

Supplementary Material

Genome-wide identification and functional analysis of NADPH oxidase

family genes in wheat during development and environmental stress

responses

Chun-Hong Hu^{1, 2†}, Xiao-Yong Wei^{1†}, Bo Yuan¹, Lin-Bo Yao¹, Tian-Tian Ma¹, Peng-Peng Zhang¹, Xiang Wang¹, Peng-Qi Wang¹, Wen-Ting Liu¹, Wen-Qiang Li¹, Lai-Sheng Meng³, Kun-Ming Chen^{1*}

¹, State Key Laboratory of Crop Stress Biology in Arid Areas, College of Life Sciences, Northwest A&F University, Yangling 712100, Shaanxi, P. R. China.

², Department of general biology, College of Life Science and Agronomy, Zhoukou Normal University, Zhoukou 466000, henan, P. R. China.

³, The Key Laboratory of Biotechnology for Medicinal Plant of Jiangsu Province, School of Life Sciences, Jiangsu Normal University, Xuzhou 221116, Jiangsu, P. R. China.

† , These authors contributed equally to this work.

*Correspondence author: Kunming Chen

State Key Laboratory of Crop Stress Biology in Arid Area/College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, 712100, China.

Tel.: +86-29-8708-1178; Fax: +86-29-8709-2262.

Email: kunmingchen@nwsuaf.edu.cn; kunmingchen@163.com

E-mail addresses of authors: Chun-Hong Hu, ourcarrot@163.com; Xiao-Yong Wei, weixiaoyongshengwu@163.com; Bo Yuan, jax911911@gmail.com; Lin-Bo Yao, 675053332@qq.com; Tian-Tian Ma, 9447w@sina.com; Peng-Peng Zhang, 15256849763@163.com; Xiang Wang, wangxiang1342@126.com; Peng-Qi Wang, wangpengqi@nwafu.edu.cn; Wen-Ting Liu, lwt1635@yahoo.com; Wen-Qiang Li, wqli@nwsuaf.edu.cn; Lai-Sheng Meng, menglsh@jsnu.edu.cn; Kun-Ming Chen, kunmingchen@nwsuaf.edu.cn

1. Supplementary Figures and Tables

1.1 Supplementary Figures

Figure S1. Chromosomal locations of wheat TaNOXs. The graph was drawn by Adobe Photoshop CS6 according to the information from scaffolds and Gene ID of each wheat NOX genes and the original sketchs were gotten from the WHEAT URGI website (https://wheat-urgi.versailles.inra.fr/Seq-Repository). TaNOX_S positions are indicated by short horizontal lines and numbers.

Figure S2. The cis-elements responding to abiotic-/biotic-stresses and hormone treatments in the antisense strands of the TaNOX promoters. The cis-elements are as follows. Anaerobic responsive element (ARE): ARE and GC-motif; LTR: low-temperature responsiveness; MYB responsive element (MRE): MBS and MRE; HSE: heat stress responsiveness; TC-rich repeats: defense and stress responsiveness; Light responsive element (LRE): G-box, SPI, MNF1,4cl-CMA2b,GT1-motif, ACE, AAC-motif; 3AF1-binding site and BoxI elements; CAAT-box: common cis-acting element in promoter and enhancer regions; TATA-box: core promoter element around -30 of transcription start; Auxin-responsive element (AuRE): AuXRR-core and TGA-element; TCA-element: salicylic acid (SA) responsiveness; Abscisic acid responsive element (ABRE): ABRE, motif IIb and CE3 elements; Gibberellins (GA)-responsive element (GARE): GARE, TATC-box and p-box; ERE: ethylene-responsive element; Methl jasmonic acid (MeJA) responsive element (JARE): TGACG-motif and CGTCA-motif.

Figure S3. Inducible patterns of TaNOXs under different biotic and abiotic stresses. (A) The expression profiles obtained from the database of TA_AFFY_WHEAT-0 as reported by Genevestigator v3, illustrating different expression levels of TaNOXs under different biotic and abiotic stresses. Results were given as heat

maps in green/red coding that reflect relative signal values; where greener represents stronger down-regulated expression and redder represents stronger up-regulated expression. (B) The inducible expression patterns of wheat NOX family genes in roots under abiotic stresses and hormone treatments. Expression levels of the TaNOXs were assayed by qRT-PCR under cold (4°C), heat (40°C), 20% PEG6000, salt (200 mM NaCl), and oxidative (30 μM MV) stresses and by ABA (100 μM), SA (500 μM), MeJA (100 μM) hormone treatments. Two-week old seedlings were used for the analysis. Data are means \pm SD (n = 3) and are representative of similar results from three independent experiments with three or four replicates for each experiment.

Figure S4. The coexpression patterns between wheat *TaNOX* genes and others at ten different developmental stages. The coexpression relationship between eight wheat *TaNOXs* and other genes were inferred from the database of Ta_mRNASeq_WHEAT_GL-0 with Genevestigator v3. The eight *TaNOX* genes including TaNOX2-1DL, TaNOX3-1DL, TaNOX7-3AS, TaNOX8-4DL, TaNOX12-5BL, TaNOX13-5AL, TaFRO3-2BL and TaFRO4-2BL, were selected for the analysis. The ten developmental stages include geinimation, sheeding growth, tillering, stem elongation, booting, inflorescence emergence, anthesis, milk development, dough debelopment and ripening. \star : Repreasants the target gene among the eight *TaNOXs*.

Figure S5. The coexpression patterns between wheat *TaNOX* **genes and others in twenty two different tissues of wheat.** The coexpression relationship between eight *TaNOXs* and other genes were inferred from the database of Ta_mRNASeq_WHEAT_GL-0 with Genevestigator v3. The eight TaNOX genes are the same to those in figure S4. The twenty two tissues include callus, microspore, coleoptile, spike, rachis, spikelet, stamen, pollen, pistil, caryopsis, endosperm, aleurone layer, ltarchy endosperm, endosperm transfer layer, outer pericarp, shoot, culm(stem), leaf, blade(lamina), flag leaf, roots. \star : Repreasants the target gene among the eight *TaNOXs*.

Genome-wide Analysis of NOX Family Genes in Wheat

Figure S6. The coexpression relationship between wheat *TaNOX* **genes and others under biotic and abiotic stresses.** The response profiles between eight *TaNOXs* and other genes in wheat were inferred from the database of Ta_mRNASeq_WHEAT_GL-0 with Genevestigator v3. The eight TaNOX genes are the same to those in figure S4. \star : Repreasants the target gene of the *TaNOXs*.

1.2 Supplementary Tables

Table S1. The detailed information about the members of wheat *NOXs* **gene family**

The data source of wheat from IWGSC (http://www.wheatgenome.org/), e!EnsemblPlants (http://plants.ensembl.org/index.html), ExPASy

(http://web.expasy.org/protparam/), Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) and TMHMM v.2.0

(http://www.cbs.dtu.dk/services/TMHMM/). A.A.: number of amino acids; MW: molecular weight; pI: theoretical isoelectric point; Chr.: chromosome; Loc: the subcellular localization; TMH: the numbers of transmembrane helix.

Table S2. *NOXs* **and** *FROs* **gene family members in eight plant species**

Subgroups	Homologoue gene pairs		Identical percentage / %	
$\mathbf I$	TaNOX11-5AL	TaNOX9-4DL	98.7	
	TaNOX9-4DL	BdNOX8	89	
	TaNOX10-5BL	AetNOX4	98.9	
	TaNOX3-1DL	AetNOX6	98.7	
	TaNOX4-1DL	OsNOX8	74.3	
	TaNOX15-6BS	TaNOX-like4	97.9	
\mathbf{I}	TaNOX12-5DL	OsNOX7	93.3%	
	TaNOX12-5DL	BdNOX9	95.1%	
	TaNOX12-5DL	ZmNOX11	95.6%	
	TaNOX6-3AL	TaNOX2-1AL	86.6	
	TaNOX6-3AL	ZmNOX8	90.8	
	TaNOX2-1AL	OsNOX2	88.4	
III	TaNOX13-5AL	TaNOX8-4DL	97.5	
	TaNOX13-5AL	ZmNOX12	88.8	
	TaNOX8-4DL	ZmNOX12	87.5	
	TaNOX14-5AS	OsNOX9	86.3	
	TaNOX7-3AS	OsNOX1	87.3	
	TaNOX7-3AS	ZmNOX5	81.2	
IV	TaNOX-like1	AetNOX7	52.1	
	TaNOX-like3	AetNOX7	59.8	
V	TaNOX5-3AL	OsNOX3	90.8	
	TaNOX5-3AL	AtRbohG	64.3	
	TaNOX1-1BL	AtRbohG	65.9	
VI	TaFRO3-2BL	OsFRO1	76.4	

Table S3. The identical percentage of amino acids sequences between homologue gene pairs

Table S4. Rates of non-synonymous (Ka) and synonymous (Ks) nucleotide substitutions and their ratios (ω) for protein-coding nucleotide sequence of paralogous and orthologous gene pairs of NOX family genes in four species*

Gene pairs	Ka	Ks	ω (Ka/Ks)	Divergence time (Mya)			
Triticum aestivum VS. Triticum aestivum (AABBDD, $2n = 6 \times 42$)							
TaNOX4-1DL VS. TaNOX6-3AL	2.1470	1.7086	1.256584338	6.1166			
TaNOX9-4DL VS. TaNOX10-5BL	1.2268	0.5256	2.334094368	3.1663			
TaNOX1-1BL VS. TaNOX12-5DL	2.3250	1.3319	1.745626549	6.2533			
TaNOX8-4DL VS. TaNOX13-5AL	0.0073	0.1414	0.051626591	0.1159			
Triticum aestivum (AABBDD, 2n = 6×= 42) VS. Aegilops tauschii (DD, 2n = 2×= 14)							
TaNOX1-1BL VS. AetNOX3	0.0087	0.0589	0.14770798	0.0606			
TaNOX3-1DL VS. AetNOX6	0.0084	0.0054	1.555555556	0.0231			
TaNOX5-3AL VS. AetNOX2	0.0054	0.0613	0.088091354	0.0551			
TaNOX10-5BL VS. AetNOX4	0.0200	0.0692	0.289017341	0.0955			
TaNOX4-1DL VS. AetNOX10	0.0090	0.0074	1.216216216	0.0259			
TaNOX7-3AS VS. AetNOX1	0.0053	0.0424	0.125	0.0420			
Triticum aestivum (AABBDD, 2n = 6×= 42) VS. Triticum urartu (AA, 2n = 2×= 14)							
TaNOX1-1BL VS. TuNOX3	0.0083	0.0757	0.109643329	0.0710			
TaNOX2-1AL VS. TuNOX6	0.0275	0.0421	0.653206651	0.0927			
TaNOX3-1DL VS. TuNOX4	0.1697	0.3622	0.468525676	0.6466			
TaNOX11-5AL VS. TuNOX5	0.0492	0.0784	0.62755102	0.1683			
TaNOX12-5DL VS. TuNOX8	0.0135	0.0810	0.166666667	0.0893			
TaNOX13-5AL VS. TuNOX7	0.0005	0.0069	0.072463768	0.0060			
TaNOX14-5AS VS. TuNOX1	0.0019	0.0039	0.487179487	0.0071			
TaNOX15-6BS VS. TuNOX2	0.0071	0.0948	0.074894515	0.0859			
Triticum aestivum (AABBDD, $2n = 6 \times = 42$) VS. Hordeum vulgare (AA, $2n = 2 \times = 14$)							
TaNOX1-1BL VS. HvNOX3	0.0388	0.1388	0.279539	0.1853			
TaNOX5-3AL VS. HvNOX4	0.0032	0.0818	0.03912	0.0650			
TaNOX6-3AL VS. HvNOX5	0.0072	0.1728	0.041667	0.1384			
TaNOX7-3AS VS. HvNOX6	0.0196	0.1017	0.192724	0.1163			
TaNOX14-5AS VS. HvNOX2	0.0175	0.0821	0.213155	0.1003			

* The non-synonymous (Ka) and synonymous (Ks) were estimated for the paralogous and orthologous gene pairs of NOX family genes in four species, *Triticum aestivum, Aegilops tauschii, Triticum urartu* and *Hordeum vulgare*, by using the bioinformatics software pamlX1.2 based on the aligned CDS sequences of each orthologous and paralogous gene pairs which the terminator codons were removed. Divergence time (T, million years ago, Mya) for each paralogous gene pair was calculated using the average Ks of λ substitutions per synonymous sites per year as T =Ks/2 λ (λ = 6.5 × 10⁻⁹).

PCC: Pearson correlation coefficient

٠

Table S6. The primers used for qRT-PCR, gene clone, and vector construction in this study.

2. Supplementary information S1 for experimental program

Sequence retrieval and identification of the NOX gene family in wheat

We retrieved the potential sequences of NOX family members including TaNOXs and TaFROs in wheat from IWGSC (http://www.wheatgenome.org/, last accessed May 25, 2017), NCBI (https://www.ncbi.nlm.nih.gov/, last accessed May 25, 2017), and e! Ensembl Plants (http://plants.ensembl.org/index.html, last accessed May 25, 2017), with *Arabidopsis* and rice NOX sequences as queries. Then we identified each NOX member by predicting the conserved domains in EMBL-EBI (http://pfam.xfam.org/search#tabview=tab1) and SMART (http://smart.embl-heidelberg.de/) websites. For further information, we analyzed the physicochemical parameters such as the molecular weight (MW) and isoelectric point (pI) of the wheat NOX and FRO candidates with the ExPASY Compute pI/Mw Program (http://web.expasy.org/compute_pi/), predicted the subcellular localization with Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) and the numbers of transmembrane helix with TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM), and performed amino acid sequence alignment by the software of BioEdit v7.0.5 and DNAMAN 6.0.

Sequence alignment and gene structure analysis

The un-rooted maximum-likelihood phylogenetic tree of wheat NOX and FRO family members was constructed by alignment the complete protein sequences with MEGA 6.06, in which the bootstrap was set to 1,000 replicates. The logos of domain organization were generated by submitting the full-length amino acid sequences to EMBL-EBI or SMART websites, and were amended with Adobe_Photoshop_CS6. The four conserved domain motifs, namely NADPH_Ox, Ferric_reduct, FAD_binding_8 and NAD_binding_6 in each NOX sequence, were generated by submitting the full-length amino acid sequences to MEME Suite (http://meme-suite.org/). These domains were used to perform multiple alignment analysis. The bits score means information content of each position in the amino acid sequence.

Phylogenetic relationships of NOX and FRO gene families in wheat and seven other plant species

Multiple sequence alignments and phylogenetic relationship analysis of NOX and FRO gene families were performed using the maximum-likelihood method with MEGA 6.06 based on the full-length protein sequences from eight plant species. Due to the fact that, the wheat NOX homologues assigned as the same number in gene denomination are similarity in gene structure and protein sizes each other, we selected the biggest ones (including 23 homologues) from those as the representations for phylogenetic analysis. All the amino acid sequences used here, including 23 from *Triticum aestivum*, 11 from *Triticum urartu*, 11 from *Brachypodium distachyon*, 9 from *Hordeum vulgare*, 13 from *Aegilops tauschii*, 16 from Zea mays, 11 from Oryza sativa, and 18 from *Arabidopsis thaliana*, were gotten from the e! Ensembl Plants website (http://plants.ensembl.org/index.html), the Arabidopsis Information Resource website (http://www.arabidopsis.org/), the Rice Genome Annotation Project website (http://rice.plantbiology.msu.edu/index.shtml), the URGI website (https://wheat-urgi.versailles.inra.fr/Seq-Repository) and the NCBI website (https://www.ncbi.nlm.nih.gov/), etc. To further analyze the phylogenetic relationships of NOX family genes, non-synonymous (Ka) and synonymous (Ks) in paralogous and/or orthologous gene pairs from four species, namely *Triticum aestivum*, *Aegilops tauschii*, *Triticum urartu* and *Hordeum vulgare*, were estimated using the bioinformatics software PAMLX 1.2 based on the aligned CDS sequences of each paralogous gene pairs with the terminator codons removed. Divergence time (T, million years ago, Mya) for each orthologous or paralogous gene pairs were estimated using the average Ks of λ substitutions per synonymous sites per year.

Subcellular localization analysis

Analysis of the subcellular location of wheat NOXs were performed in rice protoplasts using transient transformation systems with some modifications (Zhang et al., 2011). The full-length open reading frame (ORF) of *TaNOX7*, *TaNOX10, TaFRO4 and TaNOX-like4* genes were used for the construction of vectors. The merged sequences of TaNOX7::eGFP, TaNOX10::eGFP, TaFRO::eGFP and *TaNOX-like4*::eGFP were under the control of the constitutive 2×35S promoter (cauliflower mosaic virus: CaMV) in pTF486 vector, which were transformed into rice protoplasts. After 16 h incubation with gentle shaking (45 rpm) at 23°C in darkness, the transformed protoplasts were observed with a confocal microscope (A1R, Nikon, Tokyo, Japan). At least three monoclones were sent for sequencing upon gene cloning and vector construction for each examined gene.

Prediction and functional analysis of cis-regulatory elements

The 2,000 bp promoter sequences of TaNOXs (named as TaNOX-pros) selected from the genomic DNA sequences upstream of the transcriptional start sites were used to analyze the cis-acting elements. The cis-acting elements existing in the promoters were predicted by the databases: PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002), PLACE (http://www.dna.affrc.go.jp/PLACE/). Each cis-element logo and position were drawn manually according to the sketch inferred from the Gene Structure Display Server (http://gsds. cbi.pku.edu.cn) by aligning the cis-elements-out DNA sequence and the corresponding full-length DNA sequence of TaNOX-pros with "ATG" and "TAA" added at 5′ and 3′ ends, respectively. The prediction score of every cis-element is greater than or equal to 5. In general, the selected cis-elements were first drawn on the sense strands instead of the antisense strands unless there were not the cis-elements on the sense strands but on the antisense strands.

Then, seven promoter sequences were cloned by PCR using the gene-specific primers, with the isolated genomic DNA from wheat leaves as the template. The promoter sequences was then cloned into the corresponding site of the dual luciferase reporter expression vector (Basic vector) instead of the 2×35S promoter in front of the Rluc gene. The Basic vector contains both expression cassettes of firefly luciferase (Fluc) and Renilla luciferase (Rluc), and the more information about it could refer to Fig. 7A and a previous study (Gu et al., 2013). Three or more monoclones were sent for sequencing to identify each promoter sequence and related vector construction (GenScript Co., Ltd, Nanjing, China). The primers used for cloning the promoters and constructing the dual luciferase reporter expression vectors were listed in Supplementary Table S6.

The preparation of rice protoplasts and the PEG-mediated transformation were performed as described previously (Zhang et al., 2011). Following transformation, the protoplasts were resuspended in 1 mL cell culture solution I per well and plated in 12-well plates (Nest Biotech Co., Ltd, Jiangsu, China) for incubation at 45 rpm at 23°C for 16 h in the dark. In order to determine the responses of the TaNOX-pros to temperature stresses and hormone treatments, the transformed protoplasts were incubated at 23°C for 2 h and then were placed at 42°C or 4°C or treated with exogenous abscisic acid (ABA) or methyl jasmonic acid (MeJA) for 14 h, respectively, while the control groups were still placed at 23°C for 14 h until harvested for luciferase activity detection.

Luciferase activity detection was performed with the Dual-Luciferase ® Reporter Assay System (DLTAS) (www.promega.com) with some modification (Gu et al., 2013). The cultures from one well were put into a 1.5 ml of Eppendorf tube and then harvested by centrifugation at 12,000 rpm for 5 min. After being re-suspended in 30 μl potassium phosphate lysis buffer (PLB, provided by DLTAS kit), the protoplasts were freeze-thawed twice at -80℃ and then centrifuged at 12,000 rpm at 4℃ for 10 min. The resulted supernatants were then collected for luciferase activity detection. LSD's t-tests were performed using SPSS 16.0 software. Statistical differences between means (Rluc activity normalized against Fluc activity) from the test and control groups were considered significant at $P \le 0.05$ based on one-way ANOVA. All values were shown as the average of the data collected from three independent experiments, each having three replicates.

Plant materials, treatments, expression profiles and coexpression network analysis

The expression patterns of wheat NOX family genes at different developmental stages and in different tissues were carried out with the online Genevestigator v3 (https://genevestigator.com/gv/). Results were given as heat maps by submitting the probe ID of the genes into the Ta_mRNASeq_WHEAT_GL-0 database in Genevestigator v3. The inducible expression patterns of the genes were obtained in the same way, while the data from both the Ta_mRNASeq_WHEAT_GL-0 database and the TA_AFFY_WHEAT-0 database in Genevestigator v3.

To construct the coexpression patterns of wheat NOX genes with the others, the gene expression data from ten different developmental stages and/or twenty two different tissues of wheat plants were considered and the probe IDs of eight NOX genes including TaNOX2-1DL, TaNOX3-1DL, TaNOX7-3AS, TaNOX8-4DL, TaNOX12-5BL, TaNOX13-5AL, TaFRO3-2BL and TaFRO4-2BL were submitted into the Ta_mRNASeq_WHEAT_GL-0 database to search for the potential coexpression genes using the Similarity Search-Co-Expression tool in Genevestigator v3. The coexpression patterns were then constructed and presented as heat maps and also table lists.

For the analysis of the tissue-specific expression profiles of wheat NOX genes, seven tissues of wheat (*Triticum aestivum* cv. Chinese Spring) including young leaf, young root, leaf, stem, leaf sheath, young panicle and panicle were harvested to isolate the total RNA for further qRT-PCR. For analysis of the inducible expression profiles of the NOX genes, the 14 d old hydroponic seedlings that were treated with 4℃, 40℃, 200 mM NaCl, 20% PEG6000, 30 μM MV, 100 μM MeJA and 100 μM ABA respectively, for 0, 12 and 24 h, were used as the materials for RNA extraction. For the analysis of the coexpression levels of *TaNOXs* and other genes in wheat, the 7 d old seedlings treated with 42℃ for 24 h, 200 mM NaCl or 10 μM MeJA or 10 μM ABA for 6 h, soil drought for 2 d, and nature drying for 30 min respectively, were used as the materials for RNA extraction. Total RNA was extracted from the different samples using RNAiso TM Plus (Takara, Dalian, China) and treated with RNase-free DNase I (Takara, Dalian, China). The concentration and quality of the RNAs were tested by the NanoDrop 1000 Spectrophotometers (Thermo, USA) and the first-strand cDNA synthesized by using EasyScript First-Strand cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China). The gene expression levels were detected with the UltraSYBR Mixture (Kangwei, Beijing, China) using a real-time PCR detection system CFX96 (Bio-Rad, Hercules, California, America) with *TaActin* (AB181991.1) and *TaGAPDH* (ABS59297.1) as the internal transcript level controls. All the expression level represent the mean \pm SD of data collected three independent experiments, each having three or four replicates. The detailed primer sequences are shown in Supplementary Table S6.

Reference

- 1. Zhang, Y., Su, J., Duan, S., Ao, Y., Dai, J., Liu, J. et al. (2011). A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. *Plant methods*, 7, 30. doi: 10.1186/1746-4811-7-30
- 2. Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y. et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research*, 30, 325-327.
- 3. Gu, L., Han, Z., Zhang, L., Downie, B. and Zhao, T. (2013). Functional analysis of the 5' regulatory region of the maize GALACTINOL SYNTHASE2 gene. *Plant science: an international journal of experimental plant biology*, 213, 38-45. doi: 10.1016/j.plantsci.2013.09.002