A peer-reviewed version of this preprint was published in PeerJ on 19 November 2019.

View the peer-reviewed version (peerj.com/articles/8062), which is the preferred citable publication unless you specifically need to cite this preprint.

Jiang W, Yang L, He Y, Zhang H, Li W, Chen H, Ma D, Yin J. 2019. Genome-wide identification and transcriptional expression analysis of superoxide dismutase (SOD) family in wheat (*Triticum aestivum*) PeerJ 7:e8062 https://doi.org/10.7717/peerj.8062

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

Wenqiang Jiang , **Lei Yang** , **Yiqing He** , **Haotian Zhang** , **Wei Li** , **Huaigu Chen** , **Dongfang Ma** Corresp., , **Junliang Yin** Corresp.

Corresponding Authors: Dongfang Ma, Junliang Yin
Email address: madf@yangtzeu.edu.cn, yinjunliang@nwafu.edu.cn

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 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 possibly have important function in abiotic stress. Taken together, our results provide a basis for further functional research on SOD gene family in wheat and facilitate their potential applications in the genetic improvement of wheat.

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 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 possibly have important function in abiotic stress. Taken together, our results provide a basis for further functional research on SOD gene family in wheat and 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 accumulates 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 (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 ℃, the Cu/Zn-SOD activity of barley leaves did not change significantly; when the temperature dropped to -3 ℃, the Cu/Zn-SOD activity increased

78 significantly (Moses, 2012). Under drought and saline conditions, the high drought resistance
79 and salt tolerance of the transgenic *AtHDG11* gene increased, while the SOD activity increased,
80 indicating the role of SOD in plant resistance. When the Arabidopsis *CBF1* (C-repeat-binding
81 factor 1) gene were transferred to tobacco plants, the SOD activity of tobacco plants was
82 significantly higher than that of the control, which improved the tolerance of transgenic plants to
83 low temperature (Zhang et al, 2010). Overexpression of Mn-SOD in tobacco and maize
84 chloroplasts enhances the protective effect of transgenic tobacco and maize on the plasma
85 membrane and tolerance to herbicide-induced oxygen stress (Bowler et al, 1991; Breusegem et al,
86 1999). Taken together, these results indicate that enhanced SOD activity in plants can increase
87 plant resistance to a variety of stresses.

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

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-dimension modelled using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/) server at intensive mode (Kelley et al, 2015).

Phylogenetic analysis of TaSODs

The phylogenetic relationship was inferred with the Maximum Likelihood (ML) method based on 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 parameters using known SOD protein sequences, including AtSODs, OsSODs, and ZmSODs, and then apply parameters to identify conserved TaSODs: each sequence can comprise any number of non-overlapping occurrences of each motif, the number of different motifs is 20, and the motif width ranges 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 TBtools software (https://github.com/CJ-Chen/TBtools) for drawing.

Multiple conditional transcriptome analysis of TaSODs

RNA-seq data 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 (normalized by FPKM, Fragments Per Kilobase of exon model per Million mapped reads). R package “pheatmap” was used to draw the heatmap of TaSODs.

RESULTS

Identification of SODs from wheat genome

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 against wheat reference genome IWGSCv1.0 (E-value < 1e-5). The candidate hits are further confirmed by Pfam and local BLASTp with 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, 10 loci of the sub-genomics A, B, and D. At this point, 54 SODs proteins were obtained from these four plant genomes (Arabidopsis, rice, maize, and wheat), which was detailed in in supplemental information:Table S1. The sequences were renamed in ascending order based on the phylogenetic relationship of OsSODs (Liu et al, 2018). The analysis of 26 wheat SOD 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 alternative splicing isoforms.

Table 1 Predicted sequence features of TaSODs

Group	Designation	Gene ID	[a]Length	[b]MW	[c]pI	[d]Ins.	[e]Ali.	[f]GRAVY	[g]Sub.

	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.

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

174 **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 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 diversities 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 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 only one in TaSOD1.5b (4 introns). As expected, the SOD members in the same clade of phylogenetic tree demonstrated a very similar exon/intron distribution pattern. For example, the TaSOD2.1, TaSOD2.2 and TaSOD2.3 had the same numbers of exon/intron and similar length.

Information corresponding to TaSODs are 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 the wheat genome is shown

194 to be present only on chromosomes 2, 4, and 7. At the same time, we found that the density of

195 these loci in chromosome 2 is higher, accounting for 38.46% of all SOD genes (Figure 3).

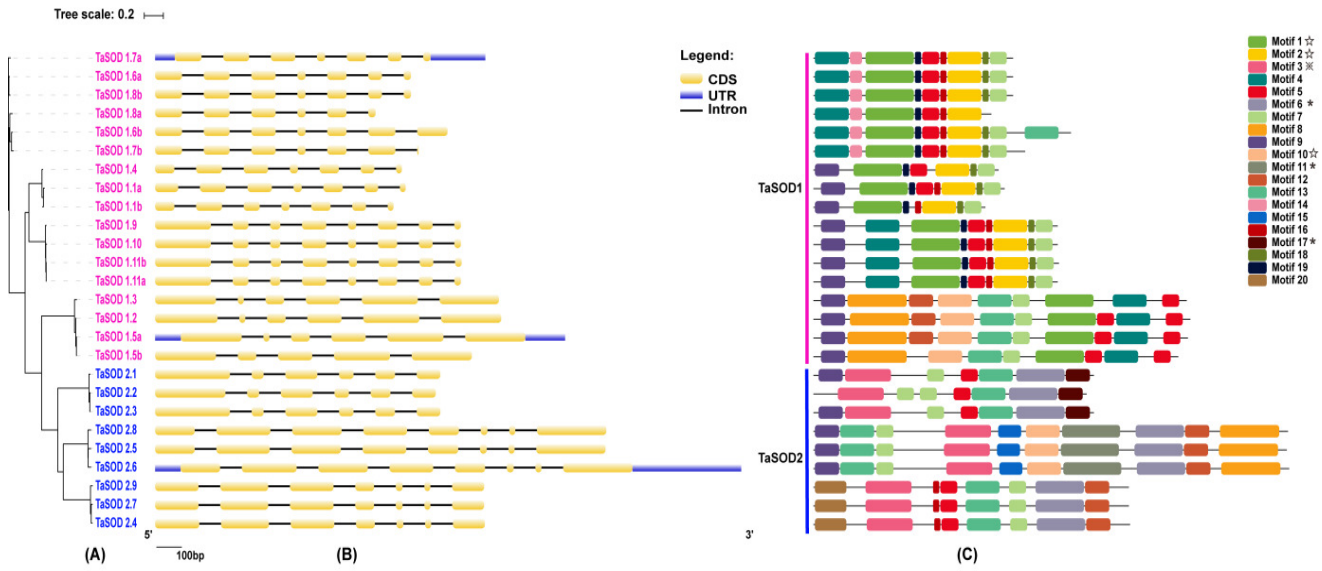
196

197 Figure 1 Multiple alignment of TaSOD proteins of functional domain. (A): TaSOD1

198 (Cu/ZnSODs) subfamily sequence alignment. The motif1 conserved domain is marked in the

199 figure. (B): TaSOD2 (Fe-SODs and Mn-SODs) subfamily sequence alignment. The motifs of

200 motif4 and motif6 are marked in the figure. And metal-binding domain are also labeled.



201

202 Figure 2 Phylogenetic analysis, gene structure, and conserved motifs of TaSODs. (A): The
203 phylogenetic tree of all SOD genes in *Triticum aestivum*. The tree was created with bootstrap of
204 1000 by maximum likelihood (ML) method in MEGA7. (B): The exon-intron structure of SOD
205 genes in *Triticum aestivum*. Exon-intron structure analyses were conducted using the GSDS
206 database. Lengths of exons and introns of each TaSOD gene are displayed proportionally
207 (Supplemental Figure S1). (C): The motif compositions of TaSODs were identified by MEME.
208 Model exhibition of motif compositions in SOD amino acid sequences using MAST. Each motif
209 is indicated with a specific color. ☆ symbol represents the Cu/Zn-SOD domain, ※ symbol
210 represents the Fe_N domain, * symbol represents the Fe_C domain.

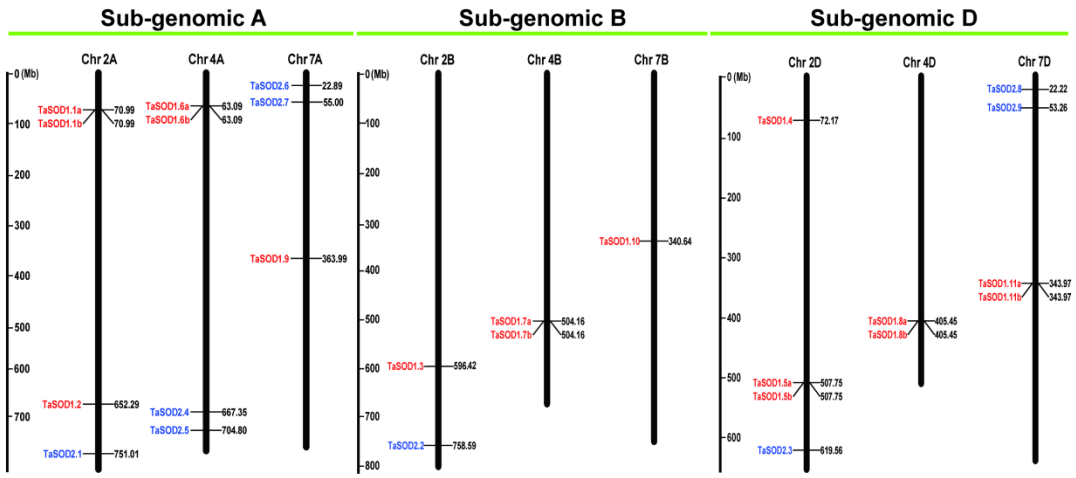


Figure 3 Chromosomal localization of the 26 TaSODs genome. Different classes of TaSODs are represented in different colors. Red represents TaSOD1 and blue represents TaSOD2. **TaSODs protein features**

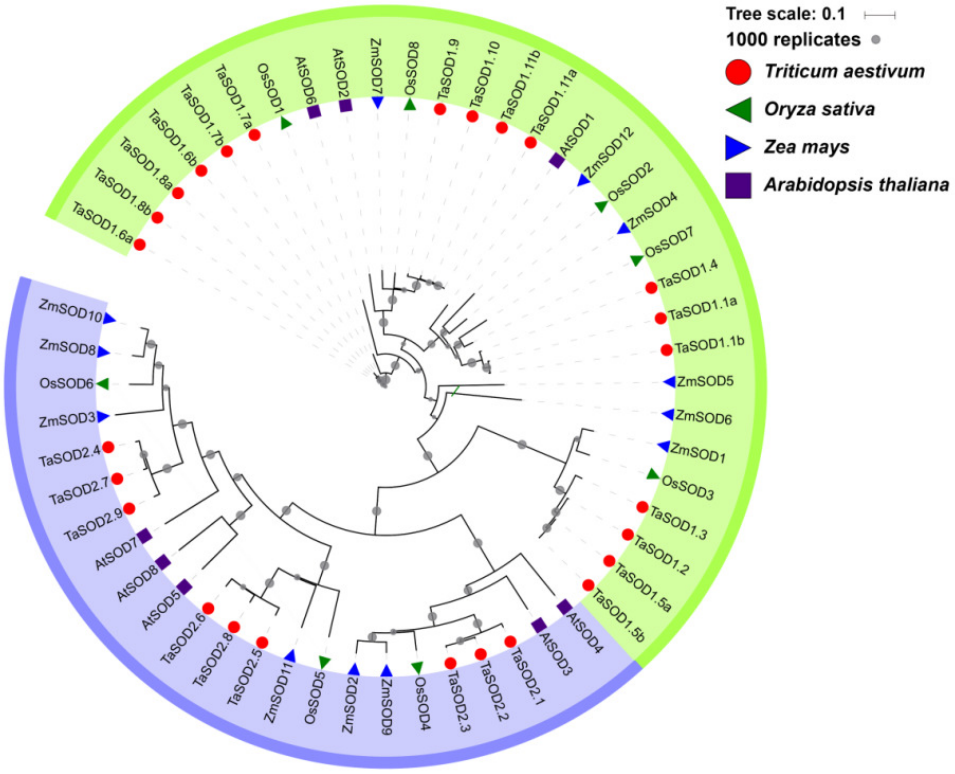
The amino acid sequences of 26 TaSODs proteins were submitted to the ExPASyServer10 (http://www.expasy.org/tools/) online analytical system for analysis of biochemical characteristics such as isoelectric point (pI), relative molecular mass (MW) and instability index (Table 1). The results showed that 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, 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

SOD2 subfamily were a hydrophilic protein, while the SOD1 subfamily had 6 SOD1 (TaSOD1.3, TaSOD1.5a, TaSOD1.9, TaSOD1.10, TaSOD1.11a, and TaSOD1.11b) as hydrophobin and the rest were hydrophilic proteins. Interestingly, all Cu/Zn-SODs of wheat predicted localization in cytoplasm were classified into acidic amino acids. And the subcellular localization prediction that all Fe-SODs in chloroplast were mostly alkaline amino acids, whereas all Mn-SODs were composed with alkaline amino acids and located in mitochondria.

Phylogenetic relationship analysis

To obtain a better understanding of the evolutionary history and evolutionary relationships of SOD family in wheat, a phylogenetic tree was generated using the maximum likelihood (ML) method using available full-length amino acid sequences. The phylogenetic tree revealed that these SOD proteins could be classified into two major groups: SOD1 and SOD2. And we found that the SOD1 subfamily is composed of Cu/Zn-SODs and the SOD2 subfamily is composed of Fe-SODs and Mn-SODs. Based on the phylogenetic tree, we could clearly observe that the SOD proteins within the same subfamily were clustered together, while Fe-SODs and Mn-SODs were divided into one sub-groups (Figure 4), implying that FeSODs and MnSODs originated from a common ancestor (Alscher, Erturk & Heath, 2002). The SOD1 group consists of 17 TaSODs (TaSOD1.1a to TaSOD1.11b), 3 from AtSODs, 6 from ZmSODs, and 5 from OsSODs. Similarly, the SOD2 proteins includes 9 TaSODs (TaSOD2.1 to TaSOD2.9), 5 AtSODs, 6 ZmSODs, and 3 OsSODs. Moreover, we also could find that the dicot SODs (Arabidopsis) have more closely

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



247

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

254 **Conserved motifs and clustering analysis of TaSODs**

To investigate the evolutionary relationship of SODs in wheat, a phylogenetic tree was constructed by aligning the 26 TaSODs. These TaSODs were clustered into two groups (TaSOD1 and TaSOD2), which was highly consistent with the types of their metal cofactors (Figure 2A). To further examine structural diversity and predict the function of the TaSODs protein, 20 motifs in TaSODs were found by MEME software and further annotated by InterProScan5 (Figure 2C). Details of these 20 motifs are shown in supplemental information: Table S2. It was reported that the SOD gene family usually contain highly conserved domain involved in metal binding (Perry et al, 2010). The motifs 1, 2, 3, 6, 10, 11, and 17 together constitute the SOD conserved sites. And the motif 1, 2, 10, and 13 are Cu/Zn-SODs conserved domains, and the motif 3, 6, 11, and 17 are conserved domains of Fe-SODs and Mn-SODs.

The same subfamily was observed with common motifs. All TaSODs in the SOD1 subfamily contain motifs 1 and 5. But motif 5 is not a conserved domain of TaSOD. In the SOD2 subfamily, all members contain motifs 3 and 6. In addition, the Cu/Zn-SODs conserved domains (motif1) were also analyzed to understand the relationship between TaSOD1 and SODs of other species. The alignment of all Cu/Zn-SODs conserved domains of 17 TaSOD1 is described as a conserved motif 1: [RT]-[VF]-[ST]-G-L-[AS]-[PE]-G-[LK]-H-G-[FW]-[HS]-[ILV]-[HN]-[AE]-[FL]-G-D-[TL]-T-[NK]-G-[CA]-[NE]-S-T-G-[PK]-[HV]-[FY]-N-P-[HN]-[GND]-[KL]-[LS]-[HS]-[GD]-[AK]-P. This result shows eight conserved Glycine (G), Leucine (L), Histidine (H), Aspartic acid (D), Serine (S), Threonine (T), Asparagine (N), Proline (P) in the Cu/Zn-SODs conserved domains motif. And the conserved motif of Fe-SODs and Mn-SODs conserved domains site is

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.

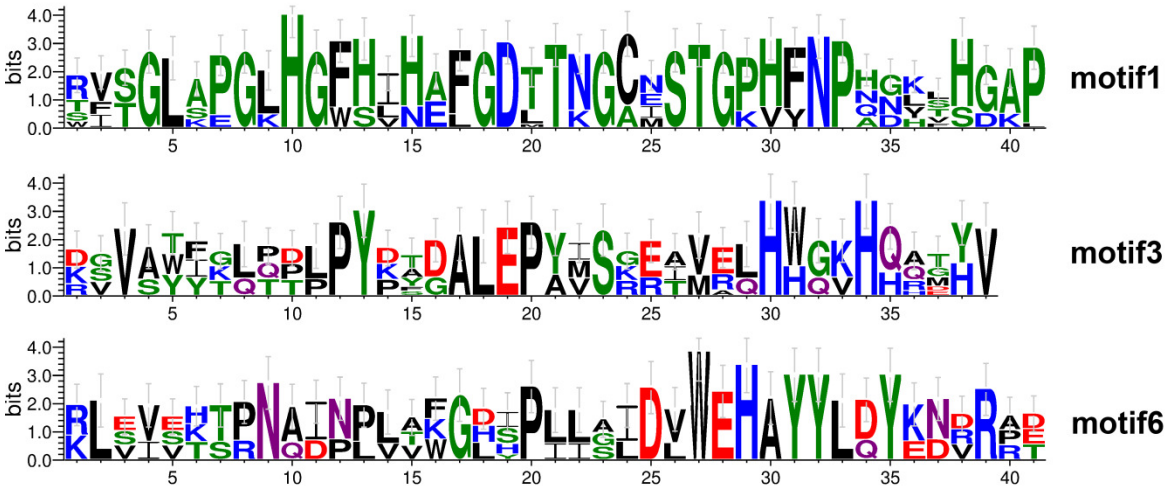
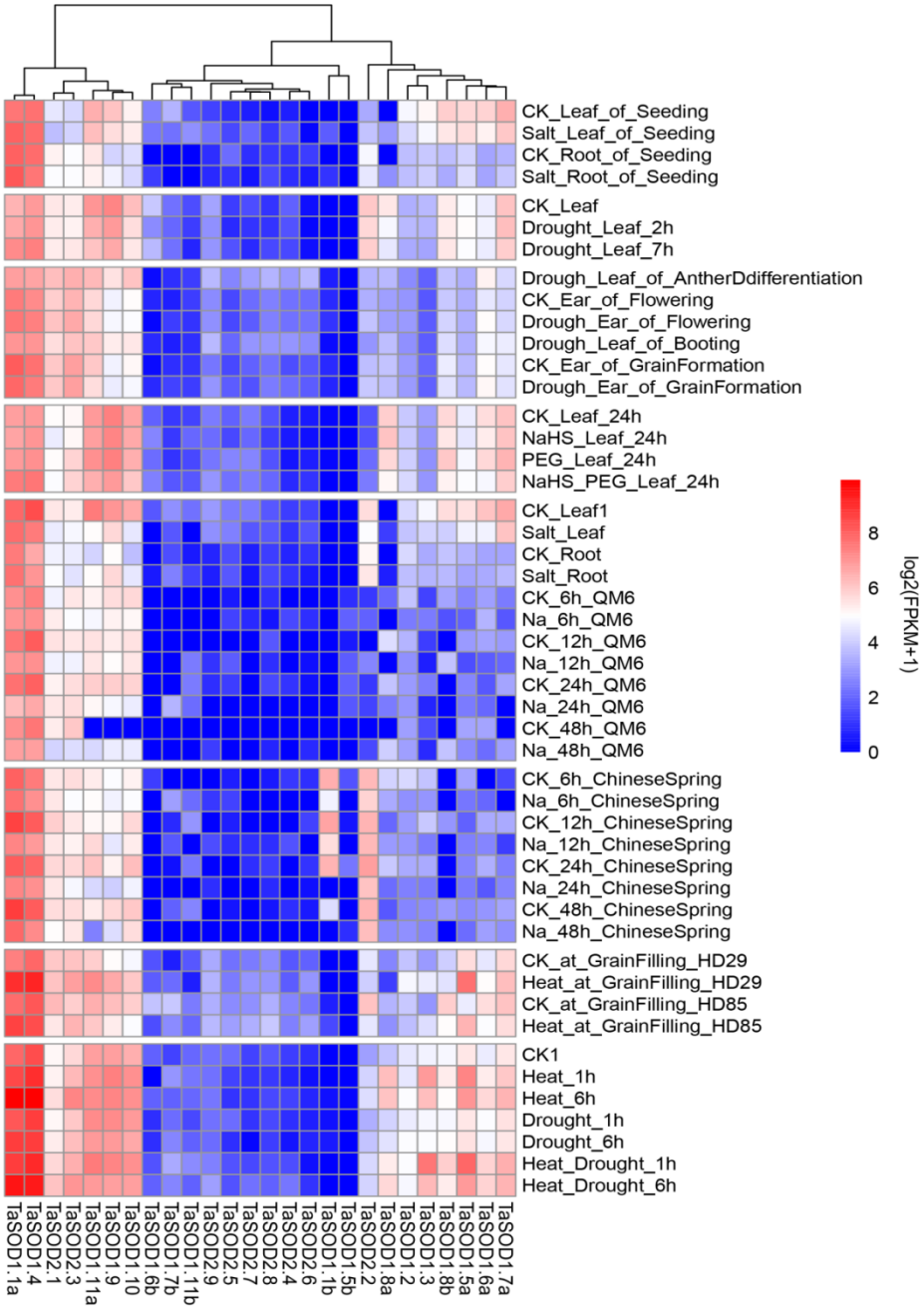


Figure 5 Conserved motifs of TaSODs. The number on x axis indicates the position of amino acid, and the number on Y axis indicates represents the conservation of the protein. The height of a letter indicates its relative frequency at the given position (x-axis) in the motif.

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

309 significantly up-regulated under drought and high temperature conditions. In particular,
310 TaSOD1.1a and TaSOD1.4 showed the highest expression levels under drought and high
311 temperature stress.



312

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.

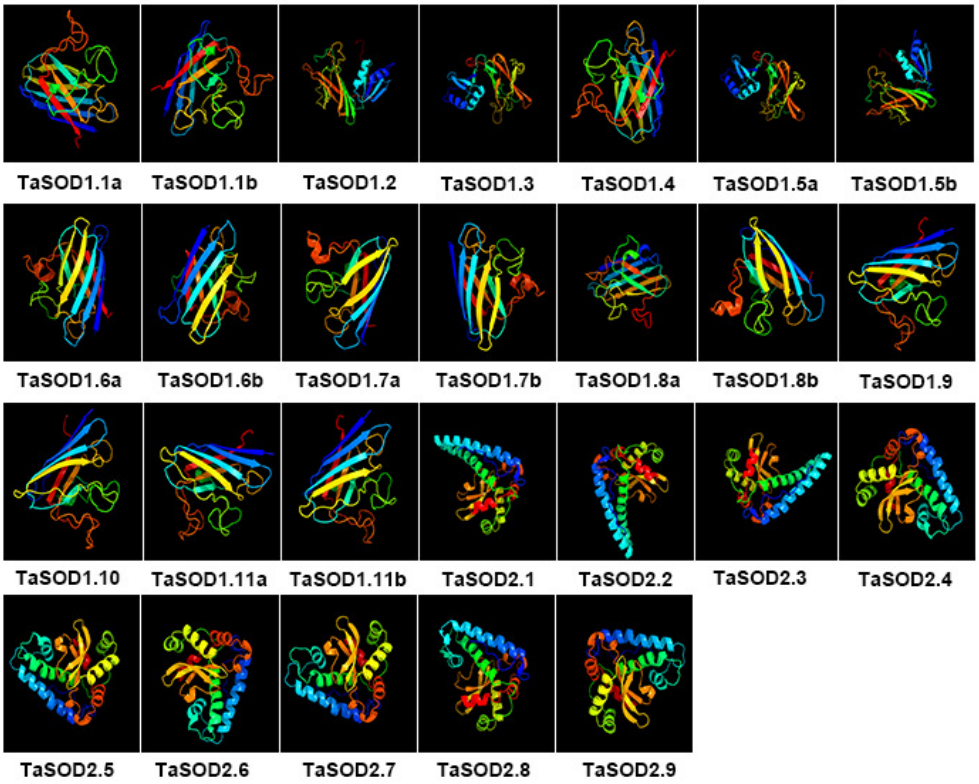


Figure 7 Predicted 3D models of TaSODs proteins. Models were generated by using Phyre2 server at intensive mode. Models were visualized by rainbow color from N to C terminus.

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.

Alscher R G, Erturk N, Heath L S. 2002. Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *Journal of Experimental Botany* **53(372)**:1331-1341 DOI 10.1093/jxb/53.372.1331.

Ahmad P, Umar S, Sharma S. 2010. Mechanism of free radical scavenging and role of phytohormones in plants under abiotic stresses. *Plant Adaptation and Phytoremediation* **2010**:99-118 DOI 10.1007/978-90-481-9370-7_5.

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* **40(Web Server issue)**:W597-W603 DOI 10.1093/nar/gks400.

Bailey T L, Johnson J, Grant C E, Noble W S. 2015. The MEME Suite. *Nucleic Acids Research* **43(Web Server issue)**:W39-W49 DOI 10.1093/nar/gkv416.

Bolwell, G. P. 1998. Comparative biochemistry of the oxidative burst produced by rose and french bean cells reveals two distinct mechanisms. *Plant Physiology* **116(4)**:1379-1385 DOI 10.1104/pp.116.4.1379.

Bowler C, Slooten L, Vandenbranden S, De Rycke R, Botterman J, Sybesma C, Van Montagu M, Inz D. 1991. Manganese superoxide dismutase can reduce cellular damage

450 mediated by oxygen radicals in transgenic plants. *Embo Journal* **10(7)**:1723-1732 DOI
451 10.1002/j.1460-2075.1991.tb07696.x.

452 **Breusegem F V , Slooten L , Stassart J M, Botterman J, Moens T, Montagu M V, Inzé D.**
453 **1999.** Effects of overproduction of tobacco MnSOD in maize chloroplasts on foliar tolerance
454 to cold and oxidative stress. *Journal of Experimental Botany* **50(330)**:71-78 DOI
455 10.1093/jexbot/50.330.71.

456 **Bubliy O A , Loeschcke V. 2005.** Correlated responses to selection for stress resistance and
457 longevity in a laboratory population of drosophila melanogaster. *Journal of Evolutionary*
458 *Biology* **18(4)**:15 DOI 10.1111/j.1420-9101.2005.00928.x.

459 **Chou K C, Shen H B. 2010.** Plant-mPLoc: A Top-Down Strategy to Augment the Power for
460 Predicting Plant Protein Subcellular Localization. *Plos One* **5(6)**:e11335 DOI
461 10.1371/journal.pone.0011335.

462 **Dehury B , Sarma K , Sarmah R ,Jagajjit S, Smita S, Mousumi S, Priyabrata S, Mahendra**
463 **K M, Gauri D S, Manabendra D C,Madhumita B. 2013.** In silico analyses of superoxide
464 dismutases (SODs) of rice *(Oryza sativa* L.). *Journal of Plant Biochemistry and*
465 *Biotechnology* **22(1)**:150-156 DOI 10.1007/s13562-012-0121-6.

466 **Ding F Z. 2008.** Research progress on application of superoxide dismutases in tobacco. *Journal*
467 *of Anhui Agricultural Sciences* **36(5)**:1897-1894 DOI 10.1016/S1872-2075(08)60071-0.

Dong L, He Y Z, Wang Y L,Dong Z Y. 2013. Research progress on application of superoxide dismutase (SOD). *Journal of Agricultural Science and Technology* **15(5)**:53-58 DOI 10.3969/j.issn.1008-0864.2013.05.08.

Dou J H, Yu S X, Fan S L, Pang C Y Song M Z. 2010. SOD and plant stress resistance. *Molecular Plant Breeding* **8(2)**: 359-364. DOI 10.3969/mpb.008.000359.

Du X M , Yin W X , Zhao Y X , Zhang H. 2001. The production and scavenging of reactive oxygen species in plants. *Sheng Wu Gong Cheng Xue Bao* **17(2)**:121-125.

Valko M, Rhodes C J , Moncol J ,Izakovic M, Mazur M. 2006. Free radicals, metals and antioxidants in oxidative stress-induced cancer. Chemico-Biological Interactions, 2006, **160(1)**:1-40.

Dupont C L, Neupane K, Shearer J, Palenik B. 2010. Diversity, function and evolution of genes coding for putative Ni-containing superoxide dismutases. *Environmental Microbiology* **10(7)**:1831-1843 DOI 10.1111/j.1462-2920.2008.01604.x.

Feng K, Yu J, Cheng Y, Ruan M, Wang R, Zhou G, Li Z, Yao Z, Yang Y, Zheng Q, Wan H. 2016. The SOD gene family in tomato: Identification, phylogenetic relationships, and expression patterns. *Frontiers in Plant Science* **7(131)**:1279 DOI 10.3389/fpls.2016.01279.

Feng X, Lai Z, Lin Y, Lai G T, Lian C L. 2015. Genome-wide identification and characterization of the superoxide dismutase gene family in Musa acuminata cv. *Tianbaojiao (AAA group)*. *BMC Genomics* **16(1)**:823 DOI 10.1186/s12864-015-2046-7.

Fink R C, Scandalios J G. 2002. Molecular evolution and structure-function relationships of the superoxide dismutase gene families in angiosperms and their relationship to other eukaryotic and prokaryotic superoxide dismutases. *Archives of Biochemistry and Biophysics* **399(1)**:19-36 DOI 10.1006/abbi.2001.2739.

Foyer C H, Noctor G. 2005. Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. *Plant Cell* **17(7)**:1866-1875 DOI 10.1105/tpc.105.033589.

Guo Y Y, Tian S S, Liu S S, Wang W Q, Sui N. 2017. Energy dissipation and antioxidant enzyme system protect photosystem II of sweet sorghum under drought stress. *Photosynthetica* **2017**:1-12. DOI 10.1007/s11099-017-0741-0.

Hu B, Jin J , Guo A Y, Zhang H, Luo J, Gao G. 2014. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31(8)**:1296 DOI 10.1093/bioinformatics/btu817.

Jiang H F, Ren X P. 2004. The effect on sod activity and protein content in groundnut leaves by drought stress. *Acta Agronomica Sinica* **3(1)**:1-4 DOI 10.1109/CNMT.2009.5374675.

Kelley L A, Mezulis S, Yates C M, Wass M N, Sternberg M J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10(6)**:845-858 DOI 10.1038/nprot.2015.053.

Keerthana S P, Kolandaivel P. 2015. Study of mutation and misfolding of Cu-Zn SOD1 protein. *Journal of Biomolecular Structure and Dynamics* **33(1)**:167-183 DOI 10.1080/07391102.2013.865104.

Krishna N, Susheel K, Roshan S P, Young N Y, Rupak T, Yu S P, Jayamati N, Puneet S C, Bijaya P, Choon H L. 2014. Developmental stagedependent differential gene expression of superoxide dismutase isoenzymes and their localization and physical interaction network in rice (*Oryza sativa* L.). *Genes and Genomics* **36(1)**:45–55 DOI 10.1007/s13258-013-0138-9.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33(7)**:1870 DOI 10.1093/molbev/msw054.

Liu J l, Ouyang L J, Zeng J l, Fu J R, He H H, Zhu C L,SunX T, Zhou D H, Hu L F. 2018. Genome-wide analysis of rice SOD gene family and expression research under stress. *Molecular Plant Breeding* **16(9)**:2753-2760 DOI 10.13271/j.mpb.016.002753.

Liu L, White M J, Macrae T H. 2010. Transcription factors and their genes in higher plants. *Febs Journal* **262(2)**: 247-257 DOI 10.1046/j.1432-1327.1999.00349.x

Lin Q B, Ding Y G, Wang H B, Zhang Z Q, Mao X Y, Le X Y. 2009. The Effects of the SOD Mimics on the Biomass and SOD Activities of Rice Seedlings under Salt Stress[J]. *Chinese Journal of Soil Science* DOI 10.19336/j.cnki.trtb.2009.05.040.

Lin Y L, Lai Z X. 2013. Superoxide dismutase multigene family in longan somatic embryos: a comparison of CuZn-SOD, Fe-SOD, and Mn-SOD gene structure, splicing, phylogeny, and expression. *Molecular Breeding* **32(3)**:595-615 DOI 10.1007/s11032-013-9892-2.

McCord J M, Fridovich I. 1969. Superoxide dismutase an enzymic function for erythrocuprein (hemocuprein). *Journal of Biological Chemistry* **244(22)**:6049-6055 DOI 10.1016/0019-2791(69)90289-4.

Molina R, Juan J, Tsai C J , Kirby E G. 2013. The populus superoxide dismutase gene family and its responses to drought stress in transgenic poplar overexpressing a pine cytosolic glutamine synthetase (GS1a). *PLoS One* **8(2)**:e56421 DOI 10.1371/journal.pone.0056421.

Moses J C. 2012. Cu/Zn superoxide dismutase activity and respective gene expression during cold acclimation and freezing stress in barley cultivars. *Biologia Plantarum* **56(4)**:693-698 DOI 10.1007/s10535-012-0143-x.

Mittler R. 2002. Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science* **7(9)**:405-410 DOI 10.1016/S1360-1385(02)02312-9.

Nielsen H. 2017. Predicting Secretory Proteins with SignalP. *Methodsin Molecular Biology* **1611(2017)**:59-73. DOI 10.1007/978-1-4939-7015-5_6.

538 **Perry J J, Shin D S, Getzoff E D,Tainer J A. 2010.** The structural biochemistry of the
539 superoxide dismutases. *Biochimica Et Biophysica Acta* **1804(2)**:245-262 DOI
540 10.1016/j.bbapap.2009.11.004.

541 **Pitzschke A, Forzani C , Hirt H. 2006.** Reactive Oxygen Species Signaling in Plants.
542 *Antioxidants and Redox Signaling* **8(9-10)**:1757-1764 DOI 10.1089/ars.2006.8.1757.

543 **Quan L J, Zhang B, Shi W W, Li H Y. 2010.** Hydrogen peroxide in plants:A versatile
544 molecule of the reactive oxygen species network. *Journal of Integrative Plant Biology*
545 **50(1)**:2-18 DOI 10.1111/j.1744-7909.2007.00599.x.

546 **Qu L J, Zhu Y X. 2006.** Transcription factor families in Arabidopsis: Major progress and
547 outstanding issues for future research. *Current Opinion in Plant Biology* **9(5)**:544-549 DOI
548 10.1016/j.pbi.2006.07.005.

549 **Razali N, Aziz A A, Lim C Y, Junit S M. 2015.** Investigation into the effects of antioxidant-
550 rich extract of *Tamarindus indica* leaf on antioxidant enzyme activities, oxidative stress and
551 gene expression profiles in HepG2 cells. *Peerj* **3(9)**:1227–1238 DOI 10.7717/peerj.1292.

552 **Sugimoto M, Oono Y, Gusev O, Matsumoto T, Yazawa T, Levinskikh M A, Sychev VN,**
553 **Bingham G E, Wheeler R, Hummerick M. 2014.** Genome-wide expression analysis of
554 reactive oxygen species gene network in mizuna plants grown in long-term spaceflight.
555 *BMC Plant Biology* **14(1)**:4-4 DOI 10.1186/1471-2229-14-4.

556 **Tepperman J M, Dunsmuir P. 1990.** Transformed plants with elevated levels of chloroplastic
557 SOD are not more resistant to superoxide toxicity. *Plant Molecular Biology* **14(4)**:501-511
558 DOI 10.1007/bf00027496.

559 **Wang L L,Wang L, Zhang Z, Ma M, Wang R Z, Qian M, Zhang S L. 2018.** Genome-wide
560 identification and comparative analysis of the superoxide dismutase gene family in pear and
561 their functions during fruit ripening. *Postharvest Biology and Technology* **143(2018)**:68-77
562 DOI 10.1016/j.postharvbio.2018.04.012

563 **Wang W, Xia M X, Chen J , Deng F N, Yuan R, Zhang X P, Shen F F. 2016b.** Data set for
564 phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODs in *Gossypium*
565 *raimondii* and *G. arboreum*. *Data in Brief* **9**:345-348 DOI 10.1016/j.dib.2016.05.025.

566 **Whittaker J W. 2010.** Metal uptake by manganese superoxide dismutase. *Biochimica Et*
567 *Biophysica Acta Proteins and Proteomics* **1804(2)**:298-307 DOI
568 10.1016/j.bbapap.2009.08.014.

569 **Xia M, Wang W, Yuan R, Den F, Shen F F. 2015.** Superoxide dismutase and its research in
570 plant stress-tolerance. *Molecular Plant Breeding* **13(11)**:2633-2646 DOI
571 10.13271/j.mpb.013.002633.

572 **Xu G, Guo C, Shan H, Kong H. 2012.** Divergence of duplicate genes in exon-intron structure.
573 *Proceedings of the National Academy of Sciences of the United States of America*
574 **109(4)**:1187-1192 DOI 10.1073/pnas.1109047109.

575 **Yan Z L, N J Y, Xi L L, Zhou H Y, Jiang J, Liu J H. 2009.** Effect of soil water on protective
576 enzyme activity and membrane lipid peroxidation in pea. *Chinese Journal of Eco-*
577 *Agriculture* **17(3)**:554-559 DOI 10.3724/SP.J.1011.2009.00554.

578 **Yin J L, Fang Z W, Sun C, Zhang P, Lu C, Wang S P, Ma D F, Zhu Y X. 2018.** Rapid
579 identification of a stripe rust resistant gene in a space-induced wheat mutant using specific
580 locus amplified fragment (SLAF) sequencing. *Scientific Reports* **8(1)**:3086 DOI
581 10.1038/s41598-018-21489-5.

582 **Zeng X C, Liu Z G, Shi P H, Xu Y Z. 2014.** Cloning and expression analysis of copper and
583 zinc superoxide dismutase (Cu/Zn-SOD) gene from *Brassica campestris* L. *Acta*
584 *Agronomica Sinica* **40(4)**:636 DOI 10.3724/SP.J.1006.2014.00636.

585 **Zhou Y, Hu L H, Wu H, Jiang L W, Liu S Q. 2017.** Genome-wide identification and
586 transcriptional expression analysis of cucumber superoxide dismutase (SOD) family in
587 response to various abiotic stresses. *International Journal of Genomics* **2017**:1-14 DOI
588 10.1155/2017/7243973.

589 **Zhang H H, Xu N ,Li X, Gu S Y. 2016a.** Overexpression of 2-Cys Prx increased salt tolerance
590 of photosystem Ⅱ(PSⅡ) in tobacco. *PeerJ Preprints* **4**:e2500v1 DOI
591 10.7287/peerj.preprints.2500v1

592 **Zhang H N, Gu J T, Lu W J, Li C D, Xiao K. 2009.** Improvement of low-temperature stress
593 tolerant capacities in transgenic tobacco plants from overexpression of wheat TaSOD1.1.

and TaSOD1.2 genes. *Scientia Agricultura Sinica* **42(1)**:10-16 DOI 10.3864/j.issn.0578-1752.2009.01.002.

Zhang J, Li B, Yang Y, Hu W, Chen F, Xie L X, Fan L Y. 2016b. Genome-wide characterization and expression profiles of the superoxide dismutase gene family in *Gossypium*. *International Journal of Genomics* **31(6)**:1-11 DOI 10.1155/2016/8740901.

Zhang Y, Tan J, Guo Z, Lu S, He S, Shu W, Zhou B. 2010. Increased abscisic acid levels in transgenic tobacco over-expressing 9 cis-epoxycarotenoid dioxygenase influence H_2O_2 and NO production and antioxidant defences. *Plant Cell and Environment* **32(5)**:509-519 DOI 10.1111/j.1365-3040.2009.01945.x.

Zhang Y T, Gao J M, Zhang Q L, Zhang A D, Wang H X, Sun J, Inner M. 2016c. Research progress on plant superoxide dismutase. *Animal Husbandry and Feed Science* **37(9)**:28-31 DOI 10.16003/j.cnki.issn1672-5190.2016.09.009.

Figure 1

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.

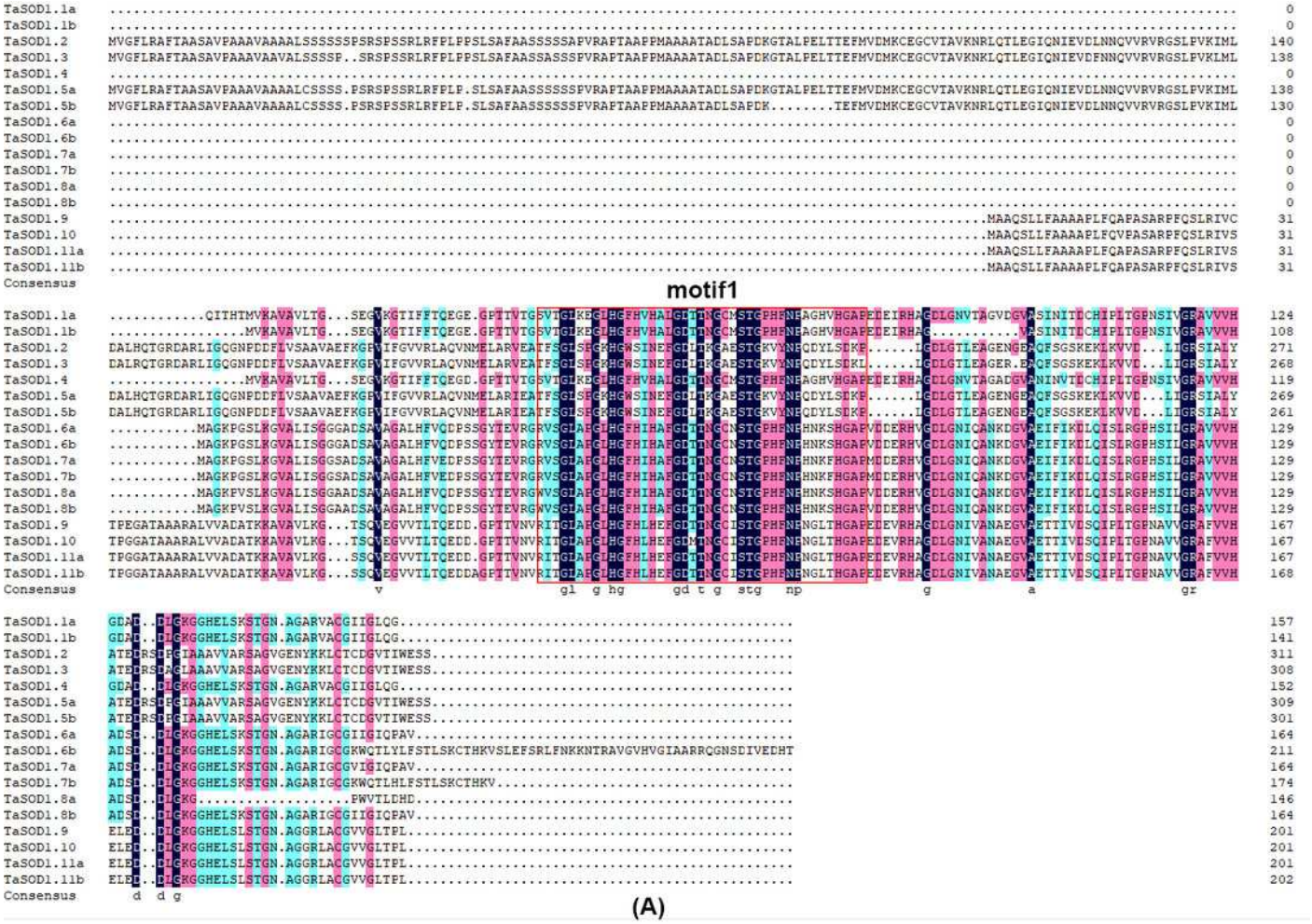
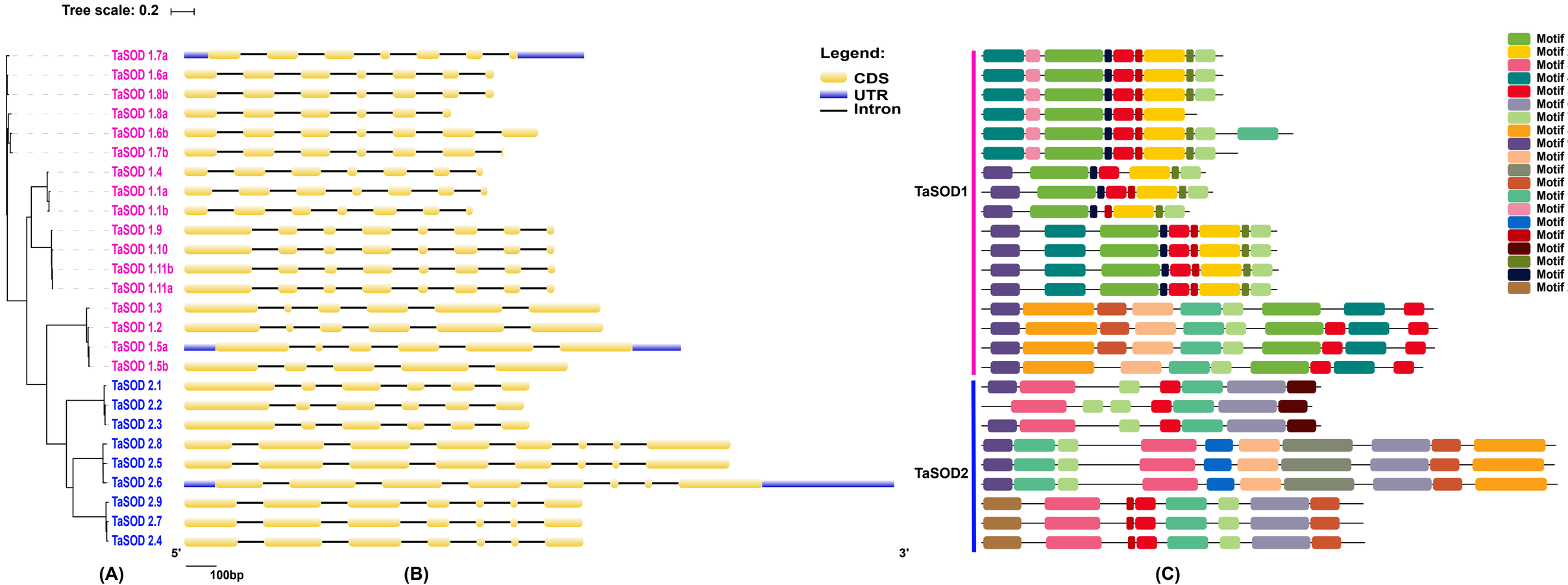


Figure 2(on next page)

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.

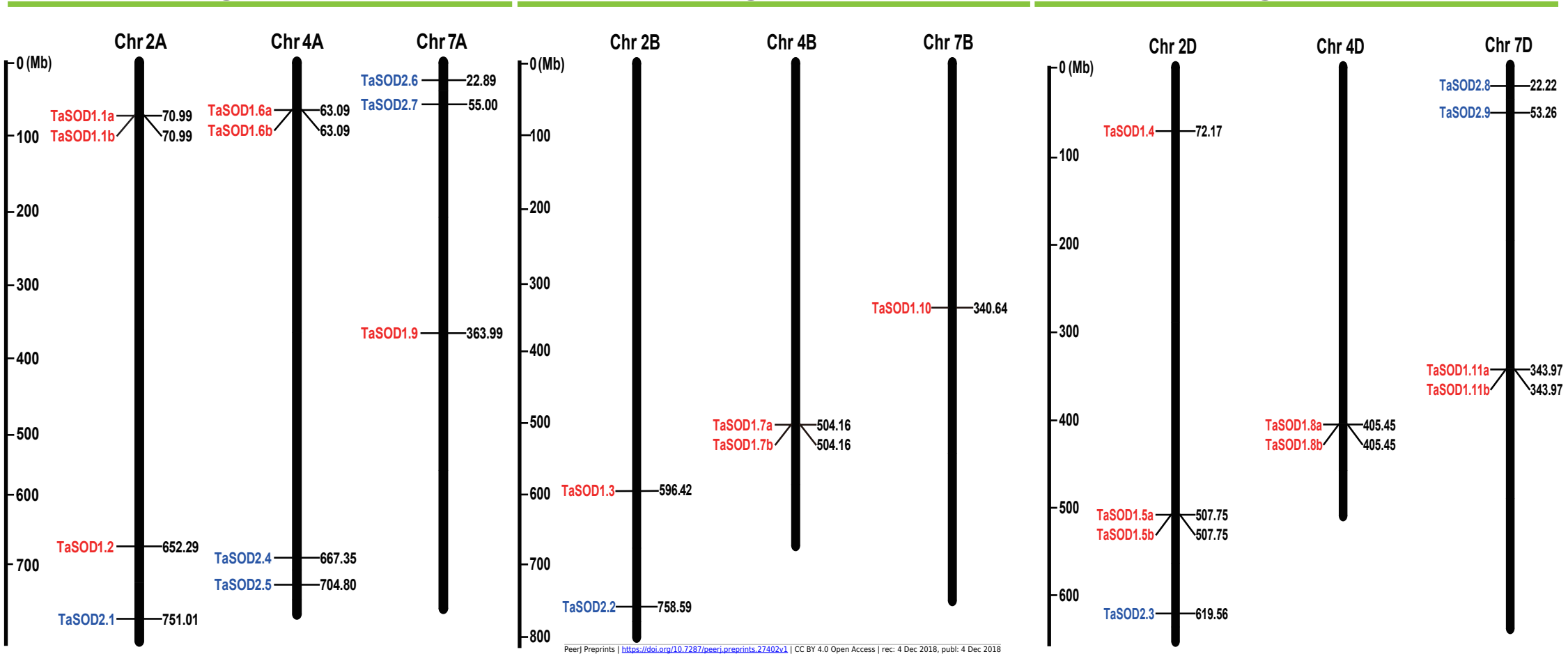


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27402v1 | CC BY 4.0 Open Access | rec: 4 Dec 2018, publ: 4 Dec 2018

Figure 3(on next page)

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.

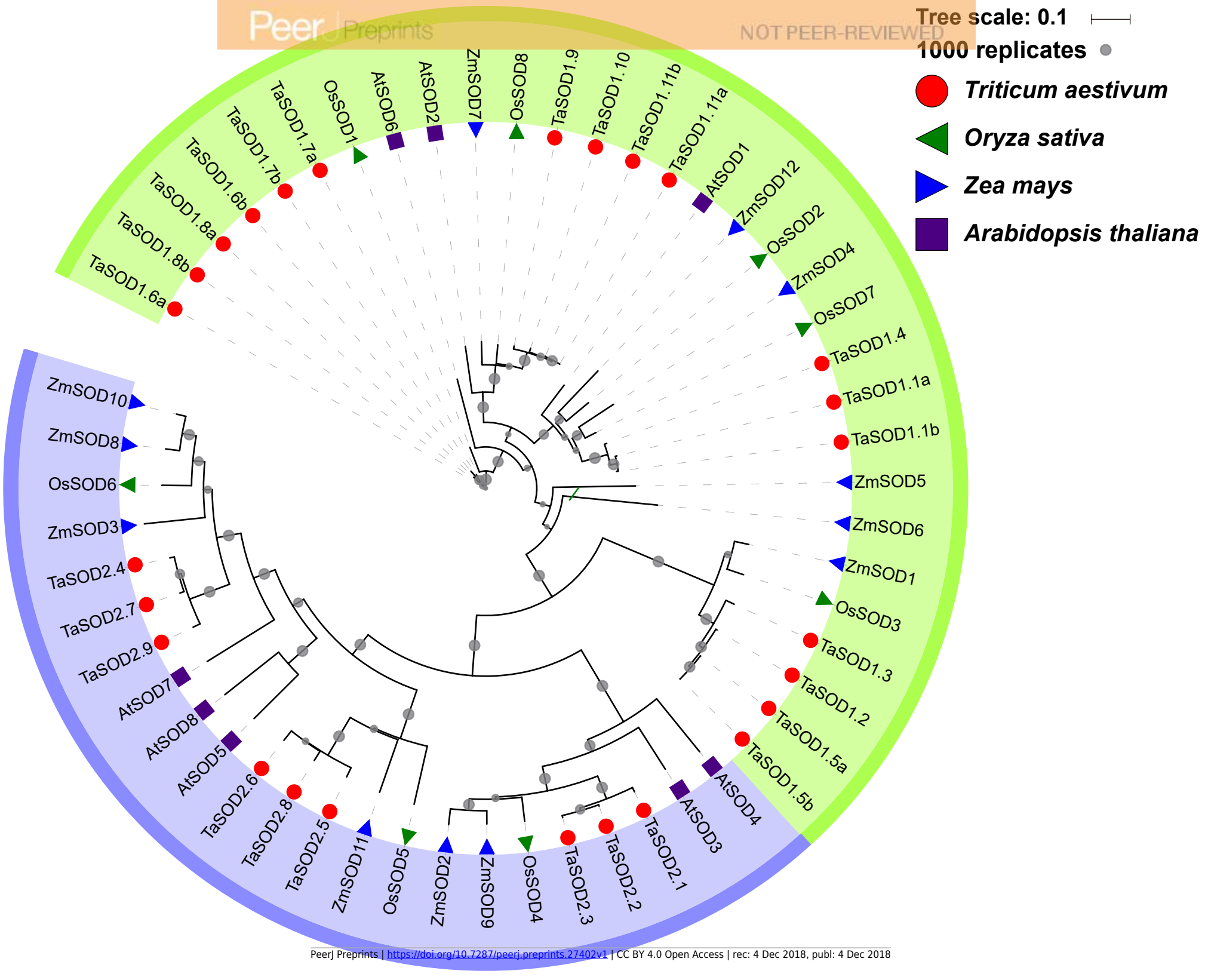


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27402v1 | CC BY 4.0 Open Access | rec: 4 Dec 2018, publ: 4 Dec 2018

Figure 4(on next page)

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.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27402v1 | CC BY 4.0 Open Access | rec: 4 Dec 2018, publ: 4 Dec 2018

Figure 5

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.

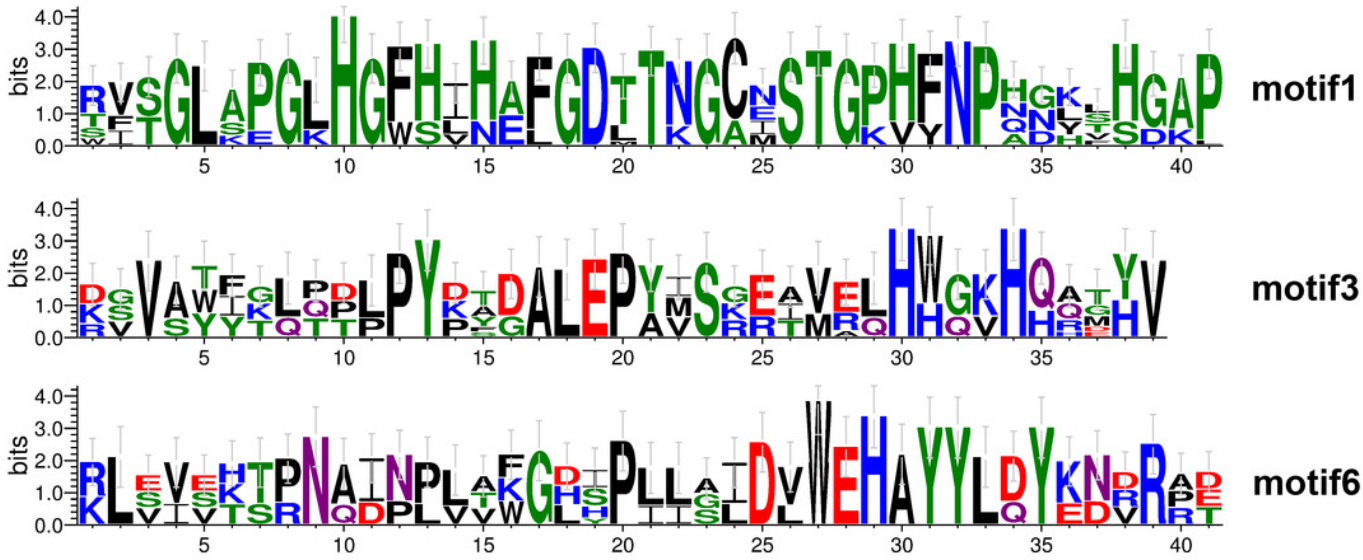
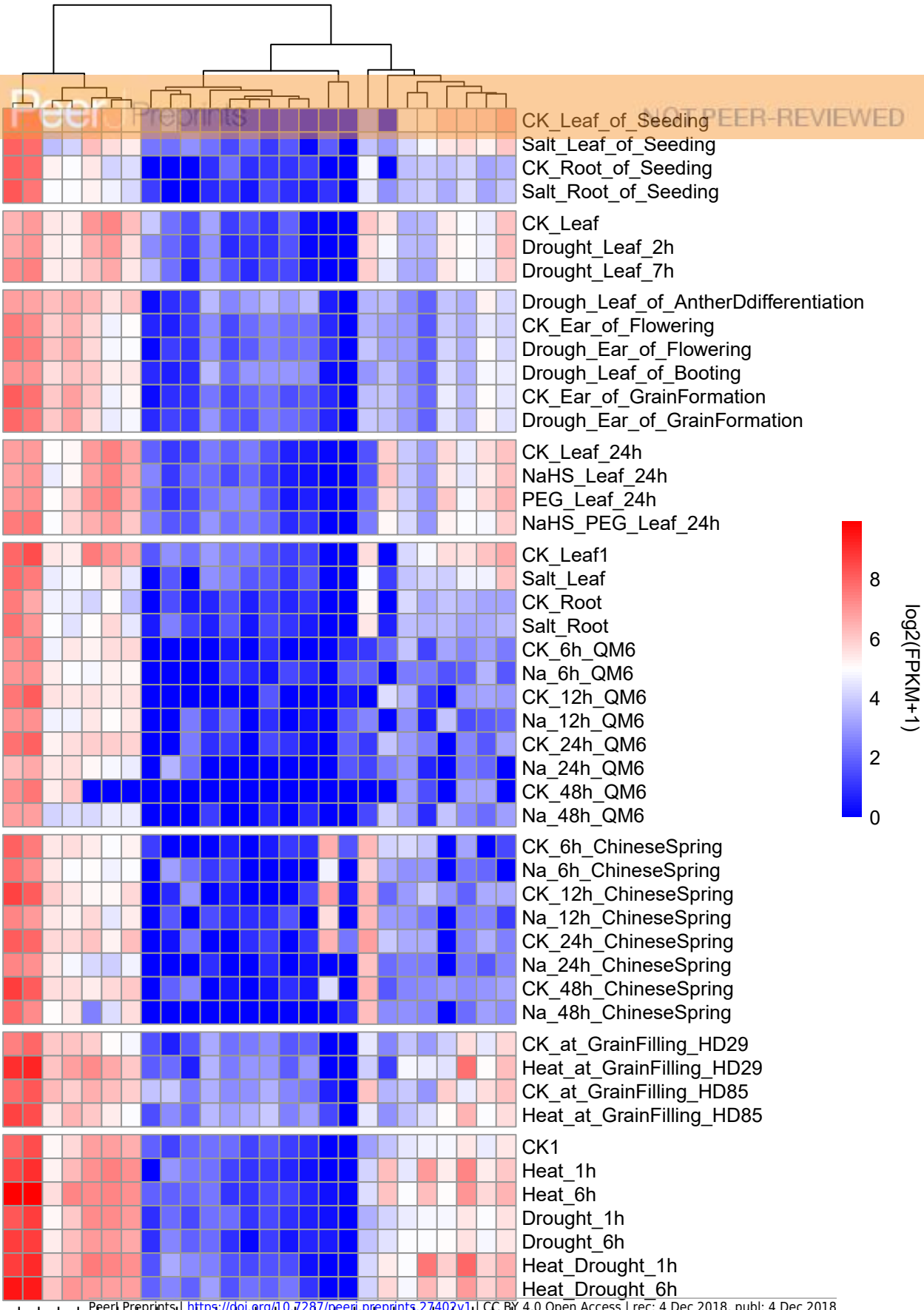


Figure 6(on next page)

Figure 6 Multi-conditional transcriptome analysis of TaSODs

The depth of the color in the figure reflects the strength of gene expression.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27402v1 | CC BY 4.0 Open Access | rec: 4 Dec 2018, publ: 4 Dec 2018

Figure 7 (on next page)

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.