Genome-wide Identification and Transcriptional Profiling of New Cobra-like Genes in Bread Wheat (*Triticum Aestivum*)

Muhammad Zaheer

COMSATS University Islamabad

Shoaib Ur Rehman

MNSUA: Muhammad Nawaz Shareef University of Agriculture

Sultan Habibullah Khan

University of Agriculture Faisalabad

Shahmeer Shahid

MNSUA: Muhammad Nawaz Shareef University of Agriculture

Awais Rasheed

Quaid-i-Azam University Islamabad: Quaid-i-Azam University

Rabia Naz

COMSATS University Islamabad

Muhammad Sajjad (✉ muhammad.sajjad@comsats.edu.pk)

COMSATS University Islamabad <https://orcid.org/0000-0002-1036-2510>

Research Article

Keywords: Wheat, Relative expressions, PCR amplification, Cobra like genes, drought, GPI

Posted Date: August 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-796690/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The *COBL* gene encodes a plant-specific glycosylphosphatidylinositol (GPI)-anchored protein. Recently identified COBRA genes supposed as a key regulator of the orientation of cell expansion in the root indicating that Cobra gene family members are likely to be important new players at the plasma membrane-cell wall interface.

Methods and Results: By performing a database search and domain prediction, we identified five genes named as *TaCOBL 1*, *TaCOBL 2*, *TaCOBL 3*, *TaCOBL 4* and *TaCOBL 5*, and selected for further analysis. Chromosomal locations of each gene were drawn on karyotype. Structure of genes, promoter analysis and phylogenetic analysis were done using bioinformatics tools and databases. Set of novel SNPs were also predicted. Gene ontologies were analyzed, and pathways involving cobra genes were predicted. Whole genes of 3kb to 4kb were successfully amplified. Five set of primers were designed targeting *TaCOBL 1*, *TaCOBL 2*, *TaCOBL 3*, *TaCOBL 4* and *TaCOBL 5* genes. Expression of *COBL* genes were checked in root and shoot by using qPCR. The qRT-PCR revealed that expression *TaCOBL* genes can be regulated under abiotic stress such drought stress.

Conclusion: The comprehensive annotation and expression profiling of *COBL* genes performed in this study enhanced our understanding and these genes were found to play a significant role in drought stress. Our findings lay the groundwork for further functional investigation of *COBL* genes mechanism.

Introduction

Transcription factors (TFs) regulates gene expression by binding to certain cis-elements, which are important for plant growth and stress responses. The availability of plant genomes enables the estimation of genome-wide evolutionary footprints and the subsequent inference of plant functional regulatory maps. Systematic identification of TFs, regulatory elements, and their functional relationships would substantially aid future mechanistic research [1, 2].

The Cobra-Like (*COBL*) gene encodes a plant-specific glycosylphosphatidylinositol (GPI)-anchored protein with an ω -attachment site at the C terminal. The GPI contains, hydrophilic and current-controlled voltage source domain, a potential N-glycosylation site, N-terminal secretion signal sequence, and a predicted cellulose binding site [1]. The COBL proteins are initially secreted in Golgi apparatus and later on it is transported to peripheral region where it synthesizes cell wall [2]. Cellulose is a key component in cell wall formation and cell expansion [3]. Many other genes are also responsible for the biosynthesis of cellulose like *FLA*, *NAC*, and *MYB*. COBRA is one of the genes that has been linked to cellulose crystallization and is substantially co-expressed with the cellulose synthase complex. These genes bind to individual glucan chains as well as cellulose and are in plasma membrane. The COBRA protein, however, is devoid of recognized patterns save for a carbohydrate binding domain, and the method by which COBRA aids cellulose production is unknown [4].

Various studies revealed that *COBL* genes are involved in the formation of micro-fibrils and cellulose crystallinity in model plants like *Oryza sativa*, *Arabidopsis thaliana*, *Zea mays* and *Solanum lycopersicum* [5]. The *COBL* gene family has two subgroups in *A. thaliana*. *AtCOBL* belongs to a multigene family consisting of 12 members and two subgroups in *Arabidopsis*, the second of which is distinguished by an additional N-terminal 170 amino acid [6]. There are eleven members in *O. sativa* and nine in *Z. mays*. *OsBS1L6* is involved in ripening and enlargement of endosperm [7]. In case of seed ripening and processing, cell wall play vital role as it controls the water absorbing abilities of grains. Many genes like *OsCEA5*, *OsCEA6* and *OsCEA8* are related to cell wall synthesis which showed down regulation during germination of rice.

In *A. thaliana*, the mutant *cobra* causes abnormalities in polar longitudinal growth of root cells [17–18], while in *O. sativa*, the mutant *brittle culm 1 (bc1)* similarly exhibits a characteristic dwarf phenotype [19]. The *cobl4* mutation in *A. thaliana*, the *brittle culm 1* mutant in *O. sativa*, and the *brittle stalk 2* mutation in *Z. mays* all alter the mechanical strength of vascular bundles and significantly reduce the cellulose content [20–22]. Additionally, Liu et al. [23] reported that COBL proteins can regulate the structure of cellulose by attaching to it and influencing the crystallinity of micro fibrils.

Wheat (*Triticum aestivum* L.) is a staple food crop which is grown worldwide for the human consumption. The high adaptability to the atmosphere and high yield potential is the basic behind its global diversity. The productivity of this main crop is being reduced due to different environmental challenges. Therefore, it is necessary to find and characterize the genes responsible to combat with environmental stresses. Genome sequencing of bread wheat enables in-depth analysis of the *COBL* gene families as some other gene families have been characterized in this species [8–12]. The purpose of this study was to characterize COBL gene family in bread wheat. Phylogenetic analyses, chromosomal distribution, cis element analysis, sub-cellular localization, exon/intron structure analyses were carried out. Additionally, we determined expression patterns of three *TaCOBL* genes in different tissues and in response to drought. The findings of this study will be helpful to further dissect the molecular function of *COBL* genes in bread wheat.

Materials And Methods

Plant materials and Growth conditions

A set of 48 wheat genotypes were used to amplify TaCOBL genes. For gene expression analysis a drought tolerant (Inqalab-91) and drought susceptible variety (Seher-06) was used. For expression analysis the two varieties were grown in pots in two sets. One set was applied water to maintain 70% field capacity and the second set was applied with limited water to maintain 50% field capacity.

Database search, sequence retrieval and structural analysis

To predict the COBL candidate genes, genome wide identification were performed in hexaploid wheat genome (IWGS Ref Seq V1). Template genes were selected from the model plants like *A. thaliana*, *O. Sativa* and *Zea mays*. The protein sequences of *A. thaliana* and *O. sativa* were used as reference sequences to search against wheat database by using BLAST_P program. Parameters were set and optimized, where E value were 0.01 and scoring matrix was set

BLOSUM90. Candidate genes were selected on the basis of maximum similarity with low *E* - values. Genes were further confirmed by their conserved domains and motif using NCBI CD search tool (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Protein sequences were retrieved from Uniport database in FASTA format. (<https://www.uniprot.org/>). Wheat G-genomic DNA sequences were downloaded for further analysis (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Sequences are extended up to 2 kb from both ends in order to have the promoter regions, regulatory elements, TATA boxes, enhancers and Untranslated regions (UTRs). The biophysical properties were predicted with the ExPASy ProtParam tool (<http://us.expasy.org/tools/protparam.html>). Glycosylation sites were predicted by NetNGlyc. (<http://www.cbs.dtu.dk/services>). Protein sequences were retrieved from NCBI were aligned by using Multiple Sequence Alignment (MSA) tool DNA star software and analysis was done using MUSCLE tool of MEGA X software. Multiple sequences were edited and refined using BioEdit software in order to increase similarity [13].

Subcellular localization and Karyotyping

Sub-cellular locations were predicted by Softberry (<http://www.softberry.com/>). Karyotype was constructed clearly displaying the respective loci of genes and their homeologs on respective sub-genomes A, B and D and were predicted using plant ensemble (https://plants.ensembl.org/Triticum_aestivum/Location/Genome).

Exon-Intron structure, Gene Site and phylogenetic analysis

The structures of COBL genes were determined using the Gene Structure Display Server (GSDS) (<http://gsds.gao-lab.org/index.php>) [14]. Genes sites were predicted using online FGENESH tool of softberry showing TSS sites, Number of intron and exon and poly A tail sites. The evolutionary history was inferred using the Neighbor-Joining method. An unrooted Neighbor-Joining phylogenetic tree was constructed with Bootstrap method keeping 1000 repetitions using MEGA X software [15]. The Bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% Bootstrap replicates are collapsed. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. This analysis involved 82 amino acid sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There was a total of 173 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

cis -acting elements analysis and detection of SNPs

In order to examine the promoter region, upstream region up to 2 kb of Cobra gene family was retrieved from wheat genomic database of ensemble plant (https://plants.ensembl.org/Triticum_aestivum/Info/Index). Presence of Pribnow boxes that is TATA box (TATAAT) and CAAT box (GCCAAT) was checked manually. Presence of CpG Island was investigated using SMS web (http://bioinformatics.org/sms2/cpg_islands.html). Drought responsive *cis* regulatory elements were determined by online PLACE Library tool (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>) [16]. Each COBL gene was searched against wheat genome (IWGSC refseq v1.0).

RNA Isolation and PCR amplification

RNA was extracted using RNeasy Pure Plant Kit (QIAGEN, China). qPCR was conducted using Light Cycler 480 system (Roche Diagnostics), with an initial denaturation step at 94 °C for 5 min, followed by 45 cycles of denaturation at 94 °C for 10 s, annealing at 58 °C for 10 s, and extension at 72 °C for 20 s). Wheat actin was used as internal control for normalization as a reference primer. SYBR Green I Master Mix Roche Diagnostics, Indianapolis, IN, USA was used in the reaction mixture according to the manufacturer's instructions. As the Exonic regions of all the respective homeologs were almost the same in wheat sub-genomes, hence gene-specific quantitative polymerase chain reaction (qPCR) conserved primers for *COBL* genes were designed. The relative transcription level for each gene was calculated using $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

Results

Database search, sequence retrieval and structural analysis

In this study, *A. thaliana* and *O. sativa* Cobra genes was taken as reference and blasted against wheat genome. As a result, five set of orthologs were found in wheat having maximum similarity and maximum coverage. All the sets of genes had *COBRA* conserved domain as shown in Table 1 that reconfirmed the orthologs genes. Biophysical properties were retrieved such as genes position, molecular weight, and number of amino acids, Iso-electric points, and Glycosylation sites. All the information was collected using different bio-informatics tools. The molecular weight all the cobra proteins was in range of 50–75 KDa. The number of residues were in between 429 to 461. The iso- electric points of all these set of cobra protein were in average of about 8. The number of Glycosylation sites were also in average of 8.

Subcellular localization and Karyotyping

Plasma membrane was the site of localization of the cobra protein synthesized from Cobra genes in wheat. Karyotype was constructed by searching all the genes in wheat genome by their respective positions. The predicted gene were present on chromosomes 2, 4, 5, 6 and 7 as shown in Fig. 1. Each gene has their respective homeologs were present on sub-genome A, B and D. we found that the cytogenic location of each homeologs were not adjacent to each other, In case of chromosome 2 and 4, they were away from each other's.

Table.1: Sets of genes and their orthologs having COBRA conserved domain with similarity and e value

Query SEq.	Orthologs	Position	Similarity	Coverage	E-Value	CD-Domain
<i>AtCOBL1</i>	<i>TaCOBL1A</i>	2A: 133673912:133678910	88%	90%	0.01	Present
	<i>TaCOBL1B</i>	2B: 575934027:575937341	83%	91%	0.5	Present
	<i>TaCOBL1D</i>	2D: 491498233:491501692	91%	97%	0.2	Present
<i>AtCOBL2</i>	<i>TaCOBL2A</i>	6A: 600402485:600407022	76%	89%	0.3	Present
	<i>TaCOBL2B</i>	6B: 690558034:690562573	82%	90%	0.1	Present
	<i>TaCOBL2D</i>	6D: 454229171:454234205	74%	89%	0.4	Present
<i>AtCOBL3</i>	<i>TaCOBL3A</i>	4A: 540440598:540445318	81%	87%	0.01	Present
	<i>TaCOBL3B</i>	4B: 83131612:83136786	79%	90%	0.01	Present
	<i>TaCOBL3D</i>	4D: 56705992:56710599	85%	97%	0.4	Present
<i>AtCOBL4</i>	<i>TaCOBL4A</i>	5A: 588374977:588379739	88%	91%	0.3	Present
	<i>TaCOBL4B</i>	5B: 574674518:574679254	87%	88%	0.2	Present
	<i>TaCOBL4D</i>	5D: 467600940–467604428	80%	85%	0.1	Present
<i>AtCOBL5</i>	<i>TaCOBL5A</i>	7A: 89071769:89075439	89%	90%	0.5	Present
	<i>TaCOBL5B</i>	7B: 36506027:36509655	95%	93%	0.01	Present
	<i>TaCOBL5D</i>	7D: 87748276:87751869	91%	90%	0.01	Present

Exon-Intron structure, Gene Site, phylogenetic tree and domain analysis

Gene structures were drawn using online GSDS server. The number of introns, exons and length of UTRs upstream and downstream can clearly be seen in the structure as shown in Fig. 2. Twelve set of genes has similar structural pattern having same of number of exon and exon as well as their chromosomal locations but in case of *TaCobra 5*, there were only two exon and only one intron has highly different structural pattern. This also has shorter length as compare to other set of genes. Gene structures was also retrieved from softberry database in order to zoom out further and deep understanding and analysis. In these structures, Transcription starts sites (TSS), Transcription stop site, all internal exons and poly A tail can be seen clearly.

The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. This analysis involved 82 amino acid sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were total of 173 positions in the final dataset. Evolutionary analyses were conducted in MEGA. The similarity index of Cobra protein with neighboring Cobra protein present in Arabidopsis, rice and maize were in range of 62–95% which shows their high confident similarity among them and their ancestral relations. Branches of clades of phylogenetic tree shows that they are all originated from a common ancestor.

Mack et al 2001, defined conserved domain as computational biologists define conserved domains based on recurring sequence patterns or motifs. Conserved domain of Cobra genes is COBRA (Accession no. pfam04833, Pssm-ID: 368147) which is consist of 167 amino acids. The residual sEq. of COBRA domain is:

YVAQVTIENYQPYRHIDNPGWNLGWEWAKKEFIWSMKGAQATDQGDCSYFKgLFISIPHCCEKRPTIVDLPPGAPYDTQVGNCCRGVLLPPTQDPSKSVSAFQMVGKAPF

Cis -acting elements analysis and detection of SNPs

First of all, transcription start sites of all the genes were retrieved and then extended back up to 2kb. This region was supposed to be the promoter region of the respective gene. TATA boxes and CAAT boxes were almost present in all the gen as shown in Table 3. Five set of *cis* acting elements responsive to drought stress were screened from previous literature as shown in Table 2. Among these elements, three elements WRKY, ABRE and DRE was repeatedly present in most of the genes as shown in Table 3. The sequence of desired DNA was search in genome of wheat (*Triticum aestivum*). Nineteen set of novel SNPs were detected. Amongst these, three SNPs were Intronic and three SNPs set were Exonic. Nine SNPs were present in upstream region while three set of SNPs were present in downstream region Table 4.

Table 2

Cis-Element	Conserved Cis Motif	References
ABRE	ACGTG/ACGT	(Hattori T. et al 2002)
DRE	CCGAC/RYCGAC	(Xue, G.,.2002.)
MYC	CANNTG	(Abe H, et al 2003).
MYB	CAACNA/CAACNG/TAACNG	(Abe H. et al 1997)
WRKY	TGAC	(Pan et al 2009)

Table 3

Gene IDs	Cis Acting Elements	Nos.
TraesCS2A02G175100	WRKY, ABRE	2
TraesCS2B02G406700	ABRE, DRE, WRKY	3
TraesCS2D02G386400	DRE, WRKY	2
TraesCS6A02G379800	ABRE, DRE, WRKY	3
TraesCS6B02G418400	ABRE, DRE, WRKY	3
TraesCS6D02G364400	DRE, WRKY	2
TraesCS4A02G231200	DRE, WRKY	2
TraesCS4B02G084800	DRE, WRKY,	2
TraesCS4D02G082700	WRKY ABRE	2
TraesCS5A02G392000	ABRE	1
TraesCS5B02G396900	WRKY	1
TraesCS5D02G401900	ABRE, DRE, WRKY	3
TraesCS7A02G136600	DRE	1
TraesCS7B02G037700	WRKY	1
TraesCS7D02G136500	DRE, WRKY	1

Table 4

Sr. No	Chr. No.	Positions	SNPs		Types
			Ref. Sequence	Query Sequence	
1.	chr2A	133674024	A	C	Downstream
2.	chr2A	133674066	A	G	Downstream
3.	chr2A	133678611	A	G	Upstream
4.	chr2A	133678612	G	A	Upstream
5.	chr2A	133678724	T	C	Upstream
6.	chr2A	133678798	C	G	Upstream
7.	chr2A	133678816	G	A	Upstream
8.	chr2A	133678859	G	C	Upstream
9.	chr2A	133678872	A	G	Upstream
10.	chr4B	83135305	T	C	Exonic
11.	chr4B	83135319	G	T	Exonic
12.	chr4B	83135516	T	C	Intronic
13.	chr4B	83136481	C	A	Downstream
14.	chr5A	588376735	C	T	Exonic
15.	chr5A	588377080	C	T	Intronic
16.	chr6A	600402608	G	A	Upstream
17.	chr6A	600402740	A	G	Upstream
18.	chr6A	600402753	G	C	Upstream
19.	chr6B	690561305	T	C	Intronic

Expression Profiling of TaCOBL genes

Expression profiling of gene helps in to understand its biological function. Hence we investigated the expression of *TaCOBL* genes in two wheat varieties In-qalab-91 and Seher – 06. *TaCOBL* predicted genes were successfully validated by PCR amplification by their respective designed primers. The desired and targeted product size were in range of 3 to 4kb, all the amplified PCR products were in range of 3 to 4 kb. All the genes were used to check the tissue specific transcript abundance levels by qRT-PCR in root and shoot. All the studied genes had relatively higher transcript level as compared to control for drought treatment. All the genes showed same pattern of expression except *TaCOBL 4* which shows relatively higher transcript level as compared to other genes under drought stress .Together with these observations, *TaCOBL* genes showed obvious resistance against drought stress advocating that these genes are possible candidates for breeding drought resistant wheat cultivar.

Discussion

Present grain yield patterns in major crops are insufficient to meet the needs of the world's around 9 billion people by 2050. This is complicated further by climate change, which is expected to exacerbate extreme weather events, and it is well understood that water stress inhibits crop plant growth, development, and productivity [1]. The current genomic landscape of crop plants has been transformed by next generation sequencing methods that provide information on an infinite number of manifolds. The availability of wheat genome assemblies [1] opens up new potential for more precise genome-wide analysis.

Extensive research have been conducted to determine the function and biological significance of COBL protein in many crop species. However, no thorough comparison of *COBL* genes is studied in wheat. In this study, we analyzed the *COBL* genes in wheat this study will prove beneficial in building the groundwork for future work

The structure of the exon-intron region is important and may be influenced by insertion/deletion events. Extensive studies have been reported considerable intron gain or loss during eukaryotic diversification. The examination of duplicated genes revealed a similar exon-intron pattern. Introns also play a significant influence in the evolution of diverse plant species [27]. We discovered that all *COBL* genes have the same number of introns except *COBL 5*, which has two introns, indicating that *COBL* protein in *Triticum aestivum* was recently evolved with fewer introns, supporting previous research that intron numbers decreased over time [28] and implying that newly evolved species have fewer introns than primitive species [27]. *TaCOBL* genes have an exceptionally comparable gene structure, indicating that these genes were substantially conserved. Additionally, introns were assumed to have acquired novel functions as a result of weak selection pressure and the presence of a greater number of introns. Subcellular localization could be important in defining COBL genes function. To explain

the evolutionary relationship of *COBL* genes in wheat, we constructed a phylogenetic tree including wheat, *A. thaliana* Z. mays, and *O.sativa*. The similarity index of Cobra protein with neighboring Cobra protein present in Arabidopsis, rice and maize were in range of 62–95% which shows their high confident similarity among them and their ancestral relations.

TaCOBL genes promoter regions comprised abscisic acid-responsive elements (ABRE), drought-responsive elements, and WRKY Cis-acting elements, indicating that these genes are regulated not only by abiotic stress but also by several hormones (Table 1.2). However, the frequency of stress-responsive Cis-elements is significantly greater than the other elements, showing their important function in abiotic stress. Subcellular localization could be important in defining COBL genes function.. The structure of the COBRA protein is extremely conserved. COBRALIKE proteins contain a signal peptide at the N terminus and a GPI anchoring motif at the C terminus, both of which are thought to target the protein into the plasma membrane's outer leaflet. It's difficult to say if COBRA genes that code for proteins that lack any of these patterns are functional.. The GPI anchor motif is frequently broken after the protein has been localized to the plasma membrane's outer leaflet (Butikofer et al., 2001). As a result, it's probable that COBRA's capacity to bind cellulose is critical for its correct localization at the plasma membrane-cell wall interface.

To identify the putative function of *TaCOBL* genes in regulating drought stress, we analyzed their expression in two wheat varieties at different tissues using qPCR.. The results of ours study revealed that *TaCOBL genes* have higher transcript abundance in Shoot and roots at seedling stage, showing the positive role, suggesting their function redundancy as well as function diversity that these genes may play role under osmotic stress .Expression pattern of all the COBL genes is quite similar in both varieties Among all the genes *TaCOBL 4* shows high transcript level as compared to rest of genes. The expression of *TaCOBL* genes can considerably be up regulated by abiotic stress such as PEG. Together with these outcomes, we conclude that *TaCOBL* genes might be instrumental for breeding drought tolerant wheat.

Conclusion

In summary, this study identified five non-redundant *COBL* genes in *Triticum aestivum*. Their classification; characterization in terms of gene structure and conserved domains; and comparative phylogenetic analyses all indicate that *COBL* genes are conserved across the plant species. It was suggested that paralogous genes expanded as a result of segmental duplications, whereas the codons underwent purifying selection. Three genes were shown to be up-regulated in stress conditions, implying that they play a role in drought stress. Overall, our findings will aid in the acceleration of COBL gene functional characterization, particularly in response to abiotic stress. After that, the candidate COBL genes can be used in stress breeding programs.

Declarations

Acknowledgements:

The authors would like to thank the Institute of Biotechnology of the Urmia University (Urmia, Iran) for hosting the wet lab research.

Funding:

This research received grant from Pakistan Science Foundation (PSF-MSRT II/Agr/P-COMSATS-Vehari 11).

Conflict of interest:

The authors declare that they have no competing interests.

Availability of data and material:

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability:

Not applicable.

References

1. Roudier F, Schindelman G, DeSalle R, Benfey PN (2002) The COBRA family of putative GPI-anchored proteins in Arabidopsis. A new fellowship in expansion. *Plant Physiol* 130:538–548
2. Brady SM, Song S, Dhugga KS, Rafalski JA, Benfey PN (2007) Combining expression and comparative evolutionary analysis. The COBRA gene family. *Plant physiology* 143:172–187
3. Niu E, Shang X, Cheng C, Bao J, Zeng Y, Cai C et al (2015) Comprehensive Analysis of the COBRA-Like (COBL) Gene Family in Gossypium Identifies Two COBL s Potentially Associated with Fiber Quality. *PLoS One* 10:e0145725
4. MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. *Plant J* 62:689–703

5. Cao Y, Tang X, Giovannoni J, Xiao F, Liu Y (2012) Functional characterization of a tomato COBRA-like gene functioning in fruit development and ripening. *BMC plant biology* 12:1–15
6. Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GH, Schindelman G et al (2005) "COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation". *Plant Cell* 17:1749–1763
7. Midorikawa K, Kuroda M, Yamashita H, Tamura T, Abe K, Asakura T (2019) *Oryza sativa* Brittle Culm 1-like 6 modulates β -glucan levels in the endosperm cell wall. *PLoS one* 14:e0217212
8. Appels R, Eversole K, Stein N, Feuillet C, Keller B, Rogers J, et al., "Shifting the limits in wheat research and breeding using a fully annotated reference genome," *Science*, vol. 361, 2018
9. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H et al (2017) "Wild emmer genome architecture and diversity elucidate wheat evolution and domestication". *Science* 357:93–97
10. Shoaib M, Yang W, Shan Q, Sun L, Wang D, Sajjad M, Li X, ·Sun J, Liu D, Zhan K, Zhang A (2020) TaCKX gene family, at large, is associated with thousand-grain weight and plant height in common wheat. *Theor Appl Genet* (2020). <https://doi.org/10.1007/s00122-020-03661-6>
11. Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H et al (2018) "Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*". *Nature* 557:424–428
12. Liu L, Shang-Guan K, Zhang B, Liu X, Yan M, Zhang L et al (2013) "Brittle Culm1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils". *PLoS Genet* 9:e1003704
13. Hall T, Biosciences I, Carlsbad C (2011) "BioEdit: an important software for molecular biology". *GERF Bull Biosci* 2:60–61
14. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G, "GSDS 2.0: an upgraded gene feature visualization server," *Bioinformatics*, vol. 31, pp. 1296–1297, 2015
15. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) "MEGA X: molecular evolutionary genetics analysis across computing platforms". *Molecular biology evolution* 35:1547
16. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y et al (2002) "PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences". *Nucleic acids research* 30:325–32717.
17. D. Ben-Tov, A. Idan-Molakandov, A. Hugger, I. Ben-Shlush, M. Günl, B. Yang, et al., "The role of COBRA-LIKE 2 function, as part of the complex network of interacting pathways regulating Arabidopsis seed mucilage polysaccharide matrix organization," *The Plant Journal*, vol. 94, pp. 497–512, 2018.
18. M.-T. Hauser, A. Morikami, and P. N. Benfey, "Conditional root expansion mutants of Arabidopsis," *Development*, vol. 121, pp. 1237–1252, 1995.
19. Dai XX, You CJ, Chen GX, Li XH, Zhang QF, Wu CY. OsBC1L4 encodes a COBRA-like protein that affects cellulose synthesis in rice. *Plant Mol Biol*. 2011; 75(4–5): 333–345. doi:10.1007/s11103-011- 9730. -z PMID: 21264494
20. Hochholdinger F, Wen TJ, Zimmermann R, Chimot-Marolle P, da Costa e Silva O, Bruce W, et al. The maize (*Zea mays* L.) roothairless 3 gene encodes a putative GPI-anchored, monocot-specific, COBRALike protein that significantly affects grain yield. *Plant J*. 2008; 54(5): 888–898. doi:10.1111/j.1365- 313. X.2008.03459.x PMID: 18298667
21. Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner SR. Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell*. 2005.; 17(8): 2281–2295. PMID: 15980264
22. Li YH, Qian Q, Zhou YH, Yan MX, Sun L, Zhang M, et al. BRITTLE CULM1, which encodes a COBRALike protein, affects the mechanical properties of rice plants. *Plant Cell*. 2003; 15(9):2020–2031. PMID: 12953108.
23. Ching A, Dhugga KS, Appenzeller L, Meeley R, Bourret 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):1174–1184. PMID: 16752131
24. Liu LF, Shang-Guan KK, Zhang BC, Liu XL, Yan MX, Zhang LJ, et al. Brittle Culm1, a COBRA-Like protein, functions in cellulose assembly through binding cellulose microfibrils, *PLoS Genet*. 2013; 9(8): e1003704. doi:10.1371/journal.pgen.1003704 PMID: 23990797

Figures

Figure 1

Karyotyping of candidate genes of COBRA gene family in wheat genome with respect to their positions.

Figure 2

Gene structure display showing no of intron-exons, upstream and downstream region along Transcription start sites (TSS), Transcription stop site, and poly A tail site.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [5.jpeg](#)