Genome-wide identification and transcriptional expression analysis of superoxide dismutase (SOD) family in wheat (*Triticum aestivum*)

Wenqiang Jiang[1,2,3], Lei Yang[1], Yiqin He[1], Haotian Zhang[1,3], Wei Li[3], Huaigu Chen[3], Dongfang Ma[1,2]*, Junliang Yin[1]**

[1]Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education/Hubei Collaborative Innovation Center for Grain Industry/College of Agriculture, Yangtze University, Jingzhou 434000, Hubei, China.

[2] Institute of Plant Protection and Soil Science, Hubei Academy of Agricultural Sciences, Wuhan 430064, Hubei, China.

[3]Institute of Plant Protection, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, Jiangsu, China.

*Corresponding author:

E-mail address: madf@yangtzeu.edu.cn (D. Ma), College of Agriculture, Yangtze University, Jingzhou 434000, Hubei, China

**Corresponding author:

E-mail address: yinjunliang@nwafu.edu.cn (J. Yin), College of Agriculture, Yangtze University, Jingzhou 434000, Hubei, China

ABSTRACT

Superoxide dismutases (SODs) are a key antioxidant enzyme family, which plays a critical function in plant growth and development. Previously, this gene family has been investigated in Arabidopsis and rice. In the present study, it was the first time for us to perform a genome-wide analysis of the SOD gene family in wheat. And using bioinformatics-based methods, 26 SOD genes were identified from the whole genome of wheat, including 17 Cu/Zn-SODs, 6 Fe-SODs, and 3 Mn-SODs. The chromosomal distribution analysis revealed that SOD genes are only distributed on 2, 4 and 7 chromosomes of wheat. Phylogenetic analyses with SODs from wheat and several other species revealed that these SOD proteins can divided into two major categories. SOD1 is mainly composed of Cu/Zn-SODs, and SOD2 is mainly composed of Fe-SODs and Mn-SODs. Gene structure and motif analysis indicated that most of the SOD genes have relatively conserved exon/intron arrangement and motif composition. Analysis of transcriptional data indicated that most of the wheat SOD genes are expressed in almost all the tested tissues and ~~it~~ they possibly have important function in abiotic stress. Taken together, our results provide a basis for further functional research on the SOD gene family in wheat and may facilitate their potential applications in the genetic improvement of wheat.

Subjects Bioinformatics, Genomics, Plant Science

Key words SOD, gene structure, protein characterization, abiotic stress, expression profiles

INTRODUCTION

During the growth process, plants are affected by various adverse factors (such as drought, water damage, heat damage, cold damage, pests and diseases, heavy metal ions, etc.). A variety of abiotic and biotic stresses will result in the production of large amounts of reactive oxygen species (ROS) in plants (Razali et al, 2015). When ROS accumulate~~s~~ in plants, it causes oxidative stress, which destroys biological macromolecules, biofilms, etc., and can cause cell death in severe cases (Foyer&Noctor, 2005, Quan et al, 2010). At the same time, ROS as a signal molecule can regulate many physiological processes during plant growth and development, and participate in various biotic and abiotic stress responses (Mittler 2002; Pitzschke et al. 2006). In the long-term evolution process, plants form a complex antioxidant enzyme system that inhibits ROS accumulation, mainly by superoxide dismutase (SOD), catalase (CAT), peroxidase (POD), ascorbic acid (AsA), glutathione (GH), ascorbate peroxidase (APX), etc. (Alscher, Erturk & Heath, 2002; Valko et al, 2006; Sugimoto et al, 2014; Zhang et al, 2016c). The increase in plant stress resistance may be related to the antioxidant enzyme system in the ~~body~~ cell? (Guo et al. 2017). SOD is widely present in living organisms. As the first enzyme involved in the scavenging reaction of reactive oxygen species, it is involved in almost all physiological and biochemical reactions against various environmental stresses in organisms, and is at the core of antioxidant enzymes (Ahmad et al, 2010; Dong et al, 2013). Fridovieh and Mccor (1969) first revealed the biological function of SOD. SOD can catalyze the conversion of superoxide (O^{2-}) into oxygen (O_2) and hydrogen peroxide (H_2O_2) through disproportionation, and further convert H_2O_2 into water (H_2O) by peroxidase and oxide enzyme to achieve active oxygen removal (Tepperman & Dunsmuir, 1990). SOD plays an important role in scavenging oxygen free radicals, preventing oxygen free radicals from disrupting cell composition, structure and function, and protecting cells from oxidative damage (Ding, 2008).

SOD constitutes the first line of defense for plant body elimination of ROS. It is ubiquitous in the plant kingdom and has many types. Many plants contain a series of SOD isozymes. SOD belongs to a class of metalloproteinases. According to the different metal cofactors in the catalytic site, it can be divided into four types: Cu/Zn-SOD, Mn-SOD, Fe-SOD, and Ni-SOD (Abreu & Cabelli, 2010; Whittaker, 2010). Fe-SOD and Mn-SOD are mainly present in lower plants, and Cu/Zn-SOD is mainly present in higher plants (Xia et al, 2015; Zeng et al, 2014). Further studies found that Fe-SOD is located in chloroplasts, Mn-SOD is located in mitochondria and peroxisomes, and Cu/Zn-SOD is mainly located in chloroplasts and cytoplasm (Dupont et al, 2010).

A large number of studies have shown that the expression of plant SOD gene is controlled by various environmental stresses, and different environmental conditions lead to differences in SOD gene expressions (Xia et al, 2015; Zhang et al, 2016c). The SOD activity in rice (Lin et al, 2009) and pea (Yan et al, 2009) was increased under salt stress. In arid environment, the activity of SOD decreased in peanuts at the early stage of stress, but under severe drought stress, SOD activity increased (Jiang & Ren, 2004). At 4 °C, the Cu/Zn-SOD activity of barley leaves did not change significantly; when the temperature dropped to -3 °C, the Cu/Zn-SOD activity increased significantly (Moses, 2012). Under drought and saline conditions, ~~the~~ high er drought resistance and salt tolerance of the transgenic *AtHDG11* gene was associated with increased~~, while the SOD activity increased,~~ indicating ~~the~~ a role of SOD in plant resistance. When the Arabidopsis *CBF1* (C-repeat-binding factor 1) gene ~~were~~ was transferred to tobacco plants, the SOD activity of tobacco plants was significantly higher than that of the control, which improved the tolerance of transgenic plants to low temperature (Zhang et al, 2010). Overexpression of Mn-SOD in tobacco and maize chloroplasts enhance~~s~~d the protective effect of transgenic tobacco and maize on the plasma membrane and tolerance to herbicide-induced oxygen stress (Bowler et al, 1991; Breusegem et al, 1999). Taken together, these results indicate that enhanced SOD activity in plants can increase plant resistance to a variety of stresses.

Wheat is one of the world's most important food crops, accounting for more than half of total human consumption (Yin et al, 2018). The analysis of SOD gene s can provide ideas for wheat genetic improvement (Zhang et al, 2009). At present, the response of the wheat SOD (TaSOD) gene family and the expression of each gene under different stress conditions has not been reported at the genome-wide level. In this study, we performed genome-wide identification of the SOD gene family in wheat and comprehensively analyzed their phylogenetic relationships, genome distribution, gene structure arrangement, motifs composition, expression profiles in different tissues, and their expression patterns in response to various abiotic stresses. The identification and analysis of the wheat SOD family will lay the foundation for further research on wheat stress resistance in the future.

MATERIALS AND METHODS

Identification of wheat SOD gene family members

Computer-based method was used to identify members of the SOD gene family from wheat reference genome IWGSC RefSeqv1.0 (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies). A total of 8 Arabidopsis SODs (AtSODs), 12 maize SODs (ZmSODs) and 8 rice SODs (OsSODs) protein sequences were retrieved from the Arabidopsis Information Resource (TAIR10) database (http://www.arabidopsis.org/index.jsp), the Maize Genetics And Genomics Database (MaizeGDB) (https://www.maizegdb.org/), and the Rice Genome Annotation Project (RGAP) database (http://rice.plantbiology.msu.edu/),respectively. This information was then used to identify the SOD genes in wheat. Two methods were utilized to search the wheat protein sequences. One used a Hidden Markov Model (HMM) to search against wheat protein sequences and the other used BLASTp (E-value < 1e-5) to investigate the SOD proteins against the wheat genome, followed by Pfam (v31.05) (http://pfam.sanger.ac.uk/search) to supplement whether the obtained sequence contains a SOD specific structural conserved domain and ultimately determined the number of SOD gene family members.

Chromosomal locations and syntenic analysis

The wheat genome GFF3 gene annotation file was obtained from the wheat database IWGSCv1.0 and the gene annotation of wheat SODs (TaSODs) was extracted from the GFF3 file. The start and end location information of TaSODs in correspondence to chromosomes were used to draw the physical map via the software MapInspect. (citation)

Proteins characterization of predicted TaSODs

The characterization analysis of TaSODs was performed by using the protein identification and analysis tools on the ExPASy Server10 (https://prosite.expasy.org/) (Artimo et al, 2012). The features of protein length, is isoelectric point (pI), molecular weight (MW), instability index, atomic composition, and amino acid composition were predicted. The TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and SignalP4.1 (http://www.cbs.dtu.dk/services/SignalP/) online tools were used to predict transmembrane domains and signal peptides of TaSODs (Nielsen, 2017). Subcellular localization prediction of TaSODs was performed by Plant-mPLoc (http://www.csbio.sjtu.edu.cn/cgi-bin/PlantmPLoc.cgi) (Chou & Shen, 2010). TaSODs members were three-dimensionally modelled using the Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/) server at the intensive mode setting (Kelley et al, 2015).

Phylogenetic analysis of TaSODs

The phylogenetic relationship was inferred with the Maximum Likelihood (ML) method based on the LG model in MEGA7.0 (Kumar, Stecher & Tamura; 2016). The midpoint rooted base tree was drawn using Interactive Tree of Life (IToL, version3.2.317, http://itol.embl.de).

Analysis of TaSODs motifs and gene structures

The annotation information of TaSODs was interpreted using GSDS version 2.0 (http://gsds.cbi.pku.edu.cn/index.php) to obtain TaSODs gene structure, intron/exon distribution, and intron/exon boundaries (Hu et al, 2014). Conserved TaSODs gene sequences were identified using the MEME Suite Analysis (version 4.9.1) and MAST Primer Search (http://meme-suite.org/tools) tools (Bailey et al, 2015). Establish(ed) were parameters using known SOD protein sequences, including AtSODs, OsSODs, and ZmSODs, ~~and~~ then ~~apply parameters~~ applied to identify conserved TaSODs: each sequence ~~can~~ comprise (was) ? any number of non-overlapping occurrences of each motif, the number of different motifs ~~is~~ (was set to) 20, and the motif width range~~s~~ from 6 to 50 amino acids. The function of these predictive motifs were analyzed using InterPro (http://www.ebi.ac.uk/interpro) and SMART (http://coot.embl-heidelberg.de/SMART), then ~~use~~ drawn using TBtools software (https://github.com/CJ-Chen/TBtools) ~~for drawing~~. "an image created"

Multiple conditional transcriptome analysis of TaSODs

RNA-seq data original(ly) from multiple conditional transcriptome analys~~i~~e~~s~~ were download(ed) from NCBI and mapped to the wheat reference genome (which reference?) by hisat2. Then ~~genes~~ ? were then assembled by cufflinks to inspect the expression levels of TaSODs (normalized by FPKM, Fragments Per Kilobase of exon model per Million mapped reads). The R package "pheatmap" was used to draw the heatmap of expression profiles for TaSODs.

RESULTS

Identification of SODs from the ~~wheat~~ wheat genome [which reference NCBI or IWGSC?]

In order to identify the wheat SOD proteins (TaSODs), 28 known SOD proteins, including 8 AtSODs, 12 ZmSODs and 8 OsSODs (Kliebenstein, Monde & Last, 1998; Dehury et al. 2013; Krishna et al, 2014), were collected as query sequences to conduct BLASTp BLASTP against wheat reference genome IWGSCv1.0 (E-value < 1e-5). The candidate hits ~~are~~ were further confirmed by Pfam and local BLASTp with a core motif (E-value <1e-5) to further confirm whether the TaSODs contained the superoxide dismutase domain. Finally, our rigorous computer-based screening strategy identified 26 reliable TaSODs (Table 1), including 11, 5, and 10 loci ~~of~~ from the sub-genom~~es~~ es of A, B, and D, respectively. At this point, 54 SODs proteins were obtained from ~~these~~ the four plant genomes (Arabidopsis, rice, maize, and wheat), which ~~was~~ are detailed in ~~in~~ the supplemental information:Table S1. The sequences were renamed in ascending order based on ~~the~~ their phylogenetic relationship of the OsSODs (Liu et al, 2018). The analysis of 26 wheat SOD genes found 17 Cu/Zn-SODs (TaSOD1.1a-TaSOD1.11b), 6 Fe-SODs (TaSOD2.4-TaSOD2.9), and 3Mn-SODs (TaSOD2.1-TaSOD2.3). This is consistent with the protein annotation information. Furthermore, it was found that TaSOD1.1, TaSOD1.5, TaSOD1.6, TaSOD1.7, TaSOD1.8 and TaSOD1.11 ~~have~~ had alternative splicing isoforms.

Table 1 Predicted sequence features of TaSODs

Group	Designation	Gene ID	aLength	bMW	cpI	dIns.	eAli.	fGRAVY	gSub.

	TaSOD1.1a	TraesCS2A01G121200.1	157	15.70149	5.81	17.3	82.55	-0.015	[h]Cyt.
	TaSOD1.1b	TraesCS2A01G121200.2	141	14.1218	6.01	20.93	81.56	-0.003	Cyt.
	TaSOD1.2	TraesCS2A01G399000.1	311	32.3006	5.39	38.55	86.05	-0.001	Cyt.
	TaSOD1.3	TraesCS2B01G417000.1	308	32.15458	5.66	40.23	87.82	0.028	Cyt.
	TaSOD1.4	TraesCS2D01G123300.1	152	15.09177	5.7	17.84	80.79	-0.028	Cyt.
	TaSOD1.5a	TraesCS2D01G396500.1	309	32.16249	5.39	39.71	86.6	0.01	Cyt.
	TaSOD1.5b	TraesCS2D01G396500.2	301	31.3796	5.57	39.85	85.98	0.002	Cyt.
SOD1	TaSOD1.6a	TraesCS4A01G065800.1	164	16.57647	6.58	24.26	83.29	-0.175	Cyt.
	TaSOD1.6b	TraesCS4A01G065800.2	212	22.20282	8.81	26.24	78.21	-0.302	Cyt.
	TaSOD1.7a	TraesCS4B01G243200.1	164	16.68561	6.39	23.82	80.91	-0.171	Cyt.
	TaSOD1.7b	TraesCS4B01G243200.2	174	18.04719	7.23	23.56	76.26	-0.271	Cyt.
	TaSOD1.8a	TraesCS4D01G242800.1	146	15.1378	5.93	24.82	83.49	-0.2	Cyt.
	TaSOD1.8b	TraesCS4D01G242800.2	164	16.6626	6.39	24.91	85.67	-0.112	Cyt.
	TaSOD1.9	TraesCS7A01G292100.1	201	20352.9	5.22	24.45	93.23	0.132	Cyt.
	TaSOD1.10	TraesCS7B01G197300.1	201	20.32292	5.35	22.8	94.18	0.156	Cyt.

	TaSOD1.11a	TraesCS7D01G290700.1	201	20.25075	5.35	23.96	93.23	0.13	Cyt.
	TaSOD1.11b	TraesCS7D01G290700.2	202	20.32183	5.35	23.89	93.27	0.139	Cyt.
	TaSOD2.1	TraesCS2A01G537100.1	231	25.29893	7.89	29.8	91.73	-0.245	[i]Mit.
	TaSOD2.2	TraesCS2B01G567600.1	225	24.60303	7.14	29.35	90.71	-0.278	Mit.
	TaSOD2.3	TraesCS2D01G538300.1	231	25.27483	7.91	31.71	90.48	-0.282	Mit.
	TaSOD2.4	TraesCS4A01G390300.1	261	29.81302	7.23	59.33	82.22	-0.429	[j]Chl.
SOD2	TaSOD2.5	TraesCS4A01G434000.1	390	42.91936	9.41	50.74	71.92	-0.526	Chl.
	TaSOD2.6	TraesCS7A01G048600.1	392	43.40094	9.31	54.79	70.56	-0.544	Chl.
	TaSOD2.7	TraesCS7A01G090400.1	260	29.798	6.84	57.55	82.54	-0.427	Chl.
	TaSOD2.8	TraesCS7D01G043000.1	391	43.32193	9.17	55.37	68.98	-0.547	Chl.
	TaSOD2.9	TraesCS7D01G086400.1	260	29.83994	6.87	58.86	82.88	-0.432	Chl.

Note: [a]Length (Amino acid length); [b]MW (Molecular weight, KD); [c]pl (Isoelectric point); Ins.[d] (Instability index); [e]Ali. (Aliphatic index); [f]GRAVY (Grand average of hydropathy); [g]Sub. (Subcellular localization); [h]Cyt. (Cytoplasm); [i]Mit. (Mitochondria); [j]Chl. (Chloroplast).

Gene structure and chromosomal distribution of wheat genes encoding SOD proteins

In order to study the gene structure of the TaSODs, we analyzed their GFF3 formatted annotation and found that all TaSODs ~~have~~ has introns. The sequence alignment of 26 TaSODs by DNAMAN software revealed that the homology between the 26 proteins was low, and the highly conserved region was mainly concentrated at the C-terminus, which may be the key region for the function of TaSODs (Figure 1). Exon-intron structural diversit~~ies~~y often plays a key role in the evolution of gene families and can provide additional evidence to support phylogenetic grouping (Qu & Zhu, 2006; Liu, White & Macrae, 2010). The exon-intron structure of the TaSOD genes was further examined based on its evolutionary classification. As shown in Figure 2B, all TaSOD genes contained introns in their genomic sequences in wheat, and their intron numbers ranged from 4 to 7. ~~7~~ seven TaSOD members (TaSOD1.9, TaSOD1.10, TaSOD1.11a, TaSOD1.11b, TaSOD2.5, TaSOD2.6, and TaSOD2.8) contained the largest number of introns (7 introns), while the smallest number was found in only one ~~in~~ TaSOD1.5b (4 introns). As expected, the SOD members in the same clade of the phylogenetic tree demonstrated a very similar exon/intron distribution pattern. For example, the TaSOD2.1, TaSOD2.2 and TaSOD2.3 had the same number~~s~~ of exon/intron splicing and similar length.

were of

Information corresponding to TaSODs ~~are~~ were extracted from the GFF3 reference file of the wheat genome to determine the chromosomal location of the TaSOD genes. Based on the extracted physical location (Supplemental information: Table S3), the chromosomal map of TaSOD was constructed using the software MapInspect. The SOD gene map ~~on~~ for the wheat genome is shown to be present only on chromosomes 2, 4, and 7. At the same time, we found that the density of these loci in chromosome 2 ~~is~~ was higher, accounting for 38.46% of all SOD genes (Figure 3).

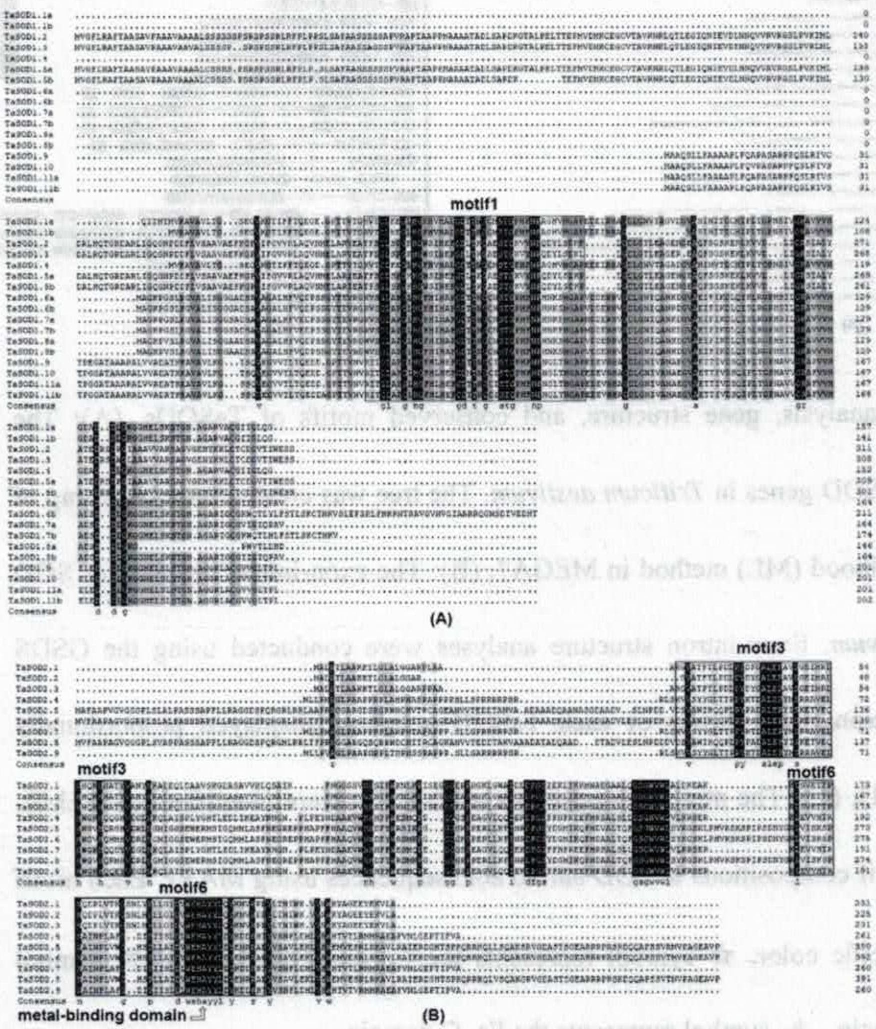

Figure 1 Multiple alignment of TaSOD proteins of functional domain. (A): TaSOD1 (Cu/ZnSODs) subfamily sequence alignment. The motif1 conserved domain is marked in the figure. (B): TaSOD2 (Fe-SODs and Mn-SODs) subfamily sequence alignment. The motifs of motif4 and motif6 are marked in the figure. And metal-binding domain are also labeled.

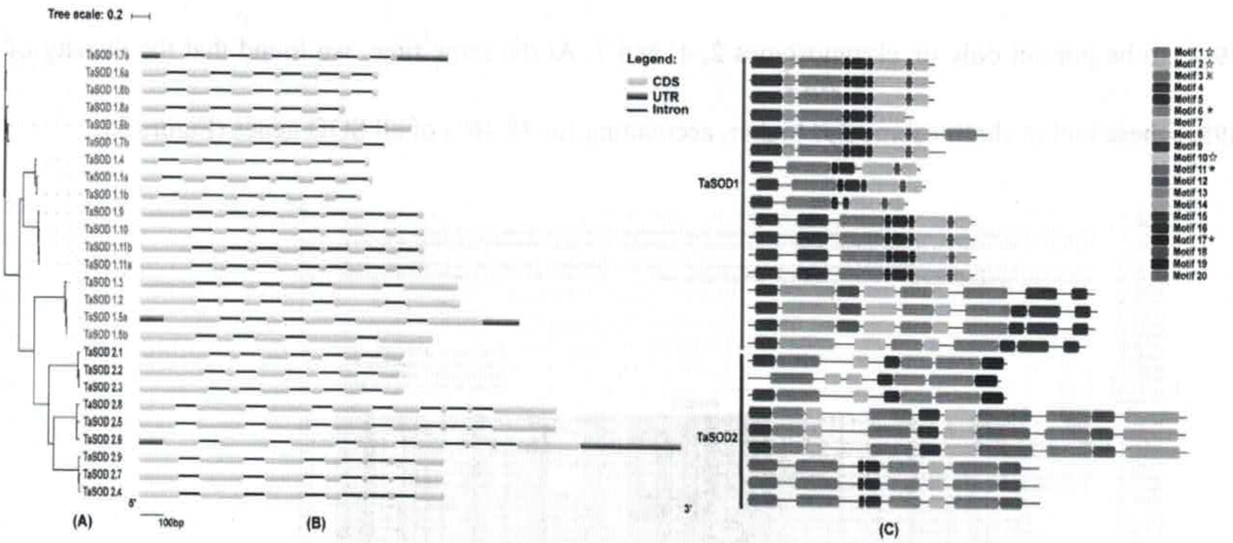

Figure 2 Phylogenetic analysis, gene structure, and conserved motifs of TaSODs. (A): The phylogenetic tree of all SOD genes in *Triticum aestivum*. The tree was created with bootstrap of 1000 by maximum likelihood (ML) method in MEGA7. (B): The exon-intron structure of SOD genes in *Triticum aestivum*. Exon-intron structure analyses were conducted using the GSDS database. Lengths of exons and introns of each TaSOD gene are displayed proportionally (Supplemental Figure S1). (C): The motif compositions of TaSODs were identified by MEME. Model exhibition of motif compositions in SOD amino acid sequences using MAST. Each motif is indicated with a specific color. ☆ symbol represents the Cu/Zn-SOD domain, ※ symbol represents the Fe_N domain, ⋆ symbol represents the Fe_C domain.

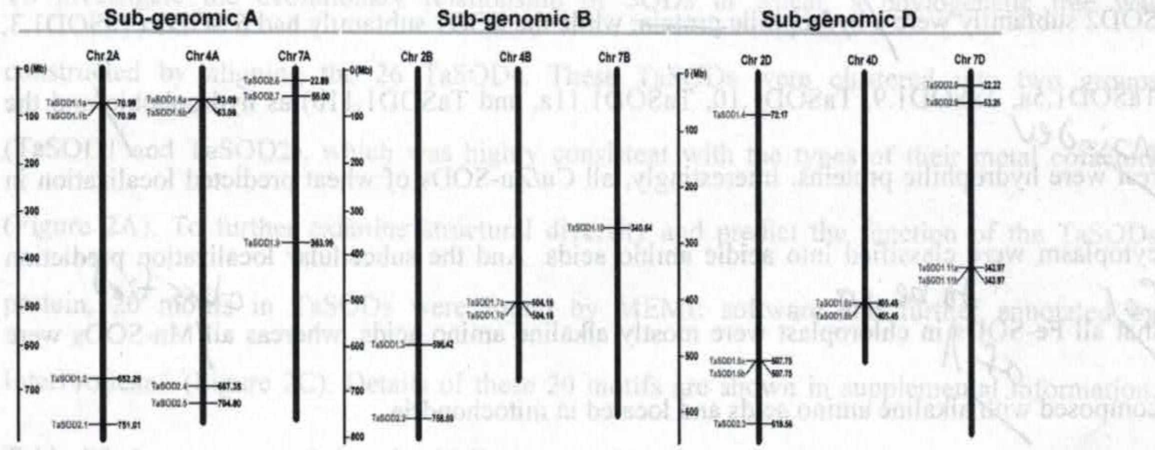

Figure 3 Chromosomal localization of the 26 TaSODs genome. Different classes of TaSODs are represented in different colors. Red represents TaSOD1 and blue represents TaSOD2.

TaSODs protein features

The amino acid sequences of 26 TaSODs proteins were submitted to the ExPASyServer10 (http://www.expasy.org/tools/) online analytical system for analysis of biochemical characteristics such as isoelectric point (pI), relative molecular mass (MW) and instability index (Table 1). The results showed that the TaSODs have an average theoretical pI of 6.69, and the range of pI spans from 5.22 to 9.42. Protein length analysis showed amino acids (aa) ranging from 141 to 392 with an average of 236 aa and an average molecular weight of 25.14396 kD (range from 14.1218 kD to 43.40094 kD). According to previous studies, all Cu/Zn-SODs are acidic in character, while FeSODs and MnSODs are basic or acidic in character (Dehury et al. 2013, Zhang et al. 2016a). In the present study, most of the SOD1 were acidic in character, classified sequences classified sequences except for two SOD1 (SOD1.6b and SOD1.7b). However, most of the SOD2 were basic proteins except for two SOD2 (SOD2.7 and SOD2.9). In addition, the GRAVY analysis showed that the when compared

phylogenetic relationship ~~related to~~ to monocot SODs (wheat, rice and maize) in each clade ~~with~~ all plants.

MARK-UP ENDED HERE

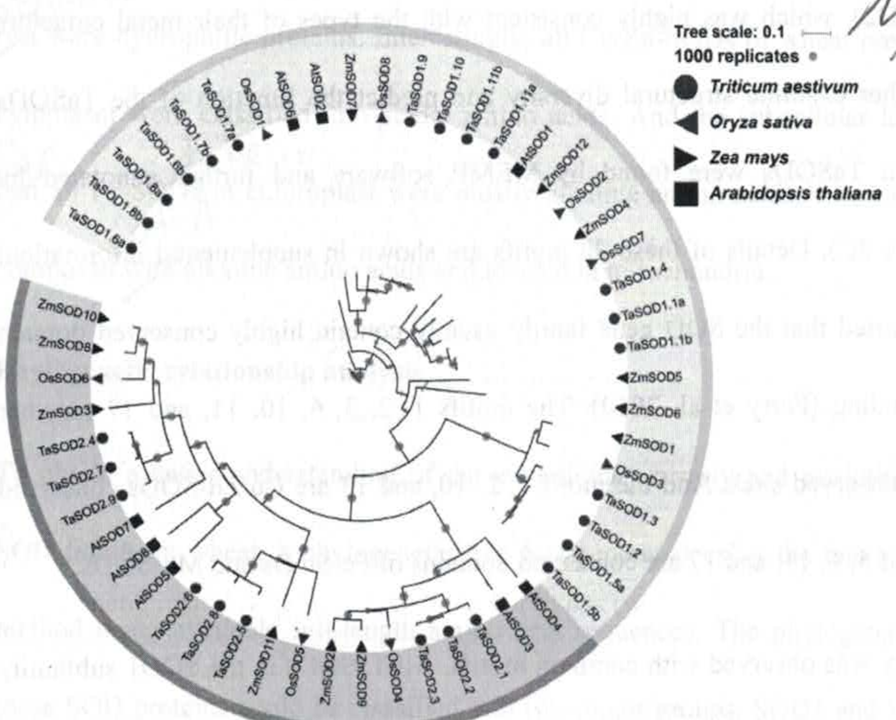

Figure 4 Phylogenetic relationship of TaSODs, OsSODs, AtSODs, and ZmSODs. Protein sequences were aligned using ClustalW2 sequence alignment program and the phylogenetic tree was constructed by software MEGA7 used to create maximum likelihood (ML) under the LG model. The tree was constructed with 1,000 bootstrap replications. Different groups were marked by different colors, and the SOD from wheat, rice, maize and Arabidopsis were distinguished with different color and shape.

Conserved motifs and clustering analysis of TaSODs

described as motif3: [DKE]-[GS]-V-[AS]-[TW]-[FI]-[GK]-[LQ]-[PQT]-[DP]-[LP]-P-Y-[DKP]-[TA]-[DG]-A-L-E-P-[YA]-[IMY]-S-[GKR]-[ER]-[AI]-[VM]-[ER]-[LQ]-H-[WH]-[GQ]-[KV]-H-[QH]-[AQ]-[TG]-[YH]-V and motif 6: [RK]-L-[ES]-[VI]-[ESV]-[HKT]-[TS]-[PR]-N-[AQ]-[ID]-[NP]-[LV]-[AT]-[FKW]-G-[DH]-[IS]-P-[LI]-[LI]-[AG]-[IL]-D-[VL]-W-E-H-A-Y-Y-L-[DQ]-Y-[KE]-[ND]-[DRV]-R-[AP]-[DET]. This result shows that motif3 has eight conserved Valine (V), Proline (P), Tyrosine (Y), Alanine (A), Leucine (L), Glutamic acid (E), Serine (S), Histidine (H). It was further found that the motif 6 included the conserved metal-binding domain "DVWEHAYY" of the Mn-SODs and Fe-SODs. The sequences, locations, and logos of the conserved motifs (motif 1, motif 3, and motif 6) in the TaSOD proteins were shown in Figure 5. The data analyses supported our results. All of the wheat genes we have identified contain conserved domains of the sod family. Congruent with previous studies in other plant species, the wheat SOD gene family contained characteristic amino acids, including a series of highly conserved active site residues that play roles in the sequence-specific binding of mental ions.

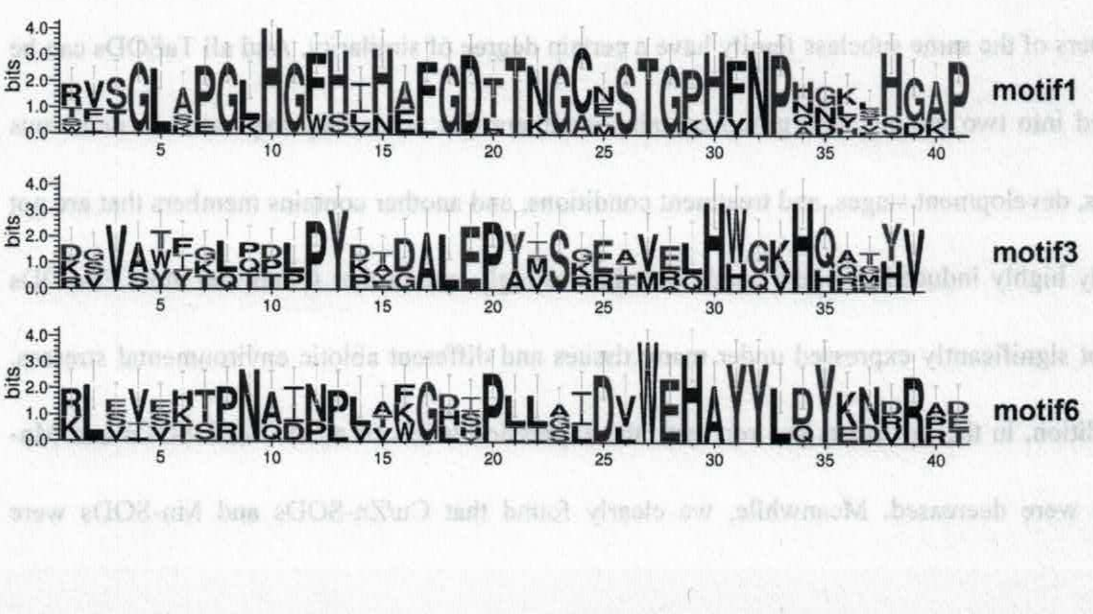

Figure 5 Conserved motifs of TaSODs. The number on x axis indicates the position of amino acid, and the number on Y axis indicates represents the conservation of the protein. The height of a letter indicates its relative frequency at the given position (x-axis) in the motif.

Multiple conditional transcriptome analysis of TaSODs

We performed comprehensive microarray analysis to estimate the expression level of each TaSODs gene in different organs. RNA-seq data (Supplemental information: Table S4) original from multiple conditional transcriptome analysis were download from NCBI and mapped to wheat reference genome by hisat2. Then genes were assembled by cufflinks to inspect the expression levels of TaSODs. R package "pheatmap" was used to draw the heatmap of wheat TaSOD genes. Previous studies have shown that different types of SOD enzyme expression regulation patterns are unique and interact with each other (Dou et al, 2010). It can be observed from the Figure 6 that the SOD gene family members are expressed in different tissues, and the expression patterns of each SOD gene family member are different. The expression patterns of members of the same subclass family have a certain degree of similarity. And all TaSODs can be divided into two groups. One group contains members that are widely expressed in numerous tissues, development stages, and treatment conditions, and another contains members that are not or only highly induced in a few conditions. Interestingly, we further found that most Fe-SODs are not significantly expressed under many tissues and different abiotic environmental stresses. In addition, in the salt stress environment, the expression levels of most Cu/Zn-SODs and Mn-SODs were decreased. Meanwhile, we clearly found that Cu/Zn-SODs and Mn-SODs were significantly up-regulated under drought and high temperature conditions. In particular, TaSOD1.1a and TaSOD1.4 showed the highest expression levels under drought and high temperature stress.

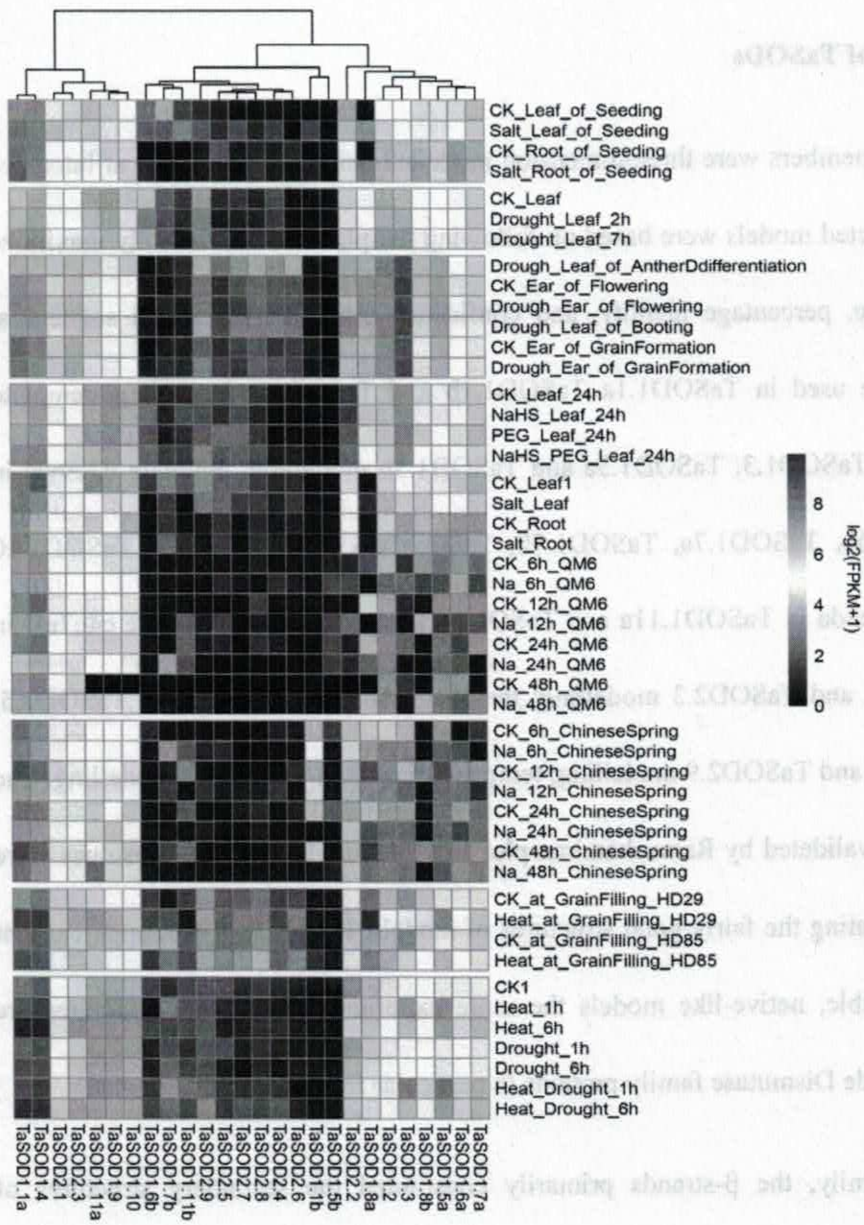

Figure 6 Multi-conditional transcriptome analysis of TaSODs. The depth of the color in the figure reflects the strength of gene expression.

Homology modelling of TaSODs

All 26 wheat TaSOD members were three-dimension modelled using Phyre2 server at intensive mode (Figure 7). Predicted models were based on following templates to heuristically maximize the alignment coverage, percentage identity, and confidence score for the tested sequences. Template c2q2IB were used in TaSOD1.1a TaSOD1.1b and TaSOD1.4 modelling, template c1jkqD in TaSOD1.2, TaSOD1.3, TaSOD1.5a and TaSOD1.5b modelling, template d2c9val in TaSOD1.6a, TaSOD1.6b, TaSOD1.7a, TaSOD1.7b, TaSOD1.8b, TaSOD1.9 and TaSOD1.10 modelling; template d1srda in TaSOD1.11a and TaSOD1.11b models .And template c4c7uB in TaSOD2.1, TaSOD2.2, and TaSOD2.3 modelling; template c6bejA in TaSOD2.4 ,TaSOD2.5, TaSOD2.6, TaSOD2.7, and TaSOD2.9 modelling; template c1xreB in TaSOD2.8 modelling. The quality of models was validated by Ramachandran plot analysis in which 80% of residues were in allowed region indicating the fairly good structures of models. However, it was apparent that to construct more reliable, native-like models the more experimentally solved structures are required from Superoxide Dismutase family proteins in particular from plant SODs.

In the SOD1 subfamily, the β-strands primarily constituted the secondary structures of modelled wheat proteins with 26-41% whereas α-helices distributed with a 3-14%. However, in the SOD2 subfamily, the α-helices primarily constituted the secondary structures of modelled wheat proteins with 45-60% whereas β-strands distributed with a 8-12%. This is similar to the results reported in previous studies (Keerthana & Kolandaivel, 2015). Moreover, to figure out the similarity or divergence of generated models, structures were superimposed to calculate the percentages of structure coverage. The superimposed SOD1 subfamily models were mainly demonstrated a 69-100% structural coverage. And the superimposed SOD2 subfamily models were mainly demonstrated a 51-89% structural coverage. In the SOD1 subfamily, we found that TaSOD1.1b and TaSOD1.4 structural coverage is 100%. However, in the SOD2 subfamily, some models such as TaSOD2.5 (51%), and TaSOD2.6 (51%) showed low structure similarity but above the twilight zone (30%). Taken together, it has been implicated that SODs from each genome donor either may have been ancestrally similar to each other or originally divergent SODs could have been stabilized during long domestication process resulting in changes on protein structures thereby on protein functions.

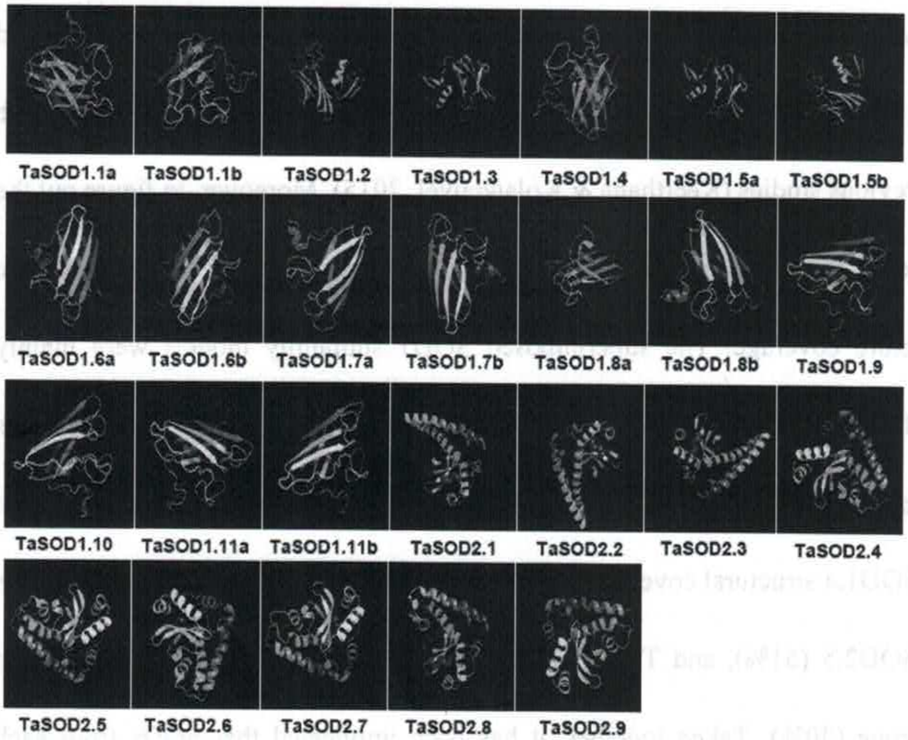

Figure 7 Predicted 3D models of TaSODs proteins. Models were generated by using Phyre2 server at intensive mode. Models were visualized by rainbow color from N to C terminus.

DISCUSSION

Superoxide dismutase (SOD) plays important roles in multiple processes of plant growth and resistance against environment stresses. However, only a tiny fraction of SOD genes have been identified in plants. Genome-wide analysis is an important approach for elucidating the biological roles of the SOD gene family members in given plant species. SOD gene family has been reported to be widely distributed in different plant species, such as Arabidopsis

(Kliebenstein, Monde & Last, 1998), longan (Lin & Lai, 2013), rice (Dehury et al, 2013, Krishna et al, 2014), poplar (Molina et al, 2013), banana (Feng et al, 2015), pear (Wang et al, 2018), tomato (Feng et al, 2015), cotton (Zhang et al, 2016a),and cucumber (Zhou et al, 2017). However, no comprehensive analysis of this gene family has been reported in wheat.

In the present study, a total of 26 SODs genes were identified in wheat, which cover all three major types of plant SOD genes, including 17 Cu/Zn-SODs, 6 Fe-SODs, and 3 Mn-SODs (Table 1). The number of SOD genes varies among plants, previous studies revealed that the numbers of SOD genes in Arabidopsis, rice, sorghum, and tomato are 8 (3 Cu/Zn-SODs, 2 Mn-SODs, and 3 Fe-SODs), 8 (5 Cu/Zn-SODs, 1 Mn-SOD,and 2 Fe-SODs), 8 (5 Cu/Zn-SODs, 1 Mn-SOD, and 2 Fe-SODs), and 9 (4 Cu/Zn-SODs, 1 Mn-SOD, and 4 Fe-SODs), respectively. There are large differences in genome size, and the number of SOD genes varies among these plant species, but does not vary proportionally along with the changes in genome size. The discrepancy in the number of SOD genes in different plant species may be attributed to gene duplication, which consists of tandem and segmental duplications and plays a crucial role in the expansion of SOD genes for diversification. Gene duplication of SOD genes was also found in different plant species (Zhang et al, 2016a; Wang et al, 2016b; Wang et al, 2016b). Therefore, these results imply that TaSOD duplication events play a principal role in gene evolution.

Gene structure analysis revealed that the intron numbers of the 26 wheat SOD genes were 4-7 (Figure 2B). A previous research reported that plant SOD genes have highly conserved intron patterns, and most cytosolic and chloroplastic SODs harbor 7 introns (Fink & Scandalios, 2002).

In our study, seven members (TaSOD1.9, TaSOD1.10, TaSOD1.11a, TaSOD1.11b, TaSOD2.5, TaSOD2.6, and TaSOD2.8) were predicted to contain 7 introns (Figure 2B). The divergence of TaSOD gene structure may be due to the mechanisms including exon/intron gain/loss, exonization/pseudoexonization, and insertion/deletion according to a previous study (Xu et al, 2012), and the SOD members in the same clade of phylogenetic tree displayed similar exon-intron organization patterns (such as TaSOD1.6a and TaSOD1.8b; TaSOD2.1 and TaSOD2.3), suggesting that they may have similar functions related to various abiotic stresses.

Phylogenetic analysis of SOD proteins between wheat and 3 other plant species (Arabidopsis, maize, and rice) showed that the SODs could be divided into two groups of Cu/Zn-SODs and FeSODs/Mn-SODs, which is consistent with the results of previous reports (Wang et al, 2016b; Liu et al, 2018). Most of the subcellular localization data of SODs supported the phylogenetic data. All Cu/Zn-SODs were grouped in the subfamily SOD1 and predicted to be located in the cytoplasm. The chloroplast Fe-SODs and mitochondrial Mn-SODs were clustered into sub-group 2, respectively. In addition, phylogenetic analysis with other species of SOD found that most of the wheat SOD can find homologous sequences in Arabidopsis, maize or rice (Figure 4), suggesting that TaSODs probably have the same functions as SODs in other plant species.

Transcriptome analysis of SOD family genes revealed that various environmental stresses had a regulatory effect on the expression of TaSOD gene. Different TaSOD genes were deferentially expressed under the same environmental stress, and there were also differences in the expression regulation of the same gene under different stresses. This also indicates that different TaSOD proteins may play different mechanisms of action under adverse defense (Bolwell, 1998; Bubliy & Loeschcke, 2005). The transcriptional group treated with various stresses found that wheat SOD had the most obvious response to high temperature and drought stress. Among the 26 TaSODs genes, only 4 TaSOD (TaSOD1.1b, TaSOD1.5b, TaSOD1.6b, and TaSOD2.6) had no obvious variety. It can be seen from the treatment of different genes that the expression levels of TaSOD1.1a and TaSOD1.4 are significantly increased under high temperature and drought treatment conditions. The promoters of these genes can be further analyzed for functional analysis of potential important stress-resistant candidate genes.

Acknowledgements This work was supported by the “National Key R&D Program of China (2018YFD0200500)”, “Open Project Program of State Key Laboratory for Biology of Plant Disease and Insect Pests (SKLOF201707)” and “Open Project Program of Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education (KF201802)”. We thank Prof. Yongli Qiao for beneficial comments on the initial project design and data analysis.

Author Contributions JunliangYin and Dongfang Ma guided the design of the experiment. Wenqiang Jiang, Lei Yang, Yiqing He directed the data analysis. Wenqiang Jiang conducted data analysis, and manuscript writing. Huaigu Chen, Wei Li, and Haotian Zhang contributed to the manuscript writing. Junliang Yin and Dongfang Ma supervised the experiment and confirmed the manuscript. Dongfang Ma is the guarantor of this work, so he can have full access to all the data in the research and is responsible for the integrity of the data and the accuracy of the data

analysis. All authors read and approved the final manuscript. Thank all the above staff for the help of this study. The authors thank the reviewers for their valuable suggestions during the revision of the early manuscripts.

Conflict of interest The authors declare that they have no competing interests.

The supplemental information in this study

Table S1 SOD genes found in Arabidopsis thaliana, *Oryza sativa*, and *Zea mays*.

Table S2 Annotation of putative of TaSODs identified by MEME

Table S3 Location TaSODs genes on Chinese Spring

Table S4 The FPKM data of TaSOD genes in different tissues and environment.

File S1 The gene sequences used in this research.

Figure S1 The exon/intron organization of TaSOD

REFERENCES

Abreu I A, and Cabelli D E. 2010. Superoxide dismutases-areview of the metal-associated mechanistic variations. *Biochim Biophys Acta* **1804(2)**: 263-274 DOI 10.1016/j.bbapap.2009.11.005.