RESEARCH ARTICLE

Genome-wide identification and expression profiling of the COBRA-like genes reveal likely roles in stem strength in rapeseed (Brassica napus L.)

Qian Yang, Shan Wang, Hao Chen, Liang You, Fangying Liu, Zhongsong Liu

College of Agronomy, Hunan Agricultural University, Changsha, Hunan, China

* zliu48@sohu.com

Abstract

The COBRA-like (COBL) genes play key roles in cell anisotropic expansion and the orientation of microfibrils. Mutations in these genes cause the brittle stem and induce pathogen responsive phenotypes in Arabidopsis and several crop plants. In this study, an in silico genome-wide analysis was performed to identify the COBL family members in Brassica. We identified 44, 20 and 23 COBL genes in B. napus and its diploid progenitor species B. rapa and B. oleracea, respectively. All the predicted COBL genes were phylogenetically clustered into two groups: the AtCOB group and the AtCOBL7 group. The conserved chromosome locations of COBLs in Arabidopsis and Brassica, together with clustering, indicated that the expansion of the COBL gene family in B. napus was primarily attributable to whole-genome triplication. Among the BnaCOBLs, 22 contained all the conserved motifs and derived from 9 of 12 subgroups. RNA-seq analysis was used to determine the tissue preferential expression patterns of various subgroups. BnaCOBL9, BnaCOBL35 and BnaCOBL41 were highly expressed in stem with high-breaking resistance, which implies these AtCOB subgroup members may be involved in stem development and stem breaking resistance of rapeseed. Our results of this study may help to elucidate the molecular properties of the COBRA gene family and provide informative clues for high stem-breaking resistance studies.

Introduction

Plant morphogenesis is dependent on the regulation of cell division and expansion. Most plant cells grow anisotropically through internal and isotropic turgor pressure yield from cell walls [1]. The plant cell wall is a dynamic, complex fibrillar network. After the plant cell expands to its final shape and the primary cell wall is formed, the secondary cell wall is formed and thickens between the primary cell wall and plasma membrane [2, 3]. The COBRA gene, which encodes a glycosylphosphatidylinositol (GPI) anchored protein [1, 4, 5], regulates microfibril deposition on the cell surface at the rapid elongation stage to guarantee a normal anisotropic expansion of the cell wall during plant morphogenesis.
The COBL family is conserved in monocots and eudicots [2, 5]. In Arabidopsis (Arabidopsis thaliana), there are 12 COBLs (AtCOBLs), which can be divided into two groups based on their protein sequences [1], one showing strong similarity to COBRA while the other exhibiting high similarity to AtCOBL7. There are 11, 11, and 10 COBLs that have been identified in the monocots rice (Oryza sativa ssp. japonica) [6], maize (Zea mays) [2], and sorghum (Sorghum bicolor) [7], respectively. Additionally, 17, 18, 24, and 33 COBLs have been reported in the eudicots tomato (Solanum lycopersicum) [8], Populus (Populus L.) [9], soybean (Glycine max) [10], and cotton (Gossypium spp.) [11], respectively. It appears that the family members increased to a certain extent in eudicots but remained almost constant in monocots. This phenomenon of expansion is presumed to be derived from whole-genome duplication [10, 11] and segmental duplication [5]. The phylogenetic relationship was similar to that of Arabidopsis among various reported species [10]. In the COBRA group, the AtCOB orthologous subgroup was predicted to be a sister clade of the AtCOBL4 subgroup and derived more recently after the division between monocots and eudicots [5].

The COBLs members have been found to mediate diverse physiological and developmental processes such as stem strength [6], pollen tube growth [12], pathogen resistance [13], and root-hair growth [4]. Silencing a COBL member, such as BRITTLE CULM1 (OsBC1) in rice, Brittle stalk 2 (ZmBk2) in maize, BRITTLE CULM1 (SbBC1) in sorghum, and TmBr1 in diploid wheat, caused plants to exhibit the brittle phenotype [6, 7, 14, 15]. Cuticle lacking, abnormal shape, and irregular size distribution were observed in the epidermal cells of a tomato mutant in which the SICOBRA-like gene was repressed. These phenotypes resulted in extensive non-uniform cracking on the surface of the immature green fruits of these plants [8]. Mutations in AtCOBL10 were observed to cause gametophytic male sterility due to reduced pollen tube growth and compromised directional sensing in the female transmitting tract [12, 16, 17].

Rapeseed (Brassica napus L. AACC, 2n = 38) which supplies approximately 13–16% of vegetable oil worldwide [18], is an allotetraploid species that was formed approximately 7,500–12,500 years ago by a spontaneous cross of the diploid progenitors B. rapa (AA, 2n = 20) and B. oleracea (CC, 2n = 18) [19]. In this study, we identified COBL genes at the genome-wide level and performed a comprehensive in silico analysis including characterization of phylogeny, gene structure, conserved motifs, and chromosomal collinearity in rapeseed and its progenitors. We also evaluated the expression patterns of these genes in various tissues as well as stems with different stem breaking resistance (SBR) by transcriptome sequencing. Our results may help to further characterize the functions of COBL family, and provide clues for stem strength in rapeseed.

**Materials and methods**

**Genome-wide identification of COBLs in Brassica napus and its both progenitor species**

The B. napus (cv. ZhongShuang11, ZS11) genome sequence was downloaded from the BnPIR database (http://cbi.hzau.edu.cn/bnpir/index.php) [18, 20]. The genome sequences, CDSs and annotation files of B. rapa (v3.0) and B. oleracea (HDEM) were retrieved from the Brassica Database (BRAD, http://brassicadb.cn). The Arabidopsis COBL protein sequences were
obtained from TAIR (http://www.arabidopsis.org) [5], and used as the query to identify COBL homologs in *B. napus*, *B. rapa* and *B. oleracea* by BLASTP [21], with the e-value being 1E-10. After redundant sequences and incomplete sequences were removed, the remaining protein sequences were submitted to SMART tools and the NCBI Conserved Domain Search Database to confirm the presence of previously characterized domains in the candidate sequences; sequences without COBRA domains were excluded from the downstream analysis [22].

The physicochemical parameters of BnaCOBL proteins, including the molecular weights (in kDa) and isoelectric points (pIs), were calculated by ExPASy [23]. The subcellular location of COBL proteins were predicted by Cell-PLoc v2.0 (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/).

**Multiple alignments and phylogenetic analysis of COBLs from Brassica and Arabidopsis**

All the predicted COBL protein sequences of *B. napus*, *B. rapa*, and *B. oleracea*, and the AtCOBLS protein sequences were aligned by Multiple Sequence Comparison by Log-Expectation (MUSCLE) [24]. The phylogenetic tree was generated in IQ-tree [25] software using the maximum likelihood (ML) method with 10,000 bootstrap replicates. The “figtree” (http://tree.bio.ed.ac.uk/software/figtree/) was used to draw the phylogenetic tree of COBL protein in four genomes.

**Chromosomal locations and syntenic analyses of COBLs in Brassica napus and its both progenitor species**

The chromosomal positions of the BnaCOBLs were obtained from the genome annotation file of ZS11. The start and end locations of each BnaCOBL were drawn on chromosomes using MapChart [26]. The synteny relationships between the BnaCOBLs and COBLs in *B. rapa*, and *B. oleracea* were evaluated using the McScanX [27] and drawn by TBtools [28].

**Prediction of gene structures, conserved motifs, and cis-acting regulatory elements of BnaCOBLs**

The gene structures (exon-intron) of BnaCOBLs were retrieved from the genome annotation file. The COBRA domain and potential N-glycosylation sites were predicted by GenomeNet Bioinformatics Tools (https://www.genome.jp/) and the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) [29]. The signal peptide, CCVS Cys-rich domain, and potential ω-sites for GPI modification were predicted with Signal 5.0 [30] and the GPI Prediction Server Version 3.0 [31]. Hydrophilicity analysis was performed by ExPASy-ProtScale (https://web.expasy.org/protscale/) [32, 33]. TBtools was used to draw the structural map of BnaCOBLs. To further analyze the COBRA domains of BnaCOBLs, the multi-sequence alignments were carried out by MEGA v7.0 [34] and the results were displayed by GeneDoc (http://www.cris.com/~Ketchup/genedoc.shtml).

To analyze the putative cis-regulatory elements (CAREs) of BnaCOBLs, the promoter regions were defined as the 1.5-kb region upstream of the ATG start codon of each gene (i.e., the 1.5-kb downstream sequences were chosen if a gene was found to map on the opposite strand relative to the sequence strand deposited in the ZS11 genome). These sequences were used to detect the CAREs with the online database PlantCARE [35]. Next, considering the characters of plant core promoter regions, we checked the common promoter elements TATA-box and CAAT-box near the start codon (<500bp), the core promoter elements (i.e., TATA-box, CAAT-box) on the opposite strands of the corresponding genes were filtered out.
of the results because the core promoter regions are direction-sensitive [36]. We classified all the elements into core promoter elements, responsive elements, the temporal and spatial specific or unannotated elements according to their functional annotation.

**Expression analysis of BnaCOBLs in various tissues**

The RNA-seq data obtained from 12 tissues of the rapeseed cultivar ZS11, which was described in a previous study [37], were downloaded from National Center for Biotechnology Information (NCBI) (ID: PRJNA394926) to assess the tissue expression preference of different COBL family members of rapeseed.

For further evaluation of the expression profiles of BnaCOBLs in rapeseed stem, we selected previously reported [38] transcriptome expression data of four stem samples: FH (High stem breaking resistance (SBR) during Flowering), FL (Low-SBR during Flowering), SH (High-SBR during Silique development), and SL (Low-SBR during Silique development). The high SBR sample had averaged SBR of 115.49N; while the low SBR sample had averaged SBR of 31.69N [38]. The raw data were downloaded from the Short Read Archive (SRA) database of NCBI under the accession number SRP142441.

The NGSQCToolkit [39] was used to clean the raw data. The RSEM [40] and STAR [41] softwares were used to map the clean reads to the reference genome of ZS11 and calculate the transcripts per million (TPM) values of each gene, and the heat map of expression of BnaCOBL genes was drawn by TBtools.

**Plant materials and qRT-PCR analysis**

The seed of ZS11, a semi-winter rapeseed cultivar, was kindly provided by Oil Crops Research Institute, Chinese Academy of Agricultural Sciences and sown on the experimental farm of Hunan Agricultural University, Changsha. Three individual plants were harvested at the initial flowering stage. Their stems were cut into two parts, the upper (adjacent to inflorescence) and the lower (the first elongated internode). Fully expanded leaves were used as leaf samples whereas the taproot and the lateral roots were collected separately after being cleaned up.

Quantitative real-time RT–PCR (qRT–PCR) was performed to determine gene expression level. Total RNA was extracted from all sample tissues separately using an RNAqueous kit (Thermo Fisher, AM1912). The yield of RNA was determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA), and the integrity of the RNA was evaluated using agarose gel electrophoresis and staining with ethidium bromide. Each RT reaction consisted of 0.5 μg RNA, 2 μl of 5X TransScript All-in-One SuperMix for qPCR and 0.5 μl of gDNA Remover in a total volume of 10 μl. Reactions were performed in a GeneAmp PCR System 9700 (Applied Biosystems, USA) for 15 min at 42˚C and 5 s at 85˚C. The 10-μl RT reaction mix was subsequently diluted tenfold in nuclease-free water. Real-time PCR was performed using LightCycler® 480 II Real-time PCR Instrument (Roche, Swiss) with 10 μl PCR reaction mixture that included 1 μl of cDNA, 5 μl of 2X PerfectStart™ Green qPCR SuperMix, 0.2 μl of forward primer, 0.2 μl of reverse primer and 3.6 μl of nuclease-free water. Reactions were incubated in a 384-well optical plate (Roche, Swiss) at 94˚C for 30 s followed by 45 cycles of 94˚C for 5 s and 60˚C for 30 s. Each sample was repeated three times. The expression levels of mRNAs were normalized to BnaActin and were calculated using the comparative cycle threshold (Ct) method [42]. The primers were designed at the specific nucleotide among the CDSs of five BnaCOBLs and checked through electronic PCR on the CDSs of these genes. These primer sequences are listed in the S1 Table.
Results
Identification of the COBL genes in *Brassica napus* and its diploid progenitors

A total of 62 putative COBLs were identified in *B. napus* through a BLASTP search using 12 *Arabidopsis* COBL protein sequences as query. These sequences were submitted to SMART and the NCBI CDD (Conserved Domains Database) to confirm the existence of COBRA domains. Finally, 44 candidate COBLs were identified and designated as BnaCOBL1-44 in rapeseed, and their basic information is listed in the S2 Table. Among these proteins, BnaCOBL38 was determined to be the largest with 699 amino acids (aa), whereas BnaCOBL19 was the smallest with 200 aa. The molecular weights and isoelectric points of the BnaCOBLs ranged from 22.06 to 77.68 kDa and 5.25 to 10.09 (S2 Table), respectively. The BnaCOBLs were predicted to localize at the cell membrane (30), extracellular (12), and endoplasmic reticulum (2).

Similarly, we also identified 20 BraCOBLs and 23 BolCOBLs in *B. rapa* and *B. oleracea*, both progenitor species of *B. napus*, respectively. Their gene symbols and chromosomal locations are listed in S2 Table. There were approximately two times as many COBLs in *B. rapa* and *B. oleracea* as in *Arabidopsis*. The sum of COBLs in the diploid progenitors was almost equal to the quantity of BnaCOBLs.

Phylogenetic analysis of the COBL genes from *B. napus*, *B. rapa*, *B. oleracea* and *Arabidopsis*

To unravel the evolutionary relationships among the COBL genes from *B. napus*, *B. rapa*, *B. oleracea* and *Arabidopsis*, a phylogenetic tree was constructed based on whole protein sequences using ML method. As shown in Fig 1, all the COBL members were clustered into two groups, which corresponded with the AtCOB group and the AtCOBL7 group in *Arabidopsis* [5]. The AtCOB group (Group I) contained AtCOB, AtCOBL1-6, 12 BraCOBLs, 15 BolCOBLs, and 25 BnaCOBLs, while the AtCOBL7 (Group II) consisted of AtCOBL7-11, 8 BraCOBLs, 8 BolCOBLs, and 19 BnaCOBLs. Group I contained more COBLs than Group II in the four species analyzed.

Based on the bootstrap values and the topology of the phylogenetic tree, these proteins were further divided into 12 subgroups (Table 1). Each subgroup had COBLs from four species, except the BnaCOBL5/44 subgroup, which lacks COBLs from *Arabidopsis*. The subgroups AtCOBL1, AtCOBL7, AtCOBL8, and AtCOBL9 each retained two BnaCOBLs, while six and eight BnaCOBLs were retained in the subgroups AtCOBL2/3 and AtCOBL11, respectively. The other subgroups had three or five BnaCOBLs. These results indicated an unequal evolution among orthologous subgroups of BnaCOBLs when derived from corresponding AtCOBLs. The subgroup AtCOB was closer to AtCOBL5 in Group I, while the subgroup AtCOBL10 was closer to AtCOBL11 in Group II. This distribution was similar to that in *Arabidopsis*. Based on the triploidy and allotetraploidization events in the evolutionary history of rapeseed, each subgroup of this phylogenetic topology represented a class of orthologous COBLs in Brassica species derived from the corresponding AtCOBL.

Chromosomal locations of COBLs and syntenic analyses between *Brassica napus* and its progenitor

The BnaCOBLs were unevenly distributed on 16 of 19 chromosomes (except for A04, C04 and C06) of rapeseed, with one to five members on each chromosome (Fig 2 and S2 Table). The BnaCOBLs were asymmetrically distributed in subgenomes: 19 were detected in the A subgenome, and 25 were detected in the C subgenome. However, the locations of BnaCOBLs on
chromosome A01, A02, and A09 were much the same as the locations of their homologous C01, C02, and C09. Even the BnaCOBLs on A01 and BnaCOBLs at the homologous region on C01 were determined to belong to the same subgroups. This kind of gene pairs were also observed on some other homologous chromosomes.

Chalhoub et al [19] reported that 80.0% of genes in B. napus (cv. Damor) were orthologous to the genes of B. rapa and B. oleracea. Based on protein sequence identity and phylogenetic topology, we identified 18 and 21 orthologous gene pairs (S3 Table) between the subgenomes of rapeseed and their respective ancestral genomes. The locations of these orthologous pairs of COBLs showed high similarity between B. napus and B. rapa or B. oleracea, respectively (Fig 3A). We found two BraCOBLs (BraCOBL5 and BraCOBL13) and three BolCOBLs (BolCOBL6, BolCOBL16, and BolCOBL17) have lost their orthologous gene pair in B. napus. On the other hand, the five BnaCOBLs (BnaCOBL14, BnaCOBL27, BnaCOBL31, BnaCOBL43, and BnaCOBL44) did not detect orthologs in either B. rapa or B. oleracea.
Table 1. Domains of BnaCOBL proteins in rapeseed.

| Gene Name  | Chr. | Subgroup | CCVS\(^1\) | N-terminal secretion signal cleavage site | \(\omega\)-site\(^2\) | Hydrophobic C-terminal |
|------------|------|----------|------------|------------------------------------------|----------------|---------------------|
| BnaCOBL9   | A03  | AtCOB    | 233        | TEA-YD                                   | N431 2.3\text{e}-07 yes |                     |
| BnaCOBL35  | C07  | AtCOB    | 233        | TEA-YD                                   | N431 2.4\text{e}-06 yes |                     |
| BnaCOBL41  | C09  | AtCOB    | 233        | TEA-YD                                   | N431 2.2\text{e}-07 yes |                     |
| BnaCOBL32  | C05  | AtCOBL1  | 231        | ADA-YD                                   | N428 3.7\text{e}-04 yes |                     |
| BnaCOBL33  | C05  | AtCOBL1  | 232        | ADA-YD                                   | A428 2.9\text{e}-04 yes |                     |
| BnaCOBL27  | C03  | AtCOBL2/3| -          |                                          | A187 - yes            |                     |
| BnaCOBL6   | A02  | AtCOBL2/3| 223        | TEA-YD                                   | N415 2.8\text{e}-05 yes |                     |
| BnaCOBL13  | A06  | AtCOBL2/3| 223        | TEA-YD                                   | N412 2.5\text{e}-07 yes |                     |
| BnaCOBL25  | C02  | AtCOBL2/3| 223        | TEA-YD                                   | N415 2.1\text{e}-04 yes |                     |
| BnaCOBL30  | C05  | AtCOBL2/3| -          |                                          | G296 - yes            |                     |
| BnaCOBL34  | C07  | AtCOBL2/3| 223        | TEA-YD                                   | N412 1.6\text{e}-07 yes |                     |
| BnaCOBL7   | A03  | AtCOBL4  |            | W526 - yes                               |                     |                     |
| BnaCOBL19  | A10  | AtCOBL4  |            | G170 - no                                |                     |                     |
| BnaCOBL24  | C02  | AtCOBL4  |            | G215 - no                                |                     |                     |
| BnaCOBL26  | C03  | AtCOBL4  |            | M204 - no                                |                     |                     |
| BnaCOBL10  | A03  | AtCOBL5  |            | M184 - no                                |                     |                     |
| BnaCOBL18  | A09  | AtCOBL5  |            | G382 - no                                |                     |                     |
| BnaCOBL36  | C07  | AtCOBL5  |            | S209 - no                                |                     |                     |
| BnaCOBL42  | C09  | AtCOBL5  |            | T183 - no                                |                     |                     |
| BnaCOBL12  | A06  | AtCOBL6  | 222        | SHG-YD                                   | S270 - no            |                     |
| BnaCOBL16  | A08  | AtCOBL6  | 219        | THG-FD                                   | S412 1.9\text{e}-07 yes |                     |
| BnaCOBL29  | C05  | AtCOBL6  | 221        | SHG-YD                                   | R548 - no            |                     |
| BnaCOBL39  | C08  | AtCOBL6  | 218        | THG-FD                                   | S411 1.1\text{e}-07 yes |                     |
| BnaCOBL5   | A02  | -        |            | SLG-RY                                   | M186 - no            |                     |
| BnaCOBL44  | C09  | -        |            | M499 - no                                |                     |                     |
| BnaCOBL2   | A01  | AtCOBL7  | 420        | TTS-QS                                   | S635 2.3\text{e}-05 yes |                     |
| BnaCOBL21  | C01  | AtCOBL7  | 419        | TAS-QS                                   | N635 3.0\text{e}-05 yes |                     |
| BnaCOBL4   | A01  | AtCOBL8  | 427        | TSS-QP                                   | S642 6.6\text{e}-06 yes |                     |
| BnaCOBL23  | C01  | AtCOBL8  | 423        | TSS-QQ                                   | N638 9.3\text{e}-06 yes |                     |
| BnaCOBL17  | A09  | AtCOBL9  | 421        | SLS-QL                                   | G638 9.8\text{e}-05 yes |                     |
| BnaCOBL40  | C09  | AtCOBL9  | 421        | SLS-QL                                   | S638 2.3\text{e}-05 yes |                     |
| BnaCOBL3   | A01  | AtCOBL10 | 433        | CNG-QD                                   | S646 1.6\text{e}-05 yes |                     |
| BnaCOBL8   | A03  | AtCOBL10 | 432        | CNG-QD                                   | S645 1.3\text{e}-05 yes |                     |
| BnaCOBL11  | A05  | AtCOBL10 | -          |                                          | G237 - no            |                     |
| BnaCOBL22  | C01  | AtCOBL10 | 433        | CNG-QD                                   | S646 - yes            |                     |
| BnaCOBL28  | C03  | AtCOBL10 | 422        | CNG-QD                                   | S635 9.2\text{e}-06 yes |                     |
| BnaCOBL1   | A01  | AtCOBL11 | 423        | SFA-QD                                   | S635 2.4\text{e}-05 yes |                     |
| BnaCOBL14  | A07  | AtCOBL11 | -          |                                          | A237 - no            |                     |
| BnaCOBL15  | A08  | AtCOBL11 | 428        | SLA-QD                                   | Y641 - yes            |                     |
| BnaCOBL20  | C01  | AtCOBL11 | 425        | SRA-QD                                   | S637 3.0\text{e}-05 yes |                     |
| BnaCOBL31  | C05  | AtCOBL11 | -          |                                          | S198 - yes            |                     |
| BnaCOBL37  | C08  | AtCOBL11 | -          |                                          | E188 - no            |                     |
| BnaCOBL38  | C08  | AtCOBL11 | 458        | -                                        | S670 - yes            |                     |
| BnaCOBL43  | C09  | AtCOBL11 | -          |                                          | S198 - yes            |                     |

\(^1\)The start site of the CCVS domain.

\(^2\)The amino acid and location of the \(\omega\)-site (GPI attachment cleavage site). A site represented in italics means that the confidence of this prediction did not reach the threshold.

https://doi.org/10.1371/journal.pone.0260268.t001
The *BnaCOBL* gene clusters [43–45] containing two or three *BnaCOBL* genes appear on the chromosome A03, C05, C07, C08, and C09 (Fig 2). The cluster on A03, C07, and C09 exhibited the same order on the “X block” [46] of *B. rapa*, *B. oleracea* and *Arabidopsis* (Fig 3B). The cluster on C08 was detected in *B. oleracea* but not in *Arabidopsis*. The cluster on C05 was detected only in *B. napus*, and this cluster may have been formed by segmental duplication in *B. napus* according to sequence and annotation of the ZS11 genome. These results suggested that the COBL gene family in rapeseed changed little during the allotetraploidization event from *B. rapa* and *B. oleracea*.

**Structure and conserved domains of BnaCOBLs**

We characterized gene structure and motif domains of the *BnaCOBL* genes. These genes had 2–12 exons (Fig 4). The number of exons varies between two groups. In Group I, 17 of 25 members possessed over 6 exons, whereas all members in Group II were determined to have only two to four exons. However, the average length of proteins was observed to be longer in Group II than in Group I.

Fig 2. Chromosomal locations of *BnaCOBL*. The blue and the green bar represent A- and C-subgenome chromosomes, respectively. The gene name is presented to the right of each bar, while the chromosome name is to the left. To the left of the A/C subgenome is a 10-Mb bar. The clusters are shown with a red frame.

https://doi.org/10.1371/journal.pone.0260268.g002
All the BnaCOBLs were observed to have the COBRA domain (Fig 4). The COBRA domain of the Group I members is close to the N-terminal secretion signal peptide, while that of the Group II members is in the middle of the protein sequences. But there are exceptions to this rule. The multiple-sequence alnements showed eight BnaCOBLs whose COBRA domain is largely defective (S1 Fig). For example, BnaCOBL31, BnaCOBL43 and BnaCOBL37 only retained 29 to 58 amino acids at C-terminal. The COBRA domain showed significant divergence between 12 subgroups and high conservation within the subgroup although it had almost the same in one-third amino acid residues among family members. One or more potential N-glycosylation sites were distributed to all BnaCOBLs. Thirty-four of these proteins were identified beginning with an N-terminal secretion signal. The CCVS (Cys-rich) motif of 27 BnaCOBLs was observed to have seven to ten amino acids away from the C-terminal of the COBRA domain. Half of the family members (Fig 4 and Table 1) were determined to have all the conserved domains, so did the ω-sites follow by a hydrophobic C-terminal domain.

Compared to AtCOBL4, BC1, BK2, the orthologous BnaCOBLs lost the C-terminal motif of COBRA domain, CCVS, and ω-site (S2 Fig) because of the exon skipping (BnaCOBL7 and BnaCOBL26) or the intron-retention (BnaCOBL19 and BnaCOBL24). These similar alternative splicing events were also identified in other rapeseed sequenced genomes so that no complete BnaCOBL7 can be found in the rapeseed pan-genome. The three members of AtCOB subgroup were conserved among all nine rapeseed genomes, which is shown in S2 Fig. In this subgroup, only two orthologous COBLs of B. rapa and B. oleracea were defective and also disappeared from B. napus.

**Cis-acting regulatory elements in the promoter region of BnaCOBLs**

The control over gene transcription via upstream cis-acting regulatory elements (CAREs) is the most prominent mechanism governing gene expression regulation [47]. The analysis of CAREs may help elucidate the expression levels of BnaCOBLs in specific tissues and conditions [48]. To predict putative cis-elements in the BnaCOBLs, DNA sequences 1500 bp upstream of the start codon (ATG) were searched for in the PlantCARE database to identify the CAREs associated with plant growth, development, and stress response. Eighty-five CAREs were
found in all BnaCOBLs. All promoter regions of BnaCOBLs contained CAAT-box which is the major determinant of promoter efficiency. Three or more TATA-boxes were found in all the genes except two, BnaCOBL10 and BnaCOBL36 which had the TATA-less type of promoters [49].

We also analyzed the phytohormones and environment responsive elements (Fig 5 and S4 Table). The stress-related CAREs (S4 Table) were the most common and identified among all the BnaCOBLs. These stress-related CAREs included MYB, MYC, ARE and as-1, which correspond to abiotic and biotic stress. The most frequent stress CAREs were the MYCs, a dehydration-responsive element. The BnaCOBLs were probably regulated by methyl jasmonate (MeJA), ethylene (ETH), and abscisic acid (ABA) since these phytohormones-responsive CAREs were detected frequently in putative promoter regions. We found 22 light-responsive elements located in all BnaCOBLs. These discoveries indicated that BnaCOBLs could be regulated by stress-related, phytohormone-responsive, and light-induced transcription factors.

**Tissue specificity of expression of BnaCOBLs in rapeseed**

The function of COBLs has been reported in root, flower, stem, and fruit skin of Arabidopsis, rice, maize, and tomato. We collected the RNA-seq data from the Sequence Read Archive...
(SRA) to examine the tissue-specific expression of the 44 BnaCOBLs. These tissues include stem, leaf, root, flower, stamen, ovule, pistil, silique, sepal, pericarp, blossomy pistil, and wilting pistil. The TPM values are listed in the S5 Table.

The expression profiling (Fig 6) in the various tissues demonstrated that BnaCOBLs participated in biological processes in all examined tissues, especially the COBLs of the subgroups the AtCOB, AtCOBL7, and AtCOBL8, whereas the COBLs of the subgroups AtCOBL1, AtCOBL10, and AtCOBL11 were specifically expressed in floral organs. Even individual genes, such as BnaCOBL6 and BnaCOBL25, were expressed in the ovule. The expression levels in tissues were mostly conserved in the intra orthologous subgroups but were different in inter orthologous subgroups. The eight members (BnaCOBL5, BnaCOBL11, BnaCOBL14 BnaCOBL31, BnaCOBL37, BnaCOBL42, BnaCOBL43, and BnaCOBL44) were found not to be
expressed in any tissues. These expression characteristics of BnaCOBLs implied that specific subgroup members have different functions in different organs.

Considering the reported brittle culm mutants and the importance of stem stress resistance, we further compared the expression levels of BnaCOBLs in stems with different SBR levels. All three AtCOB subgroup members BnaCOBL9, BnaCOBL35, and BnaCOBL41 were most active in the stem compared with other subgroups (Fig 6) and expressed at higher levels in the High-SBR one. Among these three genes, BnaCOBL41 was expressed highest. In contrast, the AtCOBL4 subgroup members BnaCOBL7, BnaCOBL26, BnaCOBL19, and BnaCOBL24, which were found to be involved in stem breaking in cereal crops, were weakly expressed in the High-SBR stem.

To confirm these results, we selected the above three AtCOB subgroup members BnaCOBL9, BnaCOBL35, and BnaCOBL41, and two AtCOBL4 subgroup genes BnaCOBL19, and BnaCOBL24, to quantify their expression with qRT-PCR in taproots, lateral roots, flower buds, leaves, upper and lower stems of rapeseed (ZS11) at the flowering stage. The amplification...
curves showed that the average expression level of the AtCOB subgroup genes was higher in the stem than that of the AtCOBL members (Fig 7).

Furthermore, the results of qRT-PCR on various parts of stems implied that the genes Bna-COBL9, Bna-COBL35, and Bna-COBL41 have different functions not only at different developmental stages but also in different internodes.

Discussion

In this study, we identified 44 COBRA-like genes in rapeseed and analyzed their phylogenetic relationships, chromosome locations, domain composition, and putative cis-elements. Together with the tissue specific expression patterns, these characters were differentiated by subgroups which were orthologs from different COBLs of Arabidopsis.

The COBRA family has been reported in many species in the plant kingdom, even in the moss Physcomitrella patens [5]. This family has already emerged in the ancestor of Arabidopsis. In Arabidopsis, there are 12 AtCOBLs, and segmental duplication contributed to AtCOBLs, as two pairs of duplication had been identified (AtCOBL2 and AtCOBL3, AtCOBL1 and AtCOBL4) [5].

We identified 44 Bna-COBLs, 20 Bra-COBLs, and 23 Bol-COBLs in the genomes of the allotetraploid Brassica napus and its diploid progenitor species B. rapa and B. oleracea, respectively.
Brassica evolved from a Brassiceae lineage-specific whole genome triplication (WGT) [19] after diverged from a common ancestor with Arabidopsis about 20 million years ago [50, 51]. After WGT the number of BraCOBLs and BolCOBLs almost doubled compared to the number of AtCOBLs. However, the number of BnaCOBLs was close to the sum of the number of BraCOBLs and BolCOBLs after allopolyploidization. Furthermore, BnaCOBLs are highly syntenic to [52–55] and conserved in gene clusters of BraCOBLs and BolCOBLs (Fig 1). We propose that whole-genome triplication event contributed to the expansion of BnaCOBLs.

The expression profiling demonstrated expression patterns of BnaCOBLs in twelve tissues (Fig 6). As the stem with reinforced mechanical strength showed higher resistance to lodging and pathogen attack [13, 38]. We concentrated on their expression levels in stems with different breaking resistance and found all three AtCOB subgroup members BnaCOBL9, BnaCOBL35, and BnaCOBL41 were expressed at higher levels in the High-SBR stem than in the Low-SBR one. BnaCOBL9 is located near the lodging coefficient QTL on A03, and BnaCOBL41 is located 300 kb upstream of the breaking force QTL on C09 in rapeseed [56]. Both BnaCOBL9 and BnaCOBL35 are reported to be hub genes with some CesA in a co-expression module, which was predicted to be relevant to cellulose biosynthesis [38]. We postulate that the AtCOB subgroup BnaCOBLs may play a role in the formation of stem strength in rapeseed.

Contrary to the expectation, the cloned AtCOBL4 subgroup COBL genes such as BC1 in rice, BK2 in maize, which were shown to be associated with the stem-breaking resistance in the grass family [57], were weakly expressed in High-SBR stems of rapeseed, which indicates AtCOBL4 subgroup members are not involved in the formation of stem strength in rapeseed. None of all AtCOBL4 subgroup members maintained all core motifs of COBLs (Fig 4), whereas members in this subgroup of B. rapa and B. oleracea were complete (S2 Fig). This structural change brought about alternative splicing variants, that is, exon skipping (BnaCOBL7 and BnaCOBL26) or intron-retention (BnaCOBL19 and BnaCOBL24). These splicing variants were confirmed in the reported rapeseed genomes [18]. Whether do structural change and alternative splicing cause neofunctionalization and/or subfunctionalization of AtCOBL4 subgroup members in rapeseed is worth more studies.

Supporting information

**S1 Fig. Sequence alignment of the COBRA domain of COBL proteins in Arabidopsis, rice, corn and rapeseed.** BC1_rice_Japo (AAQ56120.1) and BC1_rice_Indi (AAQ56121.1) represent the Brittle Culm1 protein in Oryza sativa subsp. Indica and Oryza sativa subsp. Japonica respectively. BK2 (ABJ99754.1) encodes brittle_stalk-2 protein in corn.

**S2 Fig. Sequence alignment of AtCOB and AtCOBL4 orthologous subgroup proteins in Arabidopsis, B. rapa, B. oleracea and B. napus.** The blue square brackets contain the COBL gene across four B. rapa genomes; The green square brackets contain the COBL gene across three B. oleracea genomes; The purple-red square brackets contain the COBL gene across nine B. napus genomes. Conservative domains are in the rectangle. "M1", "M2", "M3" and "M5" are the crucial sites that were verified by mutants to BC1 of rice; "M4" point at a transposon insertion site in BK2 of corn.

**S1 Table. Primer sequences designed for qRT-PCR of selected BnaCOBLs.**
S2 Table. The basic information concerning COBLs in *B. napus*, *B. rapa*, and *B. oleracea*. (XLSX)

S3 Table. Orthologous COBL gene pairs between *B. napus* with *B. rapa* and *B. oleracea*. (XLSX)

S4 Table. *Cis*-acting regulatory elements related to stress, light and phytohormone responsiveness in the promoter region of *BnaCOBL* genes. (XLSX)

S5 Table. TPM values of *BnaCOBLs* in different tissues and stems with distinct SBR. (XLSX)

**Author Contributions**

**Conceptualization:** Hao Chen, Zhongsong Liu.

**Data curation:** Qian Yang.

**Formal analysis:** Qian Yang.

**Funding acquisition:** Hao Chen.

**Investigation:** Qian Yang.

**Methodology:** Hao Chen.

**Resources:** Zhongsong Liu.

**Software:** Shan Wang.

**Supervision:** Qian Yang.

**Validation:** Qian Yang, Liang You.

**Visualization:** Qian Yang.

**Writing – original draft:** Qian Yang.

**Writing – review & editing:** Fangying Liu, Zhongsong Liu.

**References**

1. Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GH, Schindelman G, et al. *COBRA*, an Arabidopsis extracellular glycosyl-phosphatidylinositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell*. 2005; 17(6):1749–63. https://doi.org/10.1105/tpc.105.031732 PMID: 15849274.

2. Brady SM, Song S, Dhugga KS, Rafalski JA, Benfey PN. Combining expression and comparative evolutionary analysis. The COBRA gene family. *Plant Physiol*. 2007; 143(1):172–87. https://doi.org/10.1104/pp.106.087262 PMID: 17098858.

3. Green PB. Organogenesis-A Biophysical View. *Annu Rev Plant Physiol*. 1980; 31(1):51–82. https://doi.org/10.1146/annurev.pp.31.060180.000411

4. Schindelman G, Morikami A, Jung J, Baskin TI, Carpita NC, Derbyshire P, et al. COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. *Genes Dev*. 2001; 15(9):1115–27. https://doi.org/10.1101/gad.879101 PMID: 11331607.

5. Roudier F, Schindelman G, DeSalle R, Benfey PN. The COBRA family of putative GPI-anchored proteins in *Arabidopsis*. A new fellowship in expansion. *Plant Physiol*. 2002; 130(2):538–48. https://doi.org/10.1104/pp.007468 PMID: 12376623.

6. Li Y, Qian Q, Zhou Y, Yan M, Sun L, Zhang M, et al. *BRITTLE CULM1*, which encodes a COBRA-like protein, affects the mechanical properties of rice plants. *Plant Cell*. 2003; 15(9):2020–31. https://doi.org/10.1105/tpc.011775 PMID: 12953108.
7. Li P, Liu Y, Tan W, Chen J, Zhu M, Lv Y, et al. Brittle Culm 1 encodes a COBRA-Like protein involved in secondary cell wall cellulose biosynthesis in sorghum. Plant Cell Physiol. 2019; 60(4):788–801. [https://doi.org/10.1093/pcp/pcy242] PMID: 30590744.

8. Cao Y, Tang X, Giovannoni J, Xiao F, Liu Y. Functional characterization of a tomato COBRA-like gene functioning in fruit development and ripening. BMC Plant Biol. 2012; 12:211. [https://doi.org/10.1186/1471-2229-12-211] PMID: 23140186.

9. Ye X, Kang BG, Osburn LD, Cheng ZM. The COBRA gene family in Populus and gene expression in vegetative organs and in response to hormones and environmental stresses. Plant Growth Regul. 2009; 58(2):211–23. [https://doi.org/10.1007/s10725-009-9369-9] WOS:000265826900010.

10. Sangi S, Araujo PM, Coelho FS, Gazara RK, Almeida-Silva F, Venancio TM, et al. Genome-wide analysis of the COBRA-Like gene family supports gene expansion through Whole-Genome Duplication in soybean (Glycine max). Plants (Basel). 2021; 10(1). [https://doi.org/10.3390/plants1010167] PMID: 33467151.

11. Niu E, Shang X, Cheng C, Bao J, Zeng Y, Cai C, et al. Comprehensive analysis of the COBRA-Like (COBL) gene family in Gossypium identifies two COBLs potentially associated with fiber quality. PLoS One. 2015; 10(12):e0145725. [https://doi.org/10.1371/journal.pone.0145725] PMID: 26710066.

12. Li S, Ge FR, Xu M, Zhao XY, Huang GQ, Zhou LZ, et al. Genome-wide analysis of the COBRA-Like gene family supports gene expansion through Whole-Genome Duplication in oilseed rape (Brassica napus). BMC Plant Biol. 2014; 14(1):2923–36. [https://doi.org/10.1039/jxb/erl052] PMID: 16873454.

13. Ko JH, Kim JH, Jayanty SS, Howe GA, Han KH. Loss of function of COBRA, a determinant of oriented cell expansion, invokes cellular defense responses in Arabidopsis thaliana. J Exp Bot. 2006; 57(12):2923–36. [https://doi.org/10.1093/jxb/er052] PMID: 16873454.

14. Ching A, Dhugga KS, Appenzeller L, Meeley R, Bourett TM, Howard RJ, et al. Brittle stalk 2 encodes a putative glycosylphosphatidylinositol-anchored protein that affects mechanical strength of maize tissues by altering the composition and structure of secondary cell walls. Planta. 2006; 224(5):1174–84. [https://doi.org/10.1007/s00425-006-0299-8] PMID: 16752131.

15. Deng Q, Kong Z, Wu X, Ma S, Yuan Y, Jia H, et al. Cloning of a COBL gene determining brittleness in diploid wheat using a MapRseq approach. Plant Sci. 2019; 285:141–50. [https://doi.org/10.1016/j.plantsci.2019.05.011] PMID: 31203879.

16. Dai XR, Gao XQ, Chen GH, Tang LL, Wang H, Zhang XS. ABNORMAL POLLEN TUBE GUIDANCE1, an endoplasmic reticulum-localized mannosyltransferase homolog of GLYCOSYLPHOSPHATIDYLINOSITOL-GLYCOSYLPHOSPHATIDYLINOSITOL GLYCAN ANCHOR BIOSYNTHESIS B in human, is required for Arabidopsis pollen tube micropylar guidance and embryo development. Plant Physiol. 2014; 165(4):1544–56. [https://doi.org/10.1104/pp.114.236133] PMID: 24963069.

17. Gao H, Zhang Y, Wang W, Zhao K, Liu C, Bai L, et al. Two membrane-anchored aspartic proteases contribute to pollen and ovule development. Plant Physiol. 2017; 173(1):219–39. [https://doi.org/10.1104/pp.116.01719] PMID: 27872247.

18. Song JM, Liu DX, Xie WZ, Yang Z, Guo L, Liu K, et al. BnPIR: Brassica napus pan-genome information resource for 1689 accessions. Plant Biotechnol J. 2021; 19(3):412–4. [https://doi.org/10.1111/pbi.13491] PMID: 33068485.

19. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 2014; 345(6199):950–3. [https://doi.org/10.1126/science.1253435] PMID: 25146293.

20. BnPIR [cited 2020 10–23], Available from: http://cbi.hzau.edu.cn/bnapus/.

21. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004; 32(Web Server issue):W20–5. [https://doi.org/10.1093/nar/gkh435] PMID: 15215342.

22. Marchler-Bauer A, Bo Y, Han L, He J, Lancyzycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017; 45(D1):D200–D3. [https://doi.org/10.1093/nar/gkw1129] PMID: 27899674.

23. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPaSy server. Methods Mol Biol. 1999; 112:531–52. [https://doi.org/10.1385/1-59259-584-7:531] PMID: 10027273.

24. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools API in 2019. Nucleic Acids Res. 2019; 47(W1):W836–W41. [https://doi.org/10.1093/nar/gkz688] PMID: 30976793.

25. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32(1):268–74. [https://doi.org/10.1093/molbev/msu006] PMID: 25371430.
26. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002; 93(1):77–8. https://doi.org/10.1093/jhered/93.1.77 PMID: 12011185.

27. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCoScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012; 40(7):e49. https://doi.org/10.1093/nar/gkr1293 PMID: 22217600.

28. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020; 13(6):1194–202. https://doi.org/10.1016/j.molp.2020.06.009 PMID: 30778233.

29. Pugalenthi G, Nithya V, Chou KC, Archunan G. Nglyc: A random forest method for prediction of N-glycosylation sites in eukaryotic protein sequence. Protein Pept Lett. 2020; 27(3):178–86. https://doi.org/10.2174/0929866526666191002111404 PMID: 31577193.

30. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019; 37(4):420–3. https://doi.org/10.1038/s41587-019-0036-z PMID: 30778233.

31. Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in proprotein sequences. J Mol Biol. 1999; 292(3):741–58. https://doi.org/10.1006/jmbi.1999.3069 PMID: 10497036.

32. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. 2005:571–607.

33. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982; 157(1):105–32. https://doi.org/10.1016/0022-2836(82)90515-0 PMID: 7108955.

34. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018; 35(6):1547–9. https://doi.org/10.1093/molbev/msy096 PMID: 32585190.

35. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002; 30(1):325–7. https://doi.org/10.1093/nar/30.1.325 PMID: 11752327.

36. Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, et al. Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics. 2007; 8:67. https://doi.org/10.1186/1471-2164-8-67 PMID: 17346352.

37. Sun F, Fan G, Hu Q, Zhou Y, Guan M, Tong C, et al. The high-quality genome of Brassica napus cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. Plant J. 2017; 92(3):452–68. https://doi.org/10.1111/tpj.13669 PMID: 28849613.

38. Li H, Cheng X, Zhang L, Hu J, Zhang F, Chen B, et al. An Integration of Genome-wide association study and gene co-expression network analysis identifies candidate genes of stem lodging-related traits in Brassica napus. Front Plant Sci. 2018; 9:796. https://doi.org/10.3389/fpls.2018.00796 PMID: 29946333.

39. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 2012; 7(2):e30619. https://doi.org/10.1371/journal.pone.0030619 PMID: 22312429.

40. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12(1):1–16. https://doi.org/10.1186/1471-2105-12-323 WOS:000294361700001. PMID: 21816040.

41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886.

42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25(4):402–8. https://doi.org/10.1093/meth/2001.1262 PMID: 11846609.

43. Graham GJ. Tandem genes and clustered genes. J Theor Biol. 1995; 175(1):71–87. https://doi.org/10.1006/jtbi.1995.0122 PMID: 7564393.

44. Overbeek R, Fonstein M, D’Souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci U S A. 1999; 96(6):2896–901. https://doi.org/10.1073/pnas.96.6.2896 PMID: 10077608.

45. Foflonker F, Blaby-Haas CE. Colocality to cofunctionality: eukaryotic gene neighborhoods as a resource for function discovery. Mol Biol Evol. 2021; 38(2):650–62. https://doi.org/10.1093/molbev/msaa221 PMID: 32886760.

46. Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, et al. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biol. 2014; 15(6):R77. https://doi.org/10.1186/gb-2014-15-6-r77 PMID: 24916971.
47. Korkuc P, Schippers JH, Walther D. Characterization and identification of cis-regulatory elements in *Arabidopsis* based on single-nucleotide polymorphism information. Plant Physiol. 2014; 164(1):181–200. https://doi.org/10.1104/pp.113.229716 PMID: 24204023.

48. Li Y, Chen CY, Kaye AM, Wasserman WW. The identification of cis-regulatory elements: A review from a machine learning perspective. Biosystems. 2015; 138:6–17. https://doi.org/10.1016/j.biosystems.2015.10.002 PMID: 26499213.

49. Porto MS, Pinheiro MP, Batista VG, dos Santos RC, Filho Pde A, de Lima LM. Plant promoters: an approach of structure and function. Mol Biotechnol. 2014; 56(1):38–49. https://doi.org/10.1007/s12033-013-9713-1 PMID: 24122284.

50. Wang W, Guan R, Liu X, Zhang H, Song B, Xu Q, et al. Chromosome level comparative analysis of *Brassica genomes*. Plant Mol Biol. 2019; 99(3):237–49. https://doi.org/10.1007/s11103-018-0814-x PMID: 30632049.

51. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, et al. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet. 2011; 43(5):476–81. https://doi.org/10.1038/ng.807 PMID: 21478890.

52. Wang R, Li M, Wu X, Wang J. The gene structure and expression level changes of the GH3 gene family in *Brassica napus* relative to its diploid ancestors. Genes (Basel). 2019; 10(1). https://doi.org/10.3390/genes10010058 PMID: 30658516.

53. Ke YZ, Wu YW, Zhou HJ, Chen P, Wang MM, Liu MM, et al. Genome-wide survey of the bHLH super gene family in *Brassica napus*. BMC Plant Biol. 2020; 20(1):115. https://doi.org/10.1186/s12870-020-2315-8 PMID: 32171243.

54. Zhou J, Zhou HJ, Chen P, Zhang LL, Zhu JT, Li PF, et al. Genome-wide survey and expression analysis of the KT/HAK/KUP family in *Brassica napus* and its potential roles in the response to K(+) deficiency. Int J Mol Sci. 2020; 21(24). https://doi.org/10.3390/ijms21249487 PMID: 33322211.

55. Wang T, Hu J, Ma X, Li C, Yang Q, Feng S, et al. Identification, evolution and expression analyses of whole genome-wide TLP gene family in *Brassica napus*. BMC Genomics. 2020; 21(1):264. https://doi.org/10.1186/s12864-020-06678-x PMID: 32228446.

56. Wei L, Jian H, Lu K, Yin N, Wang J, Duan X, et al. Genetic and transcriptomic analyses of lignin- and lodging-related traits in *Brassica napus*. Theor Appl Genet. 2017; 130(9):1961–73. https://doi.org/10.1007/s00122-017-2937-x PMID: 28634809.

57. Carpita NC. Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol. 1996; 47:445–76. https://doi.org/10.1146/annurev.arplant.47.1.445 PMID: 15012297.