

SEQUENCE COMPARISON OF *MUTE* PROMOTERS BETWEEN TWO WHEAT CULTIVARS REVEALED BOTH CONSERVATIVE AND DIVERGED REGIONS

Olha Lakhneko^{*1,2}, Anton Stepanenko^{1,3}, Yuliia Rymar¹, Mykola Borysyuk³, Bogdan Morgun^{1,4}

Address(es):

¹ Institute of Cell Biology and Genetic Engineering, National Academy of Sciences of Ukraine, Akademika Zabolotnoho st. 148, 03143 Kyiv, Ukraine.

² Institute of Plant Genetics and Biotechnology, Plant Science Biodiversity Center, Slovak Academy of Sciences, Akademicka 2, P.O. Box 39A, 950 07 Nitra, Slovakia.

³ School of Life Sciences, Huaiyin Normal University, Changjiang West Road 111, Huai'an 223300, China.

⁴ Institute of Plant Physiology and Genetics, National Academy of Sciences of Ukraine, Vasylykivska st. 31/17, 03022 Kyiv, Ukraine.

*Corresponding author: olha.lakhneko@savba.sk

<https://doi.org/10.55251/jmbfs.9662>

ARTICLE INFO

Received 26. 11. 2022
Revised 26. 2. 2024
Accepted 29. 2. 2024
Published 1. 6. 2024

Regular article



ABSTRACT

Modulation of stomatal biogenesis is one of the means to consider for improving crops' water use efficiency without losing productivity. Transcription factor MUTE is among the critical regulators of stomata biogenesis, particularly of the guard mother cell division. In the current study, we investigated promoter regions of three homoeologous *MUTE* genes—*MUTE-A1*, *MUTE-B1*, and *MUTE-D1* from two bread wheat cultivars of different geographic origins. Based on the available sequence of the Chinese Spring genome, *MUTE* promoters were isolated and sequenced from the Ukrainian cultivar Nataalka. Promoter regions of more than 1600 bp upstream of the predicted start codons were cloned as PCR-amplified fragments and analyzed by sequencing. The sequence alignments revealed significant subgenome and cross-cultivar heterogeneity of the analyzed promoters. The analysis showed numerous single nucleotide polymorphisms and large deletions in every Nataalka sequence compared with Chinese Spring: 276-bp deletion in *MUTE-A1*, 205-bp in *MUTE-B1*, 1291-bp in *MUTE-D1*. These changes affected a range of *cis*-regulatory elements (light- and stress-responsive elements, as well as tissue-specific elements) within the investigated DNA sequences. The study sheds new light on the regulatory variation of one of the critical factors in stomata biogenesis, widening our understanding of wheat genome plasticity with potential implications for improving the crop's performance under restricted water conditions.

Keywords: *cis*-elements, *in silico* analysis, MUTE, promoter, stomata biogenesis, transcription factor, wheat

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is considered one of the most important cereal crops, with about 200 million hectares allocated for sowing worldwide. However, global climate change, which leads to droughts, increased average annual temperatures, and decreased rainfall, challenges crop productivity (Ortiz *et al.*, 2008). The efficiency of water use by plants is linked with leaf transpiration, which accounts for up to 90% of total plant water loss and is controlled mainly through the number of stomata and stomatal aperture (Duursma *et al.*, 2019). Reducing transpiration moisture loss is considered a promising means of improving water use efficiency and conserving soil water levels (Hepworth *et al.*, 2015; Bi *et al.*, 2018).

Biotechnological modulation of stomata density on the leaf surface is the perspective approach for bread wheat improvement concerning drought without impact on yield (Hughes *et al.*, 2017; Serna, 2020). The density of stomata depends on the conversion of leaf meristemoid cells into stomatal structures. This process is controlled by differential expression of stomatal biogenesis genes, such as *EPF1*, *EPF2*, *MUTE*, *SPCH*, and others (Takata *et al.*, 2013; Caine *et al.*, 2016; Lau *et al.*, 2018; Guo *et al.*, 2021). The promising object for wheat improvement is the gene encoding the MUTE transcription factor (Le *et al.*, 2014; Qi *et al.*, 2017). Guo *et al.* showed that the expression of *MUTE* gene initiates in meristemoid cells and ensures the formation of mother guard cells and the transition to guard cells (Guo *et al.*, 2021). Furthermore, the MUTE transcription factor directly controls key genes of stomatal biogenesis and cell cycle genes (Han *et al.*, 2018; Guo *et al.*, 2021). MUTE also plays a specific role in cereals compared to other plants (Raissig *et al.*, 2016, 2017).

As an allohexaploid species, *T. aestivum* has a double set of three subgenomes (A, B, and D). Each subgenome comprises seven pairs of homologous chromosomes, and the entire wheat genome consists of 42 chromosomes (Šramková *et al.*, 2021). On the one hand, sequencing of subgenomes and their further detailed study are significantly complicated because of the complexity of the genome structure and the high level of ploidy. On the other hand, the existence of groups of homologous genes from different subgenomes can make the regulation of gene expression more flexible and manageable.

Gene expression is governed by a multi-level regulation system with gene promoters playing a pivotal role at the stage of DNA transcription into messenger

RNA. DNA elements around the transcription initiation start point are bound with specific transcription factors to control selective gene transcription (Das & Bansal, 2019; Andersson & Sandelin, 2020). The typical eukaryotic promoter consists of proximal and distal regions. The proximal promoter region (from -250 to +250) contains core and proximal promoter elements (Porto *et al.*, 2014). The core promoter in plants, as well as in other eukaryotic organisms, has basic *cis*-elements, which are the binding sites for subunits of RNA polymerase II and other proteins associated with the transcription complex (Andersson & Sandelin, 2020). Regulatory sequences of core promoter usually include TATA-box and initiator element (Inr). However, a minimal core promoter may lack a TATA-box, Inr, or other regulatory downstream promoter elements (Schmitz *et al.*, 2022). In the TATA-less core promoter, Inr can compensate for the absence of this sequence and perform its functions (Andersson & Sandelin, 2020). Among other *cis*-regulatory sequences in the TATA-less proximal promoter, the CAAT-box is common. This regulatory sequence is sensitive to mutations, directly affecting the efficiency of transcription (Porto *et al.*, 2014; Ricci *et al.*, 2019).

A distal promoter region contains numerous *cis*-regulatory elements, including enhancers and silencers, which regulate gene functions in concert with transcription factors (Vo Ngoc *et al.*, 2019). Transcription factors interact with *cis*-elements, which leads to reprogramming transcription patterns (Zou *et al.*, 2011). The observed suppression of single nucleotide polymorphisms (SNPs) frequency within promoters' *cis*-elements sequences highlights the evolutionary and functional importance of preserving these regulatory elements' integrity (Korkuć *et al.*, 2014).

The aim of the current work was to perform a comparative characterization of promoter regions of homoeologous genes encoding transcription factor MUTE, regulating stomata development, in three subgenomes of the referent wheat cultivar Chinese Spring (CS) and an Ukrainian cultivar Nataalka in order to pave the way for future modulation of stomata biogenesis.

MATERIAL AND METHODS

Plant material

The cultivar (cv) Nataalka, which originated at the Institute of Plant Physiology and Genetics of the National Academy of Sciences of Ukraine, was selected to

investigate *MUTE* promoters' structure. Natalka is a cold-tolerant and drought-susceptible cultivar with a high average yield compared to the local standards. It is recommended to be grown in forest, forest-steppe and steppe zones of Ukraine. Its grain has fine bread-making quality (Morgun et al., 2014).

Isolation of promoter, sequence amplification, and sequencing

Total DNA was extracted and purified from one grain according to the standard rapid high-yield protocol based on CTAB detergent (Murray & Thompson, 1980). Specific primers for amplification of *MUTE* promoter sequences (sequences upstream of the predicted start codons) were designed manually based on reference

sequences from cv Chinese Spring (IWGSC RefSeq v2.1, https://wheat-urgi.versailles.inra.fr) (Alaux et al., 2018). The sequences of primers and their positions are specified in Table 1. Conditions of PCR were the following: denaturation–94°C, 30 sec, hybridization–60°C, 30 sec, elongation–68°C, 1 min. In summary, polymerase chain reaction (PCR) of 20 µl included 0.5 µM of forward and reverse primers each (Metabion, Germany), GCI Reaction Buffer (Takara Biomedical Technology, China) containing 2 mM MgCl₂, 0.2 µM of each deoxyribonucleoside triphosphate (Takara Biomedical Technology, China), 1 unit of LA Taq DNA polymerase (Takara Biomedical Technology, China), and 30 ng of total plant DNA.

Table 1 The characteristics of the designed primers used for promoters isolation

Promoter of gene	Chromosome	Coordinates on chromosomes according to IWGSC RefSeq v2.1	Size of region, bp	Primers for isolation 5'→3'
<i>MUTE-A1</i>	2A	75458022-75461680	3659	(i) MUTE-A1prF1 ATTATGCCCTATCGCAAC and MUTE-A1prR1 GGTGTTTGAAATTGCGTG, (ii) MUTE-A1prF2 TATTTTCGCAACTGCAGGC and MUTE-A1prR2 TACAGTGAACACAGTTGT, (iii) MUTE-A1prF3 AAGGCCTCAACGGTCACA and MUTEprR CAGCGCTTGATGTAGAAGG
<i>MUTE-B1</i>	2B	115572709-115574355	1647	MUTE-B1prF CAAGTGAATATAAGCAACG and MUTEprR
<i>MUTE-D1</i>	2D	74106782-74108166	2585	MUTE-D1prF GCCACGACTGAACTATACAAA and MUTEprR

Following PCR, the amplified fragments were separated with agarose gel electrophoreses in 1×TAE running buffer (40 mM Tris-acetate, 1 mM EDTA), purified from agarose with AxyPrep DNA Gel Extraction Kit (Axygen Scientific, USA) and ligated into the vector pMD-19 (Takara Biomedical Technology, China) according to the manufacturer instruction. The ligation products were transformed into *Escherichia coli* dH5α cells following standard calcium chloride protocol (Sambrook & Russell, 2001). Three clones for each PCR reaction were custom-sequenced (Sangon Biotech, Shanghai, China) using a combination of standard forward and reverse sequencing primers and a range of internal primers.

In silico analysis of promoter regions

The software CLC Main Workbench 6.9.2 (Qiagen, USA) was used for multiple global alignments and pairwise comparisons of reference and experimental sequences. Pairwise sequence comparison presented values of distance and percent identity. The distance was given as the proportion between identical and overlapping alignment positions between the two sequences using the Jukes-Cantor correction. Percent identity represented the percentage of identical residues in alignment positions to overlapping alignment positions between two sequences.

The promoter sequences were processed with the online tool New PLACE version 30.0 to identify *cis*-regulatory elements among 469 database entries (Higo et al., 1999).

RESULTS AND DISCUSSION

Intra- and inter-genomic sequence polymorphism of *MUTE* promoter in wheat cultivars Chinese Spring and Natalka

Our preliminary study on the expression of *MUTE* genes (*MUTE-A1*, *MUTE-B1*, and *MUTE-D1*) revealed diversified patterns among Ukrainian cultivars (data not published). This observation inspired us to investigate the variability of the genes' promoter regions. The reference genome of the cultivar CS (IWGSC RefSeq v2.1) was taken as a template for subgenome polymorphism evaluation. As a result, we observed substantial subgenome diversity in the promoter regions upstream of the transcription start compared to the protein-coding downstream sequences. The primers positioning aimed to isolate large promoter regions for comparing the similarity and variability of the sequences in wheat subgenomes.



Figure 1 Promoter sequence comparison of *MUTE* genes from subgenomes A (*MUTE-A1*), B (*MUTE-B1*), and D (*MUTE-D1*) of cultivars Chinese Spring and Natalka. **A.** Homoeologous sequence alignment of the reference cultivar Chinese Spring (IWGSC RefSeq v2.1). Green arrows mark the position of primers used for PCR amplification and sequencing of promoter regions. **B.** Homoeologous sequence alignment of cultivar Natalka. Conservation graphs indicate variable-conserved (from dark red to yellow) sections of these regions. **C.** Pairwise comparison of homoeologues promoter region from CS and Natalka includes distance (above the diagonal) and percent identity (below the diagonal).

In consequence, primer pairs were designed to isolate sequences upstream of the predicted start codons based on CS genome: (i) 3659-bp of *MUTE-A1*, (ii) 1647-bp of *MUTE-B1*, and (iii) 2585-bp of *MUTE-D1*. The extracted from the database and aligned promoter regions of CS *MUTE-A1*, *MUTE-B1*, and *MUTE-D1* genes, along with the positions of primers used for the isolation, are illustrated in Figure 1A. The alignment displayed that the region of approximately 300 bp upstream of the start codon is relatively conserved among the three sequences (yellow color). It is followed by the variable region of about 2000 bp with a high rate of indels (shades of red color). However, it showed some conservation among the subgenome sequences in the left part of the alignment (the most upstream region from the predicted start codon). These similarities may seem comparable to systems of transcription initiation (conserved region adjacent to start codon) and regulation (conserved left upstream-aligned section). The sequencing data had been deposited in the GenBank under accession numbers OP913457 (*MUTE-A1* promoter region, cultivar Nataalka), OP913458 (*MUTE-B1* promoter region, cultivar Nataalka) and OP913459 (*MUTE-D1* promoter region, cultivar Nataalka). The isolated Nataalka homoeologous sequences have a similar pattern of conserved sections. In contrast, the sequence adjacent to the start codon is two times shorter in comparison with CS sequences (well-aligned approximately 150 bp, primarily yellow in Figure 1B). The summary of pairwise comparison between six studied sequences from CS and Nataalka, including distances and percent identity, is represented in Figure 1C. The comparison shows that pairs CS/Nataalka *MUTE-A1* and CS/Nataalka *MUTE-B1* had the highest percent identity values (91.45 and 87.14, respectively). Due to changes, pair CS/Nataalka *MUTE-D1* did not have a significant number of percent identity. Nataalka *MUTE-D1* demonstrated higher similarity with sequences from subgenome B of both cultivars (Nataalka *MUTE-B1*-79.88, and CS *MUTE-B1*-70.20).

Every obtained Nataalka sequence was pairwise aligned with CS reference to reveal cross-cultivar polymorphism. Numerous cultivar-specific SNPs and several indels were indicated between the sequences originated from reference and experimental cultivars. In particular, comparing experimental sequences (cv Nataalka) with reference (cv CS), we observed 35 SNPs and one 276-bp deletion in the *MUTE-A1* promoter region; 7 SNPs and one 205-bp deletion in the *MUTE-B1* promoter region; 15 SNPs, a 4-bp deletion, a 1285-bp deletion, and a 28-bp insertion were spotted across *MUTE-D1* promoter region. Counting SNPs, we included small 1-2-bp indels. Every studied Nataalka sequence had a large deletion compared with the reference; additionally, Nataalka's upstream *MUTE-D1* region had a complementary minor 28-bp insertion. Apparently, the regulation of *MUTE* genes in the genome of Nataalka is less flexible in contrast to Chinese Spring due to the elimination of numerous *cis*-regulatory elements. Having evaluated previous studies on plant promoter region diversity, it was notably spotted that Nataalka sequences of promoter and coding regions of drought-related *TaWRKY2-D1* (from subgenome D) gene were utterly identical with CS (Lakhneko et al., 2021). In the current case study, we detected insignificant conservation, which was denoted in 1285-bp deletion, 28-bp deletion, and a series of SNPs. Such differences in the level of variation between experimental and reference sequences might be related to different homoeologous chromosome groups: *MUTE* genes are allocated on chromosome group 2, while *TaWRKY2* genes belong to chromosome group 1.

Survey of regulatory DNA-elements in wheat *MUTE* promoters

To get deeper insights into the potential regulation of *MUTE* genes in wheat, the revealed six promoter sequences were subjected to detailed analysis of *cis*-regulatory motifs employing the search of 469 entries registered in the New PLACE database (Higo et al., 1999). The search demonstrated that each studied homoeologous *MUTE* promoter region had a TATA-less proximal promoter

containing both the initiator element (Inr-element) and CAAT-box. CS and Nataalka sequences had an equal number of Inr-elements (9 in *MUTE-A1*, and 4 in *MUTE-B1* and *MUTE-D1*) with one Inr-element in the section of the proximal promoter (positions -156...-166). One copy of the central regulator of transcription efficiency CAAT-box was also positioned in the proximal promoter (positions -204...-214) upstream of the start codon in every studied *MUTE* promoter of two cultivars. Thus, the homoeologous proximal promoters of *MUTE* genes had related structures due to the conservation of up to 300 bp section upstream of the start codon. Distal promoter regions also contained abundant CAAT-boxes: 44 and 40 in *MUTE-A1* of CS and Nataalka, respectively; 12 in *MUTE-B1* of both cultivars; then 16 and 11 in *MUTE-D1* of CS and Nataalka, respectively.

The rest of the *cis*-regulatory motifs were divided into five groups: stress responsiveness, compound-responsive elements, light regulation, tissue-specific regulation, and specific pathways (Table 2). The entire promoter region included various *cis*-regulatory motifs responsible for diverse families of stress-related transcription factors. Accordingly, numerous W-boxes, MYC and MYB recognition sites, ARR1-binding elements, LTRE elements, and CBFs were spotted. Four types of W-box were identified in the studied sequences—TGAC, TGACY, TTGAC, and TGACT. Tetranucleotide TGAC was the most abundant among them—12 in *MUTE-A1*, 17 in *MUTE-B1*, and 18 in *MUTE-D1* promoter regions of CS. The divergence of Nataalka sequences resulted in the reduction of the number of TGAC W-boxes—11 in *MUTE-A1* (Figure 2), 12 in *MUTE-B1* (Figure 3), and 9 in *MUTE-D1* (Figure 4). The other three W-box motifs had a similar pattern of declining amounts in Nataalka. Only the SNP at position -339 of *MUTE-A1* contributed to the formation of a new *cis*-regulatory W-box, eliminating the CCAAT-box. Related to abiotic stress responsiveness, the MYC recognition site CANNTG was abundant within the reviewed sequences. *MUTE-A1* of CS has 36 MYC motifs. One-bp insertion at the position -716...-717 led to the elimination of two motifs on the plus and minus strands of the Nataalka sequence. *MUTE-B1* and *MUTE-D1* sequences had 28 and 26 motifs equal in both cultivars. Though substitution T for C (-1795) eliminated two MYC motifs, substitution C for T set up two of them in *MUTE-D1* of cv Nataalka. Alterations also led to the elimination of response regulator ARR1-binding elements (NGATT motif) on homoeologous sequences. Thus, CS *MUTE-A1* has 23 versus 22 motifs in cv Nataalka: 1-bp insertion (-716...-717) set formation and substitutions T for C (-1446), as well as C for T (-2265), led to eliminations of two ARR1-binding elements. The 205-bp deletion (-401...-605) eliminated two ARR1 motifs in *MUTE-B1* as well as 1285-bp deletion (-410...-1694) eliminated six of them in *MUTE-D1* of cv Nataalka. The TGTC motif of disease resistance response element BIHD1OS was also variable between CS and Nataalka sequences. As a result, 276-bp deletion (-1866...-2141) removed one element from *MUTE-A1* (six remained); 205-bp deletion (-401...-605) eliminated two of eight BIHD1OS elements in *MUTE-B1*; and 1285-bp deletion (-410...-1694) eliminated two and substitution C for T at -2392 created one BIHD1OS motif in *MUTE-D1* (totally 5 versus 6 in CS).

The MYB recognition site (YAACKG motif), which is specific for comprehensive abiotic stress responsiveness, including drought (Yang et al., 2016), did not show much variability among sequences. Both CS and Nataalka had five such elements within *MUTE-A1* and two within *MUTE-B1*. The substitution G for T set up the fifth additional site in *MUTE-D1* of cv Nataalka. A low-temperature responsive element (LTRE element) with consensus motif CCGAC was presented in one copy in the *MUTE-B1* sequence both in CS and Nataalka. The SNP A for G at position -219 led to the formation of the single motif in the *MUTE-D1* promoter in cv Nataalka. The single copy of dehydration-responsive element CBF (RYCGAC motif) was removed due to 205-bp deletion (-401...-605) *MUTE-B1*.

Table 2 The collection of most common and most represented *cis*-regulatory elements in the promoter of homoeologues *MUTE* genes of two cultivars

Type of motive	Sequence	Function (Higo et al., 1999)	Total number of occurrences					
			<i>MUTE-A1</i>		<i>MUTE-B1</i>		<i>MUTE-D1</i>	
			CS	Nataalka	CS	Nataalka	CS	Nataalka
<i>Regulation of transcription</i>								
CAAT-box	CAAT	Determine the efficiency of transcription	44	40	12	12	16	11
Inr-element	YTCANTYY	Initiator; light-responsive transcription; TATA-less promoter	9	9	4	4	4	4
	TTATTT		16	16	4	4	9	7
TATA-enhancers	TATTTAA	Transcription initiation / enhancer	5	4	4	3	5	4
	TATAAAT		2	2	1	1	2	1
	TTTATATA		1	2	0	0	0	0
	TATATAA		1	2	1	1	0	0
<i>Stress responsiveness</i>								
W-box	TGAC	Abiotic/biotic stress responsiveness	12	11	17	12	18	9
	TGACY		4	3	7	4	5	2
	TTGAC		3	4	3	3	6	3
	TGACT		1	2	3	1	2	0
MYC recognition site	CANNTG	Abiotic stress responsiveness	36	34	28	28	26	26

Continue Table 2

ARR1-binding element	NGATT	Response regulator	23	22	5	3	10	4
BIHD1OS	TGTCA	Disease resistance responses	7	6	8	6	6	5
MYB recognition site	YAACKG	Abiotic stress responsiveness	5	5	2	2	4	5
LTRE element	CCGAC	Low temperature responsive element, abscisic acid-responsiveness	0	0	1	1	0	1
CBF	RYCGAC	Dehydration-responsive element	0	0	1	0	0	0
<i>Compound-responsive element</i>								
CGCG-box	VCGCGB	Ca ²⁺ -responsive expression	32	26	24	24	37	24
DPBF-element	ACACNNG	Abscisic acid-responsive element	5	5	6	4	7	6
ABRE-related sequence	MACGYGB	Ca ²⁺ -responsive expression	5	5	4	4	6	4
CPBCSPOR	TATTAG	Cytokinin-enhanced protein binding	2	1	1	1	0	0
GARE	TAACAAR	Gibberellin acid-responsive element	1	1	0	0	1	0
ASF-1 binding site	TGACG	Auxin- and salicylic acid-responsive element	1	2	2	2	7	2
<i>Light regulation</i>								
GT-1 binding site	GRWAAW	Light-regulated transcription	32	31	8	6	8	7
GATA-box	GATA	Required for high level, light regulated, and tissue specific expression	20	18	9	8	11	8
ACGT sequence	ACGT	Etiolation-induced expression	12	12	4	4	14	8
I-box	GATAA	Light-regulated transcription	9	8	5	4	5	5
CIACADIANLELHC	CAANNNNAT C	Circadian expression	4	3	0	0	0	0
T-box	ACTTTG	Light-activated transcription	3	2	0	0	0	0
Z-DNA-forming sequence	ATACGTGT	Light-dependent developmental expression	0	1	0	0	1	0
<i>Tissue-specific regulation</i>								
CACT-element	YACT	Mesophyll-specific expression	62	53	24	21	30	20
DOF-element	AAAG	Regulator of plant growth and development	37	31	9	8	12	8
CGACG element	CGACG	Element for G-box	3	2	2	2	8	2
REbeta	CGGATA	Phytochrome regulation	1	0	1	1	1	1
<i>Specific pathways</i>								
CuRE	GTAC	Oxygen response	16	14	6	6	10	6
CCAAT-box	CCAAT	Increase promoter activity, key roles in specific pathways	16	15	5	5	5	4
Site II element	TGGGCY	Element of oxidative phosphorylation machinery	2	3	2	2	1	0
Total number			432	395	213	187	277	187
Total promoter region length			1647	1440	2585	1324	1647	1440
Motif density (total number / total promoter region length)			0.262	0.274	0.082	0.141	0.168	0.130

Legend: N=A/C/T/G, V=A/C/G, B=G/T/C, M=C/A, Y=C/T, W=A/T, R=A/G, CS-Chinese Spring

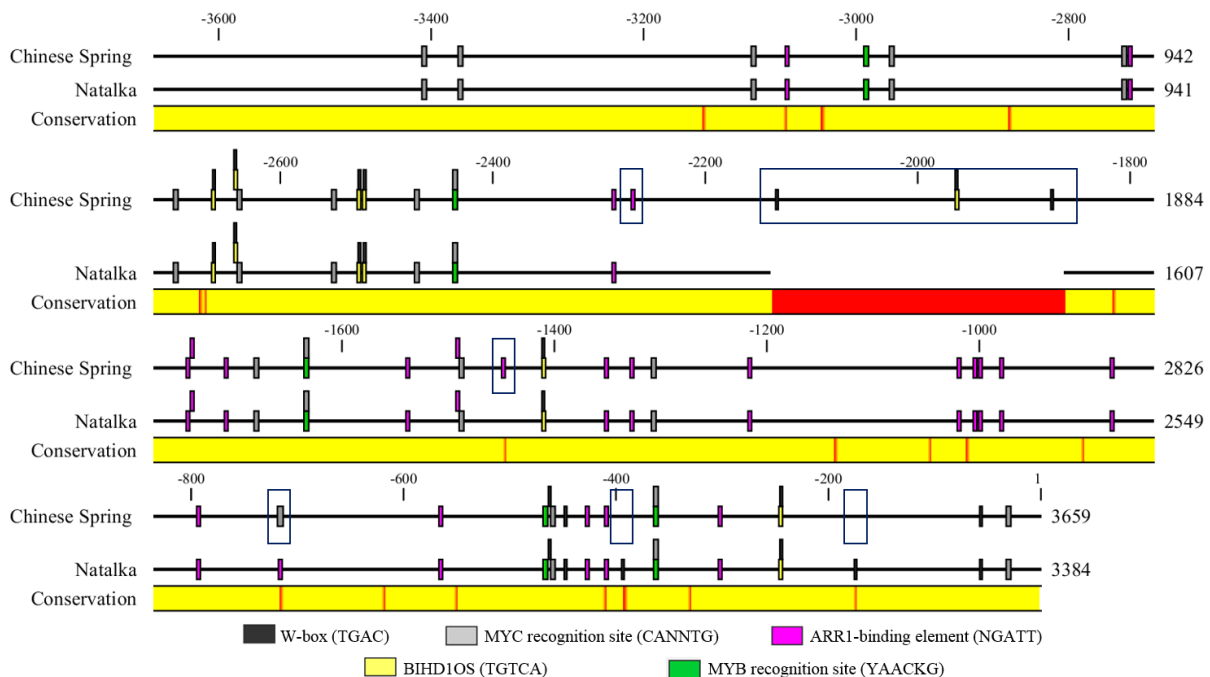


Figure 2 *MUTE-A1* promoter region alignment of the reference Chinese Spring and the isolated Natalka sequences denotes variability in sequence and stress-responsive *cis*-acting elements. Frames point to the diverse composition of *cis*-elements between two sequences. Conservation graphs indicate variable-conserved (from dark red to yellow) sections of these regions.

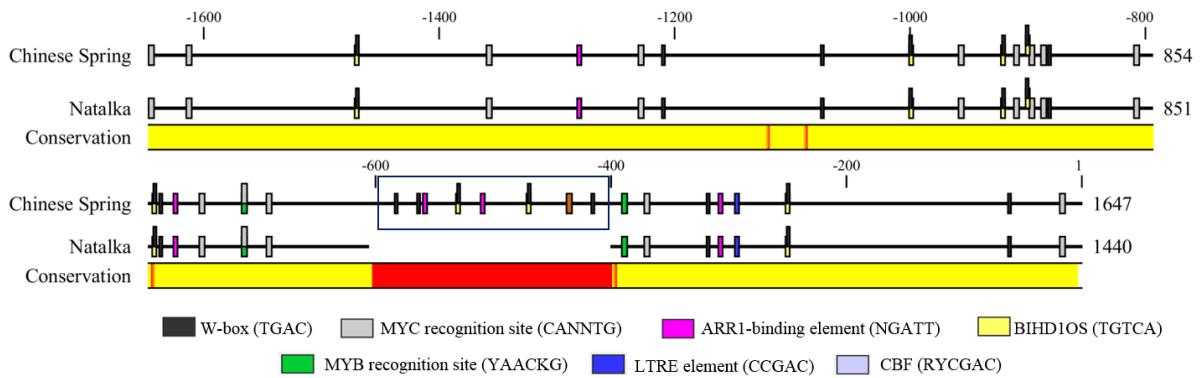


Figure 3 *MUTE-B1* promoter region alignment of the reference Chinese Spring and the isolated Nataalka sequences denotes variability in sequence and stress-responsive *cis*-acting elements. Frames point to the diverse composition of *cis*-elements between two sequences. Conservation graphs indicate variable-conserved (from dark red to yellow) sections of these regions.

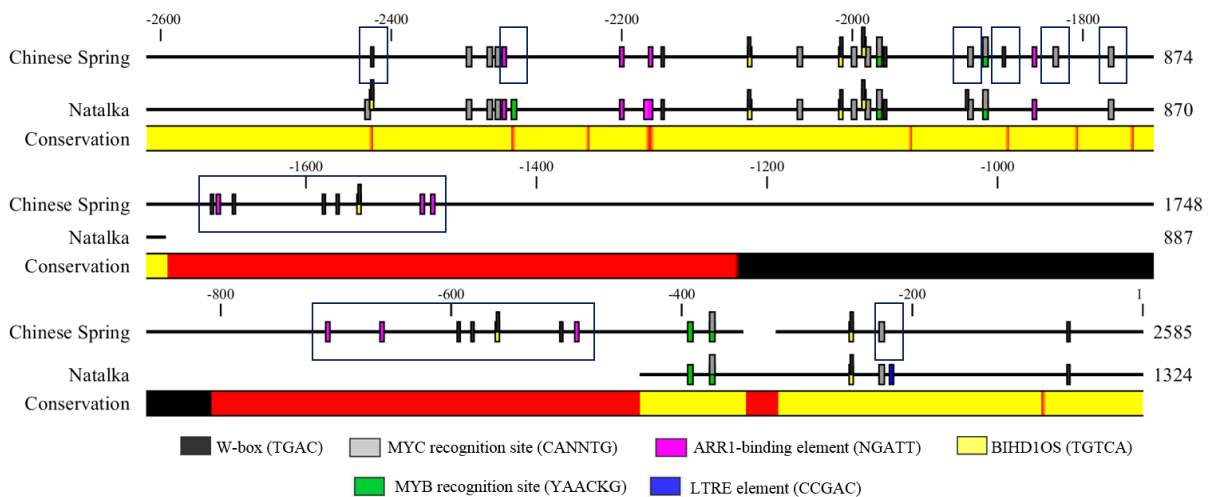


Figure 4 *MUTE-D1* promoter region alignment of the reference Chinese Spring and the isolated Nataalka sequences denotes variability in sequence and stress-responsive *cis*-acting elements. Frames point to the diverse composition of *cis*-elements between two sequences. Conservation graphs indicate variable-conserved (from dark red to yellow) sections of these regions; the black section shows not well-resolved region in Chinese Spring.

The composition of all three *MUTE* proximal promoters, which contains both initiator element (Inr-element) and CAAT-box, corresponds with the general structure of promoters in eukaryotic organisms, in particular plants (Komarnytsky & Borisjuk, 2003; Porto et al., 2014). The homoeologues proximal promoters of *MUTE* genes had related structures due to the conservation of up to 300-bp section before the start codons. Hypothetically, expression of *MUTE-A1* and *MUTE-D1* might be less upregulated in Nataalka than in CS due to reduced copies of the CAAT-box. Also, numerous TATA-based enhancers, which are AT-rich sequences located at different distances from the core promoter, can contribute to the expression efficiency (Singer et al., 1990; Komarnytsky & Borisjuk, 2003). Considering the abundance of various types of *cis*-acting elements, we distinguish that stress-responsive elements, elements for light regulation, and the few representatives of tissue-specific elements contributed more to the total number. Thus, W-boxes are binding sites of WRKY proteins, an extensive family of plant transcription factors that play a crucial role in stress responses, particularly in bread wheat (Ning et al., 2017; Baillo et al., 2020). The CANNTG motif was described as a consensus MYC recognition site in *Arabidopsis* dehydration and cold-responsive genes (Higo et al., 1999). In wheat, MYC transcription factors play a role in plant development and stress response. The CANNTG motif was highly representative across three studied sequences, which denoted a complex regulation network. The multiple calmodulin-binding CGCG-boxes were found to be richly represented across every studied *MUTE* promoter region with higher density in *MUTE-D1*. That pointed to the dependence of expression regulation from the Ca²⁺-based secondary signals indirectly guiding plant development and stress responses (T. Yang & Poovaiah, 2002; F. Yang et al., 2020).

The studied sequences contained a notable amount of light-responsive elements, common for plants, especially for their sequences to be expressed in leaves (Komarnytsky & Borisjuk, 2003; Libantova et al., 2021). It was reported that widespread GT-binding sites were found in a broad range of plant promoters (Zhou, 1999). The cooperation of GT-1 and GATA-box may lead to advanced responsiveness to a broader light spectrum (Puente et al., 1996; Chattopadhyay et al., 1998). The tissue-specific and highly widespread tetranucleotide YACT (CACT-element) was described as the motif denoting mesophyll-specific expression in C4 plants. However, CACT was also supposed to be a *cis*-regulatory element in C3 plants but with no cell specificity in reference C3 promoter from *Flaveria pringlei* (Gowik et al., 2004). The expression of wheat DNA-binding

with one-finger (DOF) proteins exhibits specific patterns in different organs and developmental stages and is impacted by various abiotic stresses (Yanagisawa, 2004; Fang et al., 2020; Liu et al., 2020). The binding motif for DOF proteins (Yanagisawa & Schmidt, 1999) was quite noticeable, especially in the *MUTE-A1* promoter region, potentially contributing to the plasticity of expression regulation. The CCAAT-box binds with transcription factors named Nuclear Factor Y. They are conserved among eukaryotes and play a fundamental biological role in plants (Qu et al., 2015; Yadav et al., 2015; Chaves-Sanjuan et al., 2021). Some of these TFs are expressed ubiquitously, and others are expressed in an environmental- or tissue-specific manner in bread wheat (Stephenson et al., 2007; Khurana et al., 2013; Qu et al., 2015).

Motif density (Table 2) was calculated to describe the total abundance of *cis*-acting elements and to evaluate the significance of lost fragments in cv Nataalka. Only in the case of *MUTE-D1*, we observed the decline of motif density from 0.168 in CS to 0.130 in cv Nataalka even though the reference sequence contained not well-resolved section of 417 bp (-811...-1227) in mentioned large indel which did not contribute to the number of motifs. Vice versa, the detected alteration resulted in the growth of this characteristic for *MUTE-A1* and *MUTE-B1*, supposing the lost fragments were not crucial and did not have a high capacity of elements. The motif density is higher in the *MUTE-A1* promoter region than in other subgenomes, which may denote more flexible or even more complex regulation of expression.

CONCLUSION

Sequence comparison of subgenomic promoter regions of homoeologues *MUTE* genes between two wheat cultivars revealed regions of both high similarity and significant divergence. The promoter sequence similarities between cv Chinese Spring and Nataalka are of different geographic origins and likely represent fundamental principles of the gene's transcription initiation and regulation. Simultaneously, the considerable deletions revealed in the promoter from cv Nataalka compared to Chinese Spring, demonstrate opportunities for significant transcriptional flexibility in regulating stomata biogenesis between wheat cultivars.

Acknowledgments: This study was supported by the National Academy of Sciences of Ukraine project 0122U002213, the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project 09I03-03-V01-00005 to OL as well as by an individual grant from the Huaiyin Normal University (Huai'an, China) to MB.

REFERENCES

- Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., Mohellibi, N., Durand, S., Kimmel, E., Michotey, C., Guerche, C., Loaec, M., Lainé, M., Steinbach, D., Choulet, F., Rimbart, H., Leroy, P., Guilhot, N., Salse, J., ... International Wheat Genome Sequencing Consortium. (2018). Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology*, 19(1), 111. <https://doi.org/10.1186/s13059-018-1491-4>
- Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, 21(2), 71–87. <https://doi.org/10.1038/s41576-019-0173-8>
- Baillou, E. H., Hanif, M. S., Guo, Y., Zhang, Z., Xu, P., & Algam, S. A. (2020). Genome-wide identification of WRKY transcription factor family members in sorghum (*Sorghum bicolor* (L.) moench). *PLOS ONE*, 15(8), e0236651. <https://doi.org/10.1371/journal.pone.0236651>
- Bi, H., Shi, J., Kovalchuk, N., Luang, S., Bazanova, N., Chirkova, L., Zhang, D., Shavruk, Y., Stepanenko, A., Tricker, P., Langridge, P., Hrmova, M., Lopato, S., & Borisjuk, N. (2018). Overexpression of the TaSHN1 transcription factor in bread wheat leads to leaf surface modifications, improved drought tolerance, and no yield penalty under controlled growth conditions. *Plant, Cell & Environment*, 41(11), 2549–2566. <https://doi.org/10.1111/pce.13339>
- Caine, R. S., Chater, C. C., Kamisugi, Y., Cuming, A. C., Beerling, D. J., Gray, J. E., & Fleming, A. J. (2016). An ancestral stomatal patterning module revealed in the non-vascular land plant *Physcomitrella patens*. *Development*, 143(18), 3306–3314. <https://doi.org/10.1242/dev.135038>
- Chattopadhyay, S., Puente, P., Deng, X.-W., & Wei, N. (1998). Combinatorial interaction of light-responsive elements plays a critical role in determining the response characteristics of light-regulated promoters in Arabidopsis. *The Plant Journal*, 15(1), 69–77. <https://doi.org/10.1046/j.1365-313X.1998.00180.x>
- Chaves-Sanjuán, A., Gnesutta, N., Gobbi, A., Martignago, D., Bernardini, A., Fornara, F., Mantovani, R., & Nardini, M. (2021). Structural determinants for NF-Y subunit organization and NF-Y/DNA association in plants. *The Plant Journal*, 105(1), 49–61. <https://doi.org/10.1111/tpj.15038>
- Das, S., & Bansal, M. (2019). Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. *PLOS ONE*, 14(3), e0212678. <https://doi.org/10.1371/journal.pone.0212678>
- Duursma, R. A., Blackman, C. J., López, R., Martin-StPaul, N. K., Cochard, H., & Medlyn, B. E. (2019). On the minimum leaf conductance: Its role in models of plant water use, and ecological and environmental controls. *New Phytologist*, 221(2), 693–705. <https://doi.org/10.1111/nph.15395>
- Fang, Z., Jiang, W., He, Y., Ma, D., Liu, Y., Wang, S., Zhang, Y., & Yin, J. (2020). Genome-wide identification, structure characterization, and expression profiling of Dof transcription factor gene family in wheat (*Triticum aestivum* L.). *Agronomy*, 10(2), 294. <https://doi.org/10.3390/agronomy10020294>
- Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., & Westhoff, P. (2004). Cis-regulatory elements for mesophyll-specific gene expression in the C4 plant *Flaveria trinervia*, the promoter of the C4 phosphoenolpyruvate carboxylase Gene. *The Plant Cell*, 16(5), 1077–1090. <https://doi.org/10.1105/tpc.019729>
- Guo, X., Wang, L., & Dong, J. (2021). Establishing asymmetry: Stomatal division and differentiation in plants. *New Phytologist*, 232(1), 60–67. <https://doi.org/10.1111/nph.17613>
- Han, S.-K., Qi, X., Sugihara, K., Dang, J. H., Endo, T. A., Miller, K. L., Kim, E.-D., Miura, T., & Torii, K. U. (2018). MUTE directly orchestrates cell-state switch and the single symmetric division to create stomata. *Developmental Cell*, 45(3), 303–315.e5. <https://doi.org/10.1016/j.devcel.2018.04.010>
- Hepworth, C., Doheny-Adams, T., Hunt, L., Cameron, D. D., & Gray, J. E. (2015). Manipulating stomatal density enhances drought tolerance without deleterious effect on nutrient uptake. *New Phytologist*, 208(2), 336–341. <https://doi.org/10.1111/nph.13598>
- Higo, K., Ugawa, Y., Iwamoto, M., & Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research*, 27(1), 297–300. <https://doi.org/10.1093/nar/27.1.297>
- Hughes, J., Hepworth, C., Dutton, C., Dunn, J. A., Hunt, L., Stephens, J., Waugh, R., Cameron, D. D., & Gray, J. E. (2017). Reducing stomatal density in barley improves drought tolerance without impacting on yield. *Plant Physiology*, 174(2), 776–787. <https://doi.org/10.1104/pp.16.01844>
- Khurana, N., Chauhan, H., & Khurana, P. (2013). Wheat chloroplast targeted sHSP26 promoter confers heat and abiotic stress inducible expression in transgenic *Arabidopsis* plants. *PLOS ONE*, 8(1), e54418. <https://doi.org/10.1371/journal.pone.0054418>
- Komarnytsky, S., & Borisjuk, N. (2003). *Functional analysis of promoter elements in plants*. 25, 113–141. https://doi.org/10.1007/978-1-4615-0073-5_6
- Korkuč, P., Schippers, J. H. M., & Walther, D. (2014). Characterization and identification of cis-regulatory elements in Arabidopsis based on single-nucleotide polymorphism information. *Plant Physiology*, 164(1), 181–200. <https://doi.org/10.1104/pp.113.229716>
- Lakhneko, O., Stepanenko, A., Kuzminskiy, Y., Borisjuk, N., & Morgun, B. (2021). Survey of drought-associated *TaWRKY2-D1* gene diversity in bread wheat and wheat relatives. *Molecular Biotechnology*, 63(10), 953–962. <https://doi.org/10.1007/s12033-021-00350-7>
- Lau, O. S., Song, Z., Zhou, Z., Davies, K. A., Chang, J., Yang, X., Wang, S., Lucyshyn, D., Tay, I. H. Z., Wigge, P. A., & Bergmann, D. C. (2018). Direct control of SPEECHLESS by PIF4 in the high-temperature response of stomatal development. *Current Biology*, 28(8), 1273–1280.e3. <https://doi.org/10.1016/j.cub.2018.02.054>
- Le, J., Liu, X.-G., Yang, K.-Z., Chen, X.-L., Zou, J.-J., Wang, H.-Z., Wang, M., Vanneste, S., Morita, M., Tasaka, M., Ding, Z.-J., Friml, J., Beeckman, T., & Sack, F. (2014). Auxin transport and activity regulate stomatal patterning and development. *Nature Communications*, 5(1), 3090. <https://doi.org/10.1038/ncomms4090>
- Libantova, J., Frátrikova, M., Jopcik, M., Bauer, M., & Rajnec, M. (2021). *In silico* characterization of B-1,3-glucanase promoter from *Drosera binata*. *Journal of Microbiology, Biotechnology and Food Sciences*, 11(3), e2670. <https://doi.org/10.15414/jmbfs.2670>
- Liu, Y., Liu, N., Deng, X., Liu, D., Li, M., Cui, D., Hu, Y., & Yan, Y. (2020). Genome-wide analysis of wheat DNA-binding with one finger (Dof) transcription factor genes: Evolutionary characteristics and diverse abiotic stress responses. *BMC Genomics*, 21(1), 276. <https://doi.org/10.1186/s12864-020-6691-0>
- Morgun, V. V., Sanin, Ye. V., & Shvartau, V. V. (2014). *Club of 100 centners. Modern cultivars and systems of nutrition and protection of winter wheat. Edition VIII*. Lohos. (In Ukrainian).
- Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, 8(19), 4321–4325. PubMed. <https://doi.org/10.1093/nar/8.19.4321>
- Ning, P., Liu, C., Kang, J., & Lv, J. (2017). Genome-wide analysis of WRKY transcription factors in wheat (*Triticum aestivum* L.) and differential expression under water deficit condition. *PeerJ*, 5, e3232–e3232. PubMed. <https://doi.org/10.7717/peerj.3232>
- Ortiz, R., Sayre, K. D., Govaerts, B., Gupta, R., Subbarao, G. V., Ban, T., Hodson, D., Dixon, J. M., Iván Ortiz-Monasterio, J., & Reynolds, M. (2008). Climate change: Can wheat beat the heat? *Agriculture, Ecosystems & Environment*, 126(1–2), 46–58. <https://doi.org/10.1016/j.agee.2008.01.019>
- Porto, M. S., Pinheiro, M. P. N., Batista, V. G. L., dos Santos, R. C., de Albuquerque Melo Filho, P., & de Lima, L. M. (2014). Plant promoters: An approach of structure and function. *Molecular Biotechnology*, 56(1), 38–49. <https://doi.org/10.1007/s12033-013-9713-1>
- Puente, P., Wei, N., & Deng, X. W. (1996). Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis. *The EMBO Journal*, 15(14), 3732–3743. <https://doi.org/10.1002/j.1460-2075.1996.tb00743.x>
- Qi, X., Han, S.-K., Dang, J. H., Garrick, J. M., Ito, M., Hofstetter, A. K., & Torii, K. U. (2017). Autocrine regulation of stomatal differentiation potential by EPFL and ERECTA-LIKE1 ligand-receptor signaling. *eLife*, 6, e24102. <https://doi.org/10.7554/eLife.24102>
- Qu, B., He, X., Wang, J., Zhao, Y., Teng, W., Shao, A., Zhao, X., Ma, W., Wang, J., Li, B., Li, Z., & Tong, Y. (2015). A wheat CCAAT box-binding transcription factor increases the grain yield of wheat with less fertilizer input. *Plant Physiology*, 167(2), 411–423. <https://doi.org/10.1104/pp.114.246959>
- Raissig, M. T., Abrash, E., Bettadapur, A., Vogel, J. P., & Bergmann, D. C. (2016). Grasses use an alternatively wired bHLH transcription factor network to establish stomatal identity. *Proceedings of the National Academy of Sciences*, 113(29), 8326–8331. <https://doi.org/10.1073/pnas.1606728113>
- Raissig, M. T., Matos, J. L., Anleu Gil, M. X., Kornfeld, A., Bettadapur, A., Abrash, E., Allison, H. R., Badgley, G., Vogel, J. P., Berry, J. A., & Bergmann, D. C. (2017). Mobile MUTE specifies subsidiary cells to build physiologically improved grass stomata. *Science*, 355(6330), 1215–1218. <https://doi.org/10.1126/science.aal3254>
- Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., Noshay, J. M., Galli, M., Mejía-Guerra, M. K., Colomé-Tatché, M., Johannes, F., Rowley, M. J., Corces, V. G., Zhai, J., Scanlon, M. J., Buckler, E. S., Gallavotti, A., Springer, N. M., Schmitz, R. J., & Zhang, X. (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nature Plants*, 5(12), 1237–1249. <https://doi.org/10.1038/s41477-019-0547-0>
- Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. 3rd Edition. Cold Spring Harbor Laboratory Press.
- Schmitz, R. J., Grotewold, E., & Stam, M. (2022). Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *The Plant Cell*, 34(2), 718–741. <https://doi.org/10.1093/plcell/koab281>
- Serna, L. (2020). The role of grass MUTE orthologues during stomatal development. *Frontiers in Plant Science*, 11, 55. PubMed. <https://doi.org/10.3389/fpls.2020.00055>

- Singer, V. L., Wobbe, C. R., & Struhl, K. (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes & Development*, 4(4), 636–645. <https://doi.org/10.1101/gad.4.4.636>
- Šramková, Z., Gregová, E., & Šturdík, E. (2021). Genetic improvement of wheat—A review. *Nova Biotechnologica et Chimica*, 9(1), 27–52. <https://doi.org/10.36547/nbc.1285>
- Stephenson, T. J., McIntyre, C. L., Collet, C., & Xue, G.-P. (2007). Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*. *Plant Molecular Biology*, 65(1), 77–92. <https://doi.org/10.1007/s11103-007-9200-9>
- Takata, N., Yokota, K., Ohki, S., Mori, M., Taniguchi, T., & Kurita, M. (2013). Evolutionary relationship and structural characterization of the EPF/EPFL gene family. *PLOS ONE*, 8(6), e65183. <https://doi.org/10.1371/journal.pone.0065183>
- Vo Ngoc, L., Kassavetis, G. A., & Kadonaga, J. T. (2019). The RNA polymerase II core promoter in *Drosophila*. *Genetics*, 212(1), 13–24. <https://doi.org/10.1534/genetics.119.302021>
- Yadav, D., Shavrukov, Y., Bazanova, N., Chirkova, L., Borisjuk, N., Kovalchuk, N., Ismagul, A., Parent, B., Langridge, P., Hrmova, M., & Lopato, S. (2015). Constitutive overexpression of the *TaNf-YB4* gene in transgenic wheat significantly improves grain yield. *Journal of Experimental Botany*, 66(21), 6635–6650. <https://doi.org/10.1093/jxb/erv370>
- Yanagisawa, S. (2004). Dof domain proteins: Plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant and Cell Physiology*, 45(4), 386–391. <https://doi.org/10.1093/pcp/pch055>
- Yanagisawa, S., & Schmidt, R. J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. *The Plant Journal*, 17(2), 209–214. <https://doi.org/10.1046/j.1365-3113X.1999.00363.x>
- Yang, F., Dong, F., Hu, F., Liu, Y., Chai, J., Zhao, H., Lv, M., & Zhou, S. (2020). Genome-wide identification and expression analysis of the calmodulin-binding transcription activator (CAMTA) gene family in wheat (*Triticum aestivum* L.). *BMC Genetics*, 21(1), 105. <https://doi.org/10.1186/s12863-020-00916-5>
- Yang, T., & Poovaiah, B. W. (2002). A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. *Journal of Biological Chemistry*, 277(47), 45049–45058. <https://doi.org/10.1074/jbc.M207941200>
- Zhou, D.-X. (1999). Regulatory mechanism of plant gene transcription by GT-elements and GT-factors. *Trends in Plant Science*, 4(6), 210–214. [https://doi.org/10.1016/S1360-1385\(99\)01418-1](https://doi.org/10.1016/S1360-1385(99)01418-1)
- Zou, C., Sun, K., Mackaluso, J. D., Seddon, A. E., Jin, R., Thomashow, M. F., & Shiu, S.-H. (2011). Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 108(36), 14992–14997. <https://doi.org/10.1073/pnas.1103202108>