., Wang, Z., Wang, H. (2016): Development of SSR markers and assessment of genetic diversity in medicinal *Chrysanthemum morifolium* cultivars. – *Front Genet* 7: 113.
- [22] Fiaz, S., Noor, M. A., Aldosri, F. O. (2016): Achieving food security in the Kingdom of Saudi Arabia through innovation: potential role of agricultural extension. – *Journal of the Saudi Society of Agricultural Sciences*.
- [23] Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011): Full-length transcriptome assembly from RNA-Seq data without a reference genome. – *Nat Biotechnol* 29(7): 644-52.
- [24] Guan, Z., Feng, Y., Song, A., Shi, X., Mao, Y., Chen, S., Jiang, J., Ding, L., Chen, F. (2017): Expression profiling of *Chrysanthemum crassum* under salinity stress and the initiation of morphological changes. – *PLoS One* 12(4): e0175972.
- [25] Haider, M. S., Kurjogi, M. M., Khalil-Ur-Rehman, M., Fiaz, M., Pervaiz, T., Jiu, S., Haifeng, J., Chen, W., Fang, J. (2017): Grapevine immune signaling network in response

- to drought stress as revealed by transcriptomic analysis. – *Plant Physiol Biochem* 121: 187-195.
- [26] Han, G., Lu, C., Guo, J., Qiao, Z., Sui, N., Qiu, N., Wang, B. (2020): C2H2 zinc finger proteins: master regulators of abiotic stress responses in plants. – *Frontiers in Plant Science* 11(115).
- [27] Hong, Y., Chen, X., Liang, X., Liu, H., Zhou, G., Li, S., Wen, S., Holbrook, C. C., Guo, B. (2010): A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. – *BMC Plant Biol* 10: 17.
- [28] Jiang, A.-L., Xu, Z.-S., Zhao, G.-Y., Cui, X.-Y., Chen, M., Li, L.-C., Ma, Y.-Z. (2014): Genome-wide analysis of the C3H zinc finger transcription factor family and drought responses of members in *Aegilops tauschii*. – *Plant Molecular Biology Reporter* 32(6): 1241-1256.
- [29] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. (2012): KEGG for integration and interpretation of large-scale molecular data sets. – *Nucleic Acids Res* 40(Database issue): D109-14.
- [30] Kang, Z., Babar, M. A., Khan, N., Guo, J., Khan, J., Islam, S., Shrestha, S., Shahi, D. (2019): Comparative metabolomic profiling in the roots and leaves in contrasting genotypes reveals complex mechanisms involved in post-anthesis drought tolerance in wheat. – *PLoS One* 14(3): e0213502.
- [31] Kerr, S. C., Gaiti, F., Tanurdzic, M. (2019): De novo plant transcriptome assembly and annotation using illumina RNA-seq reads. – *Methods Mol Biol* 1933: 265-275.
- [32] Khraiweh, B., Zhu, J. K., Zhu, J. (2012): Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. – *Biochim Biophys Acta* 1819(2): 137-48.
- [33] Kielbowicz-Matuk, A. (2012): Involvement of plant C(2)H(2)-type zinc finger transcription factors in stress responses. – *Plant Sci* 185-186: 78-85.
- [34] Kong, D., Li, M., Dong, Z., Ji, H., Li, X. (2015): Identification of TaWD40D, a wheat WD40 repeat-containing protein that is associated with plant tolerance to abiotic stresses. – *Plant Cell Rep* 34(3): 395-410.
- [35] Krasensky, J., Jonak, C. (2012): Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. – *J Exp Bot* 63(4): 1593-608.
- [36] Lee, K., Kang, H. (2016): Emerging roles of RNA-binding proteins in plant growth, development, and stress responses. – *Mol Cells* 39(3): 179-85.
- [37] Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., Wang, J. (2009): SOAP2: an improved ultrafast tool for short read alignment. – *Bioinformatics* 25(15): 1966-7.
- [38] Li, C., Ng, C. K. Y., Fan, L.-M. (2015): MYB transcription factors, active players in abiotic stress signaling. – *Environmental and Experimental Botany* 114: 80-91.
- [39] Li, W., Zhang, L., Ding, Z., Wang, G., Zhang, Y., Gong, H., Chang, T., Zhang, Y. (2017): De novo sequencing and comparative transcriptome analysis of the male and hermaphroditic flowers provide insights into the regulation of flower formation in andromonoecious *Taihangia rupestris*. – *BMC Plant Biol* 17(1): 54.
- [40] Mbandi, S. K., Hesse, U., Rees, D. J., Christoffels, A. (2014): A glance at quality score: implication for de novo transcriptome reconstruction of Illumina reads. – *Front Genet* 5: 17.
- [41] Milenkovic, V. M., Bader, S., Sudria-Lopez, D., Siebert, R., Brandl, C., Nothdurfter, C., Weber, B. H. F., Rupprecht, R., Wetzel, C. H. (2018): Effects of genetic variants in the TSPO gene on protein structure and stability. – *PLoS One* 13(4): e0195627.
- [42] Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., Shigeoka, S. (2006): Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. – *Plant J* 48(4): 535-47.
- [43] Pan, L., Zhang, X., Wang, J., Ma, X., Zhou, M., Huang, L., Nie, G., Wang, P., Yang, Z., Li, J. (2016): Transcriptional profiles of drought-related genes in modulating metabolic processes and antioxidant defenses in *Lolium multiflorum*. – *Front Plant Sci* 7: 519.

- [44] Parida, S. K., Mukerji, M., Singh, A. K., Singh, N. K., Mohapatra, T. (2012): SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. – *BMC Genomics* 13: 426.
- [45] Park, S., Ruhlman, T. A., Sabir, J. S., Mutwakil, M. H., Baeshen, M. N., Sabir, M. J., Baeshen, N. A., Jansen, R. K. (2014): Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. – *BMC Genomics* 15: 405.
- [46] Peng, S.-Q., Wu, K.-X., Huang, G.-X., Chen, S.-C. (2011): HbMyb1, a Myb transcription factor from *Hevea brasiliensis*, suppresses stress induced cell death in transgenic tobacco. – *Plant Physiology and Biochemistry* 49(12): 1429-1435.
- [47] Prash, C. M., Sonnewald, U. (2013): Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. – *Plant Physiol* 162(4): 1849-66.
- [48] Puranik, S., Jha, S., Srivastava, P. S., Sreenivasulu, N., Prasad, M. (2011): Comparative transcriptome analysis of contrasting foxtail millet cultivars in response to short-term salinity stress. – *J Plant Physiol* 168(3): 280-7.
- [49] Rafalski, A. (2002): Applications of single nucleotide polymorphisms in crop genetics. – *Curr Opin Plant Biol* 5(2): 94-100.
- [50] Ren, X., Jiang, H., Yan, Z., Chen, Y., Zhou, X., Huang, L., Lei, Y., Huang, J., Yan, L., Qi, Y., Wei, W., Liao, B. (2014): Genetic diversity and population structure of the major peanut (*Arachis hypogaea* L.) cultivars grown in China by SSR markers. – *PLoS One* 9(2): e88091.
- [51] Rizhsky, L., Liang, H., Mittler, R. (2002): The combined effect of drought stress and heat shock on gene expression in tobacco. – *Plant Physiol* 130(3): 1143-51.
- [52] Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., Mittler, R. (2004): When defense pathways collide. The response of *Arabidopsis* to a combination of drought and heat stress. – *Plant Physiol* 134(4): 1683-96.
- [53] Rodriguez, M. C., Edsgard, D., Hussain, S. S., Alquezar, D., Rasmussen, M., Gilbert, T., Nielsen, B. H., Bartels, D., Mundy, J. (2010): Transcriptomes of the desiccation-tolerant resurrection plant *Craterostigma plantagineum*. – *Plant J* 63(2): 212-28.
- [54] Sabir, J. S., Jansen, R. K., Arasappan, D., Calderon, V., Noutahi, E., Zheng, C., Park, S., Sabir, M. J., Baeshen, M. N., Hajrah, N. H., Khiyami, M. A., Baeshen, N. A., Obaid, A. Y., Al-Malki, A. L., Sankoff, D., El-Mabrouk, N., Ruhlman, T. A. (2016): The nuclear genome of *Rhazya stricta* and the evolution of alkaloid diversity in a medically relevant clade of Apocynaceae. – *Sci Rep* 6: 33782.
- [55] Sanchez, R., Zhou, M. M. (2011): The PHD finger: a versatile epigenome reader. – *Trends Biochem Sci* 36(7): 364-72.
- [56] Sarwar, M. B., Ahmad, Z., Rashid, B., Hassan, S., Gregersen, P. L., Leyva, M. O., Nagy, I., Asp, T., Husnain, T. (2019): De novo assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. – *Sci Rep* 9(1): 396.
- [57] Shirasawa, K., Ishii, K., Kim, C., Ban, T., Suzuki, M., Ito, T., Muranaka, T., Kobayashi, M., Nagata, N., Isobe, S., Tabata, S. (2013): Development of capsicum EST-SSR markers for species identification and in silico mapping onto the tomato genome sequence. – *Mol Breed* 31(1): 101-110.
- [58] Sun, M., Jia, B., Yang, J., Cui, N., Zhu, Y., Sun, X. (2017): Genome-wide identification of the PHD-finger family genes and their responses to environmental stresses in *Oryza sativa* L. – *International Journal of Molecular Sciences* 18(9): 2005.
- [59] Sun, X., Wang, Y., Sui, N. (2018): Transcriptional regulation of bHLH during plant response to stress. – *Biochem Biophys Res Commun* 503(2): 397-401.
- [60] Villordo-Pineda, E., González-Chavira, M. M., Giraldo-Carbajo, P., Acosta-Gallegos, J. A., Caballero-Pérez, J. (2015): Identification of novel drought-tolerant-associated SNPs in common bean (*Phaseolus vulgaris*). – *Front Plant Sci* 6: 546.

- [61] Wang, W., Vinocur, B., Altman, A. (2003): Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. – *Planta* 218(1): 1-14.
- [62] Wang, W., Vinocur, B., Shoseyov, O., Altman, A. (2004): Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. – *Trends Plant Sci* 9(5): 244-52.
- [63] Xu, Z. S., Xia, L. Q., Chen, M., Cheng, X. G., Zhang, R. Y., Li, L. C., Zhao, Y. X., Lu, Y., Ni, Z. Y., Liu, L., Qiu, Z. G., Ma, Y. Z. (2007): Isolation and molecular characterization of the *Triticum aestivum* L. ethylene-responsive factor 1 (TaERF1) that increases multiple stress tolerance. – *Plant Mol Biol* 65(6): 719-32.
- [64] Yates, S. A., Chernukhin, I., Alvarez-Fernandez, R., Bechtold, U., Baeshen, M., Baeshen, N., Mutwakil, M. Z., Sabir, J., Lawson, T., Mullineaux, P. M. (2014): The temporal foliar transcriptome of the perennial C3 desert plant *Rhazya stricta* in its natural environment. – *BMC Plant Biol* 14: 2.
- [65] Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., Wang, J. (2006): WEGO: a web tool for plotting GO annotations. – *Nucleic Acids Res* 34(Web Server issue): W293-7.
- [66] Zong, W., Zhong, X., You, J., Xiong, L. (2013): Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress. – *Plant Mol Biol* 81(1-2): 175-88.

ELECTRONIC APPENDIX

This article has electronic appendices.