Genome-wide analysis of cellulose synthase (CESA) and cellulose synthase-like (CSL) proteins in *Cannabis sativa* L.

Hülya Sipahi[1], Samuel Haiden[2], Gerald Berkowitz[2]

[1] Department of Agricultural Biotechnology, Faculty of Agriculture, Eskişehir, University of Eskişehir Osmangazi, Türkiye
[2] Agricultural Biotechnology Laboratory, Department of Plant Science and Landscape Architecture, University of Connecticut, Storrs, CT 06269, USA

Corresponding Author: Hülya Sipahi[1] Department of Agricultural Biotechnology, Faculty of Agriculture, Eskişehir, University of Eskişehir Osmangazi, Türkiye
Email address: hulya.sipahi@ogu.edu.tr

Abstract

The cellulose and hemicellulose components of plant cell walls are synthesized by the cellulose synthase (CESA) and cellulose synthase-like (CSL) gene families and regulated in response to growth, development, and environmental stimuli. In this study, a total of 29 CESA/CSL family members were identified in *Cannabis sativa* and were grouped into seven subfamilies (~~CESA, CSLA, CSLB, CSLC, CSLD, CSLE and CSLG~~CESA, CSLA, CSLB, CSLC, CSLD, CSLE and CSLG) according to phylogenetic relationships. The CESA/CESA proteins of *C. sativa* were closely related phylogenetically to the members of the subfamily of other species. The CESA/CSL subfamily members of *C. sativa* have unique gene structures. In addition, the expressions of 4 *CESA* and 10 *CsCSL* genes in flower, leaf, root, and stem organs of cannabis were detected using RT-qPCR. The results showed that CESA and CSL genes are expressed at varying levels in several organs. This detailed knowledge of the structural, evolutionary, and functional properties of cannabis *CESA/CSL* genes will provide a basis for designing advanced experiments for genetic manipulation of cell wall biogenesis to improve bast fibers and biofuel production.

Key words *Cannabis sativa*, Cell wall biosynthesis, Gene family, Gene expression

Introduction

Cannabis sativa L., an annual herbaceous plant, has versatile usage features as raw material in paper, textile, biofuel, automotive, and construction industries, mainly due to its cell wall content and other strains are useful as food and pharmaceuticals. Hemp is a preferable fiber source to cotton and other petroleum-derived synthetic fibers because it can be grown even in areas where water, fertilizer, and pesticide use is limited, and it yields a large amount of biomass in a short time. The hemp bast fiber, the outer part of the hemp stem known as phloem fiber, consists of cellulose microfibrils, which are contained in a matrix of hemicellulose and lignin (Behr *et al.*, 2016).

The production and processing of hemp fiber in textiles depend on the amount and distribution of cellulose, hemicellulose, and lignin components in the fiber, which affects the mechanical properties of the fiber. Hemp fibers consist of 53-91% cellulose, 4-18% hemicellulose, 1-17% pectin, and 1-21% lignin, depending on the growing conditions, harvest year and location (Liu *et al.*, 2017).

Cellulose is synthesized in the plasma membrane and located in the cell wall as microfibrils, which is the long chain structure of 1,4-D-glucose units connected by glycosidic bonds (Mujtaba *et al.*, 2017). These microfibrils constitute the scaffold of the cell wall. The stiffness and organization of cellulose microfibrils in the cell wall affect cell growth and increase the cell's resistance to osmotic pressure (Cosgrove, 2005, (Hu *et al.*, 2018). The second major structural component of the cell wall is hemicellulose. Hemicellulose, which has branched polymers consisting of 500-3000 sugar units, is a heteropolymer composed of different types of hemicellulose such as xylan, glucuronoxylan, glucomannan, and xyloglucan. After cellulose fibers are synthesized in the cell wall, they are cross-linked with pectin and hemicelluloses.

Exposing the biosynthesis, assembly, organization, and networking of the cell wall load-bearing cellulosic fibrils is complex and essential (Zhang *et al.*, 2021). Cellulose and hemicellulose polysaccharides are synthesized by the enzymes of the cellulose synthase A (CESA) family and cellulose synthase-like (CSL) family, respectively. These families belong to the *glycosyltransferase 2* (*GT2*) superfamily (Richmond and Somerville, 2000). In Arabidopsis, ten members of the celluloe synthases A (CESA1 to CESA 10) family were found and three of them (CESA1, CESA3, and either CESA2, CESA5, CESA6 or CESA9) were assembled in Cellulose Synthase Complex in the primary cell wall (Zhang *et al.*, 2021). The other Cellulose Synthase Complex, consisting of CESA4, CESA7, and CESA8, carried out cellulose synthesis in the secondary cell wall (Taylor *et al.*, 2003).

The genes of the cellulose synthase-like (CSL) family, grouping in from *CSLA* to *CSLG*, encode enzymes that generate hemicellulose, including xylans, xyloglucans, mannans, glucomannas and β-(1,3;1,4) glucan (Holland *et al.*, 2000). 30 genes belonging to the Cellulose synthase-like (CSL) family have been identified in Arabidopsis. C*SLA, CSLC* , and *CSLF* are responsible for the biosynthesis of mannan, xyloglucan, and (1 → 3; 1 → 4)-β-D-glucan, respectively (Arioli *et al.*, 1998; Richmond and Somerville, 2000, 2001; Lerouxel *et al.*, 2006; Cocuron *et al.*, 2007; Dwivany *et al.*, 2009; Doblin, Pettolino and Bacic, 2010). *CSLD* genes are involved in synthesizing xylan, homogalacturonan, and mannan (Bernal *et al.*, 2007). *CSLJ* are responsible for (1,3;1,4)-β-glucan biosynthesis (Little *et al.*, 2018).

Because *CESA* and *CSL* genes play a crucial role in plant growth, immunity responses to pathogens, and plant biomass increase, genome-wide characterization of cellulose synthase and cellulose synthase-like gene families have been studied in a variety of plants, such as rice (Hazen, Scott-Craig and Walton, 2002), flax(Guo *et al.*, 2022), pineapple (Cao *et al.*, 2019), maize (Appenzeller *et al.*, 2004), tomato (Song *et al.*, 2019), tea (Li *et al.*, 2022) and strawberry (Huang *et al.*, 2022). Subsequently, over-expression and silencing experiments of certain *CESA* and *CSL* genes were performed to identify their roles in plant development and growth, plant phenotype and defense, cell wall integrity, and osmotic stress (Held *et al.*, 2008; Zhu *et al.*, 2010; Chowdhury *et al.*, 2016; Douchkov *et al.*, 2016; Mabuchi *et al.*, 2016; Mazarei *et al.*, 2018; Huang *et al.*, 2022; Li *et al.*, 2022; Zhao *et al.*, 2022). This study aimed to identify CESA and CSL family members in the cannabis genome and reveal their structure and evolution relationships. In addition, analysis of *CsCESA/CsCSL* gene expressions may create a foundation for understanding the function of cellulose and hemicellulose.

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Francês (França)

Formatado: Francês (França)

Formatado: Francês (França)

Formatado: Francês (França)

Código de campo alterado

Código de campo alterado

Formatado: Francês (França)

Formatado: Francês (França)

Código de campo alterado

Código de campo alterado

Formatado: Francês (França)

Código de campo alterado

Comentado [A1]: Please, explain briefly what the role of CESA and CESL in these cases.

Formatado: Francês (França)

Materials & Methods

Identification of CESA/CSL gene family in *C. sativa*

The cellulose synthase (AtCESAs) and cellulose synthase like (AtCSLAs, AtCSLBs, AtCSLCs, AtCSLD, AtCSLE, AtCSLG) protein and coding sequences retrieved from TAIR (https://www.arabidopsis.org/) were used to determine *CESA/CSL* genes in three *C. sativa* genomes (NCBI Genome assembly ASM2916894v1, GeneBank: GCA_029168945.1 Pink pepper (cultivar); Genome assembly JL_Mother, GeneBank: GCA_012923435.1, Jamaican Lion ^4, isolate mother; Genome assembly JL_Father, GeneBank: GCA_013030025.1, Jamaican Lion ^4 isolate father) using BlastP (E-value of $1e^{-5}$) and TBLASTN in NCBI. The Arabidopsis loci used as query are listed TableS1. After blast analysis, the candidate CESA/CSL proteins were validated by checking Pfam domains of the cellulose synthase and cellulose synthase like using HMMER v2.43 online program (https://www.ebi.ac.uk/Tools/hmmer/). CESA family members were recognized by containing RING/U-box type zinc-binding domain (PF14569) and cellulose synthase (PF03552) domain. Those with glycosyltransferase-like family 2 (PFAM 13641) and glycosyltransferase family group 2 (PF13632) domains were characterized as CSLA and CSLC genes, respectively. Cellulose synthase-like D proteins were distinguished by containing both RING/Ubox-like zinc-binding (PF14570) and cellulose synthase (PF03552) domains. All the other cannabis cellulose synthase-like proteins contained the cellulose synthase domains (PF03552) and they were classified according to phylogenetic similarity to CSLB, CSLC, CSLD, CSLE, CSLF, CSLG, CSLH, CSLJ and CSLM family members obtained from *A. thaliana*, *Oryza sativa*, *Linum usitatissimum*, *Sorghum bicolar*, *Solanum lycopersicum*, *Zea mays*, *Glycine max*, *Setaria italica* genomes (TableS1). The identified cellulose synthase and cellulose synthase-like genes in three different cannabis genomes are listed in TableS2, S3. Further bioinformatics analyzes were performed for genes of cannabis genome ASM2916894v1 (NCBI).

Phylogenetic analysis

Amino acid sequences of CESA and CSL family members from *A. thaliana*, *C. sativa*, *Oryza sativa* and *Linum usitatissimum*, CSLMs from *Solanum lycopersicum* and *Glycine max*, and CSLJs from *Sorghum bicolar*, *Zea mays* and *Setaria italica* were used to construct phylogenetic tree. ClustalW was used to align CESA/CSL protein sequences. The evolutionary distances were calculated by the p-distance method (Nei and Kumar, 2000). The phylogenetic tree was constructed using maximum likelihood method with 1000 bootstrap replicates in MEGA11 (Tamura, Stecher and Kumar, 2021).

Gene structure, motif identification, subcellular prediction, and chromosome localization

Gene Structure Display Server v2.0 (https://gsds.gao-lab.org/) was used to analyze the exon-intron structure of these genes (Hu *et al.*, 2015). The MEME program (http:// meme. sdsc. edu/meme/cgi-bin/meme.cgi) analysed the protein sequences to detect the motifs (Bailey *et al.*, 2009). Protein subcellular localizations were predicted by WoLF PSORT with plant parameters. (https://wolfpsort.hgc.jp/). According to WoLF PSORT results, HeatMap was constructed using TBtools software (Chen *et al.*, 2020). The chromosome distribution of all *CESA/CSL* genes of cannabis ASM2916894v1 genome was visualized with TBtools software (Chen *et al.*, 2020).

Cis-Element analysis of putative promoter regions, Ka/Ks calculation, and synteny *analysis*

The 2 Kbp upstream regulatory regions upstream from the start site of translation of *CESA/CSL* genes were retrieved from the NCBI website (https://www.ncbi.nlm.nih.gov/). The PlantCARE online software (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot *et al.*,

Comentado [A2]: Which were the parameters to consider that a given gene should be retrieved?

Comentado [A3]: Default parameters?

Formatado: Realce

Comentado [A4]: Why these subclasses?

Comentado [A5]: Is this still roght?

Comentado [A6]: MEGA displays how gaps were treated. Please, explain.

Formatado: Cor da fonte: Vermelho

Formatado: Fonte: Não Itálico

Comentado [A7]: Default parameters?

Formatado: Realce

2002) was used to investigate the putative *cis*-regulatory elements in these promoter region sequences. HeatMaps were drawn using TBtools software (Chen *et al.*, 2020). Gene duplications were determined by considering the length of the aligned sequence as covering >80% of the longer gene and their similarity being >80%. The Ka/Ks ratios were calculated using TBtools softwaare (Chen *et al.*, 2020). Synteny analysis was carried out using genome files of between *C. sativa* and *A. thaliana*. Reference genome information of the species used in the synthesis analysis is given in TableS4. TBtools software (Chen *et al.*, 2020) with an e-value 1e-10 and 5 BLAST hit cutoffs was used for synteny analysis.

Plant Material

The hemp variety used in this study was 'Wife.' Cuttings of this variety were taken from 5-month-old 'Wife' female plants and aeroponically rooted in an EZ-Cloner Classic Chamber™ (Sacramento, CA, USA). Cuttings were treated with Hormodin powder and placed in rock wool cubes soaked in 20 mL/L Clonex Nutrient Solution (Growth Technologies Ltd., Taunton, UK). Cuttings were rooted for three weeks before planting. Rooted cuttings were potted in a 3-gallon container with Pro-Mix HP and fed twice a week with Botanicare™ liquid fertilizer (Vancouver, WA, USA). At the end of 8 weeks, they were transplanted to a 10-gallon container. During vegetative growth, plants were grown under a daily 18 h light/6 h dark cycle, providing liquid feed in Jack's Nutrients (Jr. Peters, Inc.) containing 100 ppm N at each irrigation. During flowering, plants were grown under a 12 h light/12 h dark cycle for seven weeks and irrigated with 15-30-15 (NPK) Jack's Nutrient (Jr. Peters, Inc.) at 100 ppm N.

RNA isolation, cDNA and RT-qPCR gene expression analysis

The amount of 100 mg of plant tissues (root, stem, leaf, and flower) was collected and immediately frozen in liquid nitrogen. RNA was extracted from leaves using The NucleoSpin Plant and Fungi RNA Isolation Kit (Macherey-Nagel). cDNA was synthesized from 2µg RNA using the iScript Reverse Transcriptase Master Mix (BioRad). qPCR analysis was carried out using Bio-Rad CFX. iTaq Universal Sybr Green Master Mix (Hercules, CA, USA) was used (Bio-Rad). CsUbiquitin (CsUBQ, NCBI GeneBank: JP465573.1) was used as the internal reference (Guo *et al.*, 2018). Selected *CESA/CSL* gene-specific primers were designed by PerlPrimer software (v1.1.21) (PerlPrimer for MicrosoftWindows, Owen J Marshall, Australia) [Marshall 2004] and listed Table S5. The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) was used for gene expression analysis. qPCR conditions included a hold time of 90°C for 3 m, 39 cycles of 95°C for 10 s, 53°C for 30 s, and 72°C for 10 s. qPCR experiments were conducted with three or four biological replications with two technical repeats.

Statical analysis

While paired-t test was used to compare the means of 2 dependent groups, F test (Repeated Measure Anova) was used to compare the means of 3 and 4 dependent groups. In Anova, Bonferroni multiple comparison test was applied for statistically significant group averages to assess variations among different tissues at a 5% significance level. Results are presented in lower case. While there is no significant difference between those marked with the same letter, there are significant differences between those marked with different letters.

Results

A total of 29, 30 and 39 *CESA/CSL* genes were identified in ASM2916894v1 (cultivar pink pepper), JL_Mother and JL_Father (cultivar Jamaican Lion ^4) genomes of *C. sativa*, respectively (TableS2, S3). Two conserved domains, namely RING/U-box type zinc-binding

Formatado: Realce

Formatado: Realce

Comentado [A8]: Provider?

Comentado [A9]: In a growth chamber? Please, explain temperature in this case.

Formatado: Realce

Comentado [A10]: How efficiency and specificity of primers were assessed?

Comentado [A11]: Which condition was the calibrator?

Formatado: Realce

Comentado [A12]: Consider using direct voice instead of passive voice

Formatado: Realce

domain (PF14569) and cellulose synthase (PF03552), were identified in the cellulose synthase A protein (CESA). The structural characteristics of the cellulose synthase-like proteins showed that CSLA proteins (glucomannan 4-beta-mannosyltransferase 9) and CSLC (xyloglucan glycosyltransferase) contained glycosyltransferase-like family 2 (PFAM 13641) and glycosyl transferase family group 2 (PF13632) domains, respectively. CSLD (cellulose-synthase-like D1) proteins included RING/Ubox like zinc-binding (PF14570) and/or cellulose synthase (PF03552) domains. Finally, the CSLB, CSLE, CSLG cellulose synthase-like proteins contained the cellulose synthase domain (PF03552). Also, all proteins included transmembrane and signal domains.

A total of 8 *CsCESA* and 21 *CsCSL* genes (3 *CsCSLA*, 2 *CsCSLB*, 4 *CsCSLC*, 5 *CsCSLD*, 5 *CsCSLE*, 2 *CsCSLG*) from ASM2916894v1 genome were numbered according to their chromosomal locations and named species names (Table S2), These 29 genes were analyzed bioinformatically.

Based on the phylogenetic classification, eight CsCESA proteins were grouped in the CESA clade, which is the largest clade. The cellulose synthase-like proteins were classified into six CsCSL subfamilies (CSLA, CSLB, CSLC, CSLD, CSLE, CSLG) (Fig. 1). Eexceptionally, while no any CsCSL proteins from the pink pepper genome were clustered in the CSLM subgroup, only one protein from the JL_Mother genome (FigS1) and two from the and JL_Father genome were clustered in this subgroup (FigS2).

Many cellulose synthases were predicted to be localized mainly in the plasma membrane and endoplasmic reticulum, followed by the nucleus, vacuole, mitochondria and Golgi body (Fig. 2). Only CsCSLB2 was present in the chloroplast and CsCESA1 and CSLA2 in the extracellular membrane.

All *CsCESA/CsCSL* genes detected in cannabis genome (ASM2916894v1) were mapped and unequally distributed on all chromosomes of *C. sativa* except chromosome 7 (Fig. 3). The highest number of genes were found on chromosome 1 with six, while the lowest number was on chromosome 2 and 9. Also, a total of two duplication events were observed.

Cis-acting elements are conserved nucleotide sequences in the gene's promoter to which transcription factors can bind, and they regulate the transcription of the gene of interest. Variations in the Cis-acting elements in the promoter regions of these genes may lead to variations in the phenotypic characteristics of the organism, such as its development and response to biotic and abiotic factors. A two kilo base upstream sequence of the *CsCESA/CsCSL* genes was searched to identify cis-acting elements. These cis-elements were categorized into four main groups: light-responsive elements, environmental stress-responsive elements, hormone-responsive elements, and development-related elements. Twenty-five light-responsive elements were found (Fig. 4A). All but one *CsCESA/CsCSL* contained the Box 4 motif, making it the most abundant light response element. Thirteen cis-acting elements were involved in environmental stress response (Fig. 4B). MYB (drought and ABA signaling), MYC (drought, salt, and stress response), STRE (multiple signaling including heat stress), LTR (low-temperature response), ARE (anaerobic induction) motifs were most abundant in *CsCESA*/*CsCSL* gene promotors. Thirteen cis-acting hormone response elements were detected (Fig. 4C). ABRE (abscisic-acid-responsive element), ERE (the ethylene-responsive element), CGTCA motifs and TGACG motifs (MeJA-responsive elements) were the most common. *CsCESA/CsCSL* genes carrying hormone response cis-regulatory elements might be upregulated by these hormone treatments. TGA-box (auxin-responsive element) was found only in one *CsCSLE4*. Finally, fourteen development response cis-acting elements were found (Fig. 4D). Among them, the

Comentado [A13]: Is it a pre-requisite in methods?

Formatado: Realce

Comentado [A14]: These characteristics were all pre requisites for classification.

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Formatado: Realce

Comentado [A15]: Please be more specific

Formatado: Realce

Formatado: Realce

Formatado: Realce

Comentado [A16]: Is Box 4 related to light?

Formatado: Realce

AAGAA-motif (the endosperm-specific negative expression), O2-site (zein metabolism regulation), and as-1 (the root-specific expression) were found to be the most abundant elements.

A total of 15 c onserved protein motifs were searched (Fig. 5B). Consistent with conserved domain analysis, CsCSLA shared similar motif composition with CsCSLC, CsCSLG with CsCSLE and CsCSLB, and CsCESA with class CsCSLD. CsCESA and CSLDs have the highest number of motifs, with 15 motifs.

Considerable diversity of the exon-intron structures of the *CsCESA*/*CsCSL* genes was noted (Fig. 5C). *CsCESA* genes usually have 14 exons, with the exciting exceptions that *CsCESA1* has only one exon, while the longest gene, *CsCESA3* has 26 exons. Although the exon-intron numbers of the *CsCSLA* and *CsCSLB* genes were the same within the members of their subfamilies, diversities were observed in the lengths of the introns. *CsCSLC* and *CsCSLD* genes have between 4-5 exons. The highest intron length variation was in the *CsCSLE* subfamily, with 8 exons.

Syntenic maps were constructed between *CESA/CSL* genes of *C. sativa* and *A. thaliana* (Fig.6, Table S4). Nine *C. sativa CESA/CSL* genes were found orthologous to *A. thaliana* genes.

The Nonsynonymous (Ka)/synonymous (Ks) ratios were calculated to reveal the evolutionary status of the two linked CsCESA/CsCSL gene pairs (Table S6). The Ka/Ks ratios of 2 linked *CsCESA* gene pairs (*CsCESA1*/*CsCESA6* and *CsCESA7*/*CsCESA8*) were found to be less than one, indicating the presence of purifying selection and, in this case, the maintenance of the number of members in this gene family.

The members of the cellulose synthase family exhibit tissue-specific expression (Hamann *et al.*, 2004). To investigate the expression patterns of *CESA/CSL* genes in different tissues (flower, leaf, root, stem) of cannabis, the expression of 14 *CESA/CSL* genes, selected based on phylogenetic groups, gene structures, and cis-elements was evaluated by RT-qPCR (Fig. 7, Table S7). All examined *CsCESA* genes were expressed in four different tissues with variable levels, suggesting that all these *CESA* genes are necessary for primary or secondary cell wall formation. One of the *CsCESA* genes (Fig.7a) was expressed relatively higher in flower and the other two (b, c) in leaf, but the difference in the expression levels of these genes between the tissues was not statistically significant. In addition, the expression level of one *CsCESA* gene (Fig.7d) was highest in stem and lowest in flower. *CsCSLA* (Fig.7f), *CsCSLB* (Fig.7g), *CsCSLG* (Fig.7m) showed good leaf-specific expression. *CsCSLC* (Fig.7h) exhibited the highest expression in the flower, whereas CsCSLD, CsCSLD, CsCSLG genes (Fig.7j, k, n) were expressed only in root and stem.

Discussion

The structure, composition, and organization of cell wall components have critical effects in terms of affecting the growth and development of the plant and its response to environmental stimuli by determining the cell's shape, strength, structural integrity and response to abiotic and biotic environmental stresses. Furthermore, the substances of the cell wall affect the quality and processing of the plant for use as paper, textiles, bioethanol, feed and food.

The plant cell wall matrix mainly comprises polysaccharides such as cellulose, hemicellulose and pectin. The cellulose is embedded in the cell wall matrix as microfibrils. Microfibrils are a bundle of 40 cellulose molecules that run parallel to each other and are attached by hydrogen bonds. Each cellulose molecule is an unbranched polymer composed of glucose units linked by an β-1,4-bond. It is estimated that hundreds of enzymes are responsible for cell wall biosynthesis (Keegstra and Raikhel, 2001; Scheible and Pauly, 2004; Liepman and

Formatado: Realce

Formatado: Realce

Formatado: Realce

Comentado [A17]: Please, develop more about this. Which were the orthologs?

Formatado: Realce

Cavalier, 2012). Cellulose Synthase (CESA) and Cellulose Synthase-Like (CSL) family proteins synthesize various β-glycan polymers. In this study, Cellulose Synthase (CESA) and Cellulose Synthase-Like (CSL) family proteins were identified and characterized, and their phylogenetic relationships were revealed by using genome sequence and bioinformatics analysis tools in cannabis, which is a fiber and biofuel source and has a crucial phytochemical content.

To better understand the CsCESA/CsCSL characteristics in cannabis, genome-wide analyses were carried out via phylogenetic relationships with existing gene structures and expression in different tissues. 8 *CsCESA* and 21 *CsCSL* (3 *CsCSLA*, 2 *CsCSLB*, 4 *CsCSLC*, 5 *CsCSLD*, 5 *CsCSLE*, 2 *CsCSLG*) genes in cannabis ASM2916894v1genome resemble the CESA/CSL phylogenetic classification in other species. As expected, the monocot-specific CSLF, CSLH and CSLJ protein subfamilies were not found in the cannabis genome.

Syntenic relationships are shaped by conservation or DNA substitution rates among taxa and genes in the genomes of different species. Gene orthology relationship determined by synteny data analysis can support the phylogenetic relationships of multiple gene families (Gabaldón et. al. 2013). The orthologous relationships of Cs*CSL* genes with *A. thaliana*, with the exception of some Cs*CESA* genes, supported their close relationship in the phylogenetic tree (Table S4, Fig. 1).

The existence of closely clustered CESA/CSL subfamily members between cannabis and other eight species used in phylogenetic tree resulted from evolutionary conservation and closer homology (Fig. 1). According to phylogenetic tree analysis (Fig. 1), CsCESA1/7 and CsCESA2 were closely clustered with AtCESA1 and AtCESA3, respectively. In addition, AtCESA5/6/2/9 are grouped with CsCESA5/8/6. AtCESA1, AtCESA3, and any of AtCESA2/ 5/ 6, or /9 complex cause primary wall cellulose accumulation with unequal enzymatic activity in cells undergoing cell division and elongation (Persson *et al.*, 2007; Hu *et al.*, 2018). The results inferred that these orthologous *CsCESA1/7/2*, or 5/8/6 genes may have similar functions.

The expression analysis showed that four *CsCESA* genes were expressed in four different tissues with varying levels. Only one *CESA* gene (Fig.7d) were expressed at highest level in the stem. These results may support that *CsCESA* genes play roles in the growth of different tissues. In addition, Arabidopsis CESA4/7/8 are involved in secondary wall thickness by increasing the amount of cellulose (Tanaka *et al.*, 2003; Taylor *et al.*, 2003; Zhong, Cui and Ye, 2019) and phylogenetically clustered together with *CsCESA4/6/3* gene homologous, respectively (Fig. 1). These results suggest that *CsCESA4/6/3* might participate in establishing secondary cell walls.

Revealing gene structures and protein motif patterns of gene family members may elucidate the evolution and diversity of their structure and function. Most CsCESA and CsCSL subfamily members, closely grouped in the phylogenetic tree according to their amino acid sequences (Fig. 5A), share the same or similar motif distributions.

The motif components of CsCESA and CsCSLD were highly similar. However, the CsCSLA and CsCSLC genes had different motif patterns. Although the members of the CsCESA/CsCSL subfamily were generally similar in terms of exon and intron number and length, they also showed some differences (Fig. 5C). A similar situation has been noted for members of the DcCSLD subfamily in *Dendrobium catenatum*, and it has been reported that intron gain/loss may occur in the evolution of *DcCSLD* genes (Xi *et al.*, 2021).

Several members of the CESA/CSL gene family have been reported to have varying expression levels depending on tissue type, growth stage, and environmental factors (Cao *et al.*, 2019, Li *et al.*, 2020, 2022; Song *et al.*, 2019; Kaur *et al.*, 2017; Yuan *et al.*, 2021; Liu *et al.*, 2022; Nawaz *et al.*, 2017; Marcotuli *et al.*, 2018); (Hou *et al.*, 2023). Likewise, the results of

Formatado: Realce
Formatado: Realce
Formatado: Realce
Comentado [A18]: Which ones?
Formatado: Realce
Formatado: Francês (França)
Formatado: Francês (França)
Formatado: Francês (França)
Formatado: Francês (França)
Formatado: Francês (França)
Formatado: Francês (França)
Formatado: Francês (França)
Código de campo alterado
Código de campo alterado
Código de campo alterado
Código de campo alterado
Código de campo alterado
Formatado: Francês (França)
Formatado: Francês (França)
Formatado: Francês (França)
Formatado: Francês (França)
Código de campo alterado
Código de campo alterado
Código de campo alterado

this study identified different expression levels of several *CsCESA/CsCSL* genes in four tissues. *CSLA* (Fig. 7f), *CSLB* (Fig. 7g) and *CSLG* (Fig. 7m) were expressed mainly in the leaf. In contrast, *CsCSLD* (Fig. 7j,k) and *CsCSLG* (Fig. 7n) were expressed only in root and stem, whereas *CSLC* (Fig. 7h) was mainly expressed in the flower. In the previous study conducted to determine transcriptomic changes associated with bast fiber development stage in textile hemp, certain CES/CSL family members showed differential expression patterns in the stem's upper, middle, and lower internodes (Guerriero *et al.*, 2017). For example, the annotation against the *Arabidopsis* database showed that some contigs annotated with *CSLC5* were more highly expressed in the upper internode, with *CSLC04* in the middle internode, with *CSLE1, CSLG1* and *CSLB04* in the lower internode. The progressive decrease in expression from the top to the bottom of the stem was detected in *CSLC04* and *CSLC5* genes, while progressive increase in expression along the stem axis for *CSLE1, CSLG1, IRX1, CSLG3, CSLE1, CSLB04* genes.

In another important fiber plant, flax, expression differences were also found in different growth stages and stem parts, confirming the role of *CESA/CSL* genes in cell wall thickening (Guo *et al.*, 2022). The flax *CESA3/8* (Lus10007538, Lus10007296) and *CSLD4* (Lus10008225) genes have been shown to be active during early fiber development. These genes are specifically expressed at the stage of fiber development when there is an increased amount of secondary cell wall deposition during the period of rapid growth. The flax *CESA6* genes (*Lus10006161.g* and *Lus10041063.g*) were found to be specifically expressed in the stem during fiber maturation. The expression levels of *LusCSLE1* (*Lus10016625*) and the two Lus*CSLG3* (*Lus10023056.g* and *Lus10023057.g*) genes were higher in 30 cm plants than in 50 cm plants. The opposite trend occurred for *CESA6* (*Lus10006161.g* and *Lus10041063.g*) genes. The phylogenetic tree of flax and cannabis *CESA/CSL* genes showed that the flax *CESA/CSL* genes were closely clustered with those of cannabis (Fig. 1). This may indicate that flax and cannabis *CESA/CSL* genes may have similar functions during the fiber development stages.

Fibre quality and yield traits could be improved by controlling cellulose biogenesis (Gipson, 1986). Therefore, quantitative trait loci (QTLs) affecting fiber yield and quality and *CESA* genes associated with these QTLs have been identified in cotton. 18 *GhCesA* genes were found to be linked to 74 fiber-quality QTLs. A few cotton CESA genes were differentially expressed in ovules at 0-3 days post-anthesis between two backcrossed inbred lines with different fiber lengths. Also, the positive regulatory role of *GhCESA4* for fiber length and strength in cotton was demonstrated by linkage analysis of QTLs (Liú et. al. 2023).

In addition to improving fiber quality, a mutant allele of *CESA3* gene was developed using CRISPR/Cas9 base editing to improve herbicide resistance in Arabidopsis plants, and this mutant allele was found to confer plant resistance to the herbicide C17 (Zhubing et. al. 2019). Shortly, editing *CESA/CSL* genes could lead to advances in areas such as enhancing plant growth and development, and improving disease and pest resistance. Also, the silencing or overexpression of genes associated with cell wall synthesis may indicate changes in plant-pathogen interactions in the cell wall, which may shed light on whether cannabinoid production is affected. In this study, the genome-wide identification of *CsCESA/CsCSL* genes will provide the basis for future investigation of their role.

Conclusions

In this study, 8 *CsCESA* and 21 *CsCSL* genes in *C. sativa* genome (ASM2916894v1) were identified according to their conserved domains and motifs. Their features were characterized, including gene structure, chromosome location, phylogenetic analysis and syntenic relationships. The cis-element analysis identified the multifunctional role of *CsCESA/CsCSL* genes in growth,

Formatado: Realce

Formatado: Realce

Comentado [A19]: I don't think this is relevant for this manuscript

Formatado: Realce

hormone responsiveness, and biotic and abiotic stress responsiveness. *CsCESA/CsCSL* genes exhibited diverse expression patterns in flower, leaf, root, and stem tissues. The detailed characterization of CESA/CSL in cannabis may aid in designing experiments for future genetic manipulation of cellulose and hemicellulose synthesis genes to breed cultivars with high fiber quality and bioethanol yield.

References

Appenzeller, L. *et al.* (2004) 'Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family', *Cellulose*, 11, pp. 287–299.

Arioli, T. *et al.* (1998) ... 'response: How many cellulose synthase-like gene products actually make cellulose?', *Trends in Plant Science*, 3(5), pp. 165–166.

Bailey, T.L. *et al.* (2009) 'MEME SUITE: tools for motif discovery and searching', *Nucleic acids research*, 37(suppl_2), pp. W202–W208.

Behr, M. *et al.* (2016) 'Studying secondary growth and bast fiber development: the hemp hypocotyl peeks behind the wall', *Frontiers in Plant Science*, 7, p. 1733.

Bernal, A.J. *et al.* (2007) 'Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in Arabidopsis', *The Plant Journal*, 52(5), pp. 791–802.

Cao, S. *et al.* (2019) 'Genome-wide identification, expression pattern analysis and evolution of the Ces/Csl gene superfamily in pineapple (Ananas comosus)', *Plants*, 8(8), p. 275.

Chen, C. *et al.* (2020) 'TBtools: an integrative toolkit developed for interactive analyses of big biological data', *Molecular plant*, 13(8), pp. 1194–1202.

Chowdhury, J. *et al.* (2016) 'Down-regulation of the glucan synthase-like 6 gene (HvGsl6) in barley leads to decreased callose accumulation and increased cell wall penetration by Blumeria graminis f. sp. hordei', *New Phytologist*, 212(2), pp. 434–443.

Cocuron, J.-C. *et al.* (2007) 'A gene from the cellulose synthase-like C family encodes a β-1, 4 glucan synthase', *Proceedings of the National Academy of Sciences*, 104(20), pp. 8550–8555.

Cosgrove, D.J. (2005) 'Growth of the plant cell wall', *Nature reviews molecular cell biology*, 6(11), pp. 850–861.

Doblin, M.S., Pettolino, F. and Bacic, A. (2010) 'Plant cell walls: the skeleton of the plant world', *Functional Plant Biology*, 37(5), pp. 357–381.

Douchkov, D. *et al.* (2016) 'The barley (Hordeum vulgare) cellulose synthase-like D2 gene (HvCslD2) mediates penetration resistance to host-adapted and nonhost isolates of the powdery mildew fungus', *New Phytologist*, 212(2), pp. 421–433.

Dwivany, F.M. *et al.* (2009) 'The CELLULOSE-SYNTHASE LIKE C (CSLC) family of barley includes members that are integral membrane proteins targeted to the plasma membrane', *Molecular Plant*, 2(5), pp. 1025–1039.

Guerriero, G. *et al.* (2017) 'Transcriptomic profiling of hemp bast fibres at different developmental stages', *Scientific reports*, 7(1), p. 4961.

Guo, R. *et al.* (2018) 'Evaluation of reference genes for RT-qPCR analysis in wild and cultivated Cannabis', *Bioscience, Biotechnology, and Biochemistry*, 82(11), pp. 1902–1910.

Guo, Y. *et al.* (2022) 'Comparative transcriptomic analysis identifies key cellulose synthase genes (CESA) and cellulose synthase-like genes (CSL) in fast growth period of flax stem (Linum usitatissimum L.)', *Journal of Natural Fibers*, 19(15), pp. 10431–10446.

Hamann, T. *et al.* (2004) 'Global expression analysis of CESA and CSL genes in Arabidopsis', *Cellulose*, 11, pp. 279–286.

Formatado: Francês (França)

Formatado: Francês (França)

Formatado: Francês (França)

Formatado: Francês (França)

Formatado: Francês (França)

Hazen, S.P., Scott-Craig, J.S. and Walton, J.D. (2002) 'Cellulose synthase-like genes of rice', *Plant physiology*, 128(2), pp. 336–340.
Held, M.A. *et al.* (2008) 'Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley', *Proceedings of the National Academy of Sciences*, 105(51), pp. 20534–20539.
Holland, N. *et al.* (2000) 'A comparative analysis of the plant cellulose synthase (CesA) gene family', *Plant physiology*, 123(4), pp. 1313–1324.
Hou, Y. *et al.* (2023) 'Identification, Classification and Expression Analysis of the CesA Gene Family from Pinus massoniana', *Forests*, 14(5), p. 1035.
Hu, B. *et al.* (2015) 'GSDS 2.0: an upgraded gene feature visualization server', *Bioinformatics*, 31(8), pp. 1296–1297.
Hu, H. *et al.* (2018) 'Three AtCesA6-like members enhance biomass production by distinctively promoting cell growth in Arabidopsis', *Plant biotechnology journal*, 16(5), pp. 976–988.
Huang, H. *et al.* (2022) 'Genome-wide identification and functional analysis of Cellulose synthase gene superfamily in Fragaria vesca', *Frontiers in Plant Science*, 13, p. 1044029.
Kaur, S. *et al.* (2017) 'Genome-wide analysis of the cellulose synthase-like (Csl) gene family in bread wheat (Triticum aestivum L.)', *BMC plant biology*, 17(1), pp. 1–17.
Keegstra, K. and Raikhel, N. (2001) 'Plant glycosyltransferases', *Current opinion in plant biology*, 4(3), pp. 219–224.
Lerouxel, O. *et al.* (2006) 'Biosynthesis of plant cell wall polysaccharides—a complex process', *Current opinion in plant biology*, 9(6), pp. 621–630.
Lescot, M. *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(1), pp. 325–327.
Li, G. *et al.* (2020) 'Genome-wide characterization of the cellulose synthase gene superfamily in Pyrus bretschneideri and reveal its potential role in stone cell formation', *Functional & integrative genomics*, 20, pp. 723–738.
Li, Q. *et al.* (2022) 'Genome-Wide Characterization of the Cellulose Synthase Gene Superfamily in Tea Plants (Camellia sinensis).', *Phyton (0031-9457)*, 91(10).
Liepman, A.H. and Cavalier, D.M. (2012) 'The cellulose synthase-like A and cellulose synthase-like C families: recent advances and future perspectives', *Frontiers in plant science*, 3, p. 109.
Little, A. *et al.* (2018) 'Revised phylogeny of the cellulose synthase gene superfamily: insights into cell wall evolution', *Plant physiology*, 177(3), pp. 1124–1141.
Liu, M. *et al.* (2017) 'Targeted pre-treatment of hemp bast fibres for optimal performance in biocomposite materials: A review', *Industrial crops and products*, 108, pp. 660–683.
Liu, X. *et al.* (2022) 'Genome-wide bioinformatics analysis of Cellulose Synthase gene family in common bean (Phaseolus vulgaris L.) and the expression in the pod development', *BMC Genomic Data*, 23(1), pp. 1–15.
Liú, R., (2023). 'Genetic linkage analysis of stable QTLs in Gossypium hirsutum RIL population revealed function of GhCesA4 in fiber development'. *Journal of Advanced Research*.
Livak, K.J. and Schmittgen, T.D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method', *Methods*, 25(4), pp. 402–408. Available at: https://doi.org/https://doi.org/10.1006/meth.2001.1262.
Mabuchi, A. *et al.* (2016) 'Phenotypic screening of Arabidopsis T-DNA insertion lines for cell wall mechanical properties revealed ANTHOCYANINLESS2, a cell wall-related gene', *Journal of plant physiology*, 191, pp. 29–35.

Formatado: Francês (França)

Marcotuli, I. *et al.* (2018) 'Expression analysis of cellulose synthase-like genes in durum wheat', *Scientific Reports*, 8(1), p. 15675.
Mazarei, M. *et al.* (2018) 'Functional analysis of cellulose synthase CesA4 and CesA6 genes in switchgrass (Panicum virgatum) by overexpression and RNAi-mediated gene silencing', *Frontiers in plant science*, 9, p. 1114.
Mujtaba, M. *et al.* (2017) 'Detailed adsorption mechanism of plasmid DNA by newly isolated cellulose from waste flower spikes of Thypa latifolia using quantum chemical calculations', *International journal of biological macromolecules*, 102, pp. 914–923.
Nawaz, M.A. *et al.* (2017) 'Genome and transcriptome-wide analyses of cellulose synthase gene superfamily in soybean', *Journal of plant physiology*, 215, pp. 163–175.
Nei, M. and Kumar, S. (2000) *Molecular evolution and phylogenetics*. Oxford University Press, USA.
Persson, S. *et al.* (2007) 'Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis', *Proceedings of the National Academy of Sciences*, 104(39), pp. 15566–15571.
Richmond, T.A. and Somerville, C.R. (2000) 'The cellulose synthase superfamily', *Plant physiology*, 124(2), pp. 495–498.
Richmond, T.A. and Somerville, C.R. (2001) 'Integrative approaches to determining Csl function', *Plant cell walls*, pp. 131–143.
Scheible, W.-R. and Pauly, M. (2004) 'Glycosyltransferases and cell wall biosynthesis: novel players and insights', *Current opinion in plant biology*, 7(3), pp. 285–295.
Song, X. *et al.* (2019) 'Genome-wide characterization of the cellulose synthase gene superfamily in Solanum lycopersicum', *Gene*, 688, pp. 71–83.
Tamura, K., Stecher, G. and Kumar, S. (2021) 'MEGA11: molecular evolutionary genetics analysis version 11', *Molecular biology and evolution*, 38(7), pp. 3022–3027.
Tanaka, K. *et al.* (2003) 'Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall', *Plant physiology*, 133(1), pp. 73–83.
Taylor, N.G. *et al.* (2003) 'Interactions among three distinct CesA proteins essential for cellulose synthesis', *Proceedings of the National Academy of Sciences*, 100(3), pp. 1450–1455.
Xi, H. *et al.* (2021) 'Genome-wide identification of Cellulose-like synthase D gene family in Dendrobium catenatum', *Biotechnology & Biotechnological Equipment*, 35(1), pp. 1163–1176.
Yuan, W. *et al.* (2021) 'Genome-wide identification of banana csl gene family and their different responses to low temperature between chilling-sensitive and tolerant cultivars', *Plants*, 10(1), p. 122.
Zhang, B. *et al.* (2021) 'The plant cell wall: Biosynthesis, construction, and functions', *Journal of Integrative Plant Biology*, 63(1), pp. 251–272.
Zhao, H. *et al.* (2022) 'Cellulose synthase-like protein OsCSLD4 plays an important role in the response of rice to salt stress by mediating abscisic acid biosynthesis to regulate osmotic stress tolerance', *Plant Biotechnology Journal*, 20(3), pp. 468–484.
Zhong, R., Cui, D. and Ye, Z. (2019) 'Secondary cell wall biosynthesis', *New Phytologist*, 221(4), pp. 1703–1723.
Zhu, Jianhua *et al.* (2010) 'A cellulose synthase-like protein is required for osmotic stress tolerance in Arabidopsis', *The Plant Journal*, 63(1), pp. 128–140.

Formatado: Francês (França)

Formatado: Francês (França)