Genome-wide bioinformatics analysis of Cellulose Synthase gene family in common bean (*Phaseolus vulgaris* L.) and the expression in the pod development

Xiaoqing Liu, Hongmei Zhang, Wei Zhang, Wenjing Xu, Songsong Li, Xin Chen* and Huatao Chen*

Abstract

Background: *CesA* and *Cs*l gene families, which belong to the cellulose synthase gene superfamily, plays an important role in the biosynthesis of the plant cell wall. Although researchers have investigated this gene superfamily in several model plants, to date, no comprehensive analysis has been conducted in the common bean.

Results: In this study, we identified 39 putative cellulose synthase genes from the common bean genome sequence. Then, we performed a bioinformatics analysis of this gene family involving sequence alignment, phylogenetic analysis, gene structure, collinearity analysis and chromosome location. We found all members possess a cellulose_synt domain. Phylogenetic analysis revealed that these cellulose synthase genes may be classified into five subfamilies, and that members in the same subfamily share conserved exon-intron distribution and motif compositions. Abundant and distinct *cis*-acting elements in the 2 k basepairs upstream regulatory regions indicate that the cellulose synthase gene family may plays a vital role in the growth and development of common bean. Moreover, the 39 cellulose synthase genes are distributed on 10 of the 11 chromosomes. Additionally expression analysis shows that all *CesA/Csl* genes selected are constitutively expressed in the pod development.

Conclusions: This research reveals both the putative biochemical and physiological functions of cellulose synthase genes in common bean and implies the importance of studying non-model plants to understand the breadth and diversity of cellulose synthase genes.

Keywords: Cellulose synthase genes, Common bean, Phylogenetic relationships, Gene expression

Background

In plants, there exists a cellulose synthase superfamily including *CesA* (cellulose synthase) and *Csl* (cellulose synthase-like) gene family, both of which belong to the glycosyltransferase GT2 family and have a similar protein sequence structure. The encoded proteins all have glycosyltransferase activity and are key enzymes essential for cellulose and hemicellulose synthesis [1, 2], which are main components of the primary and secondary cell wall.

The cellulose synthase (*CesA*) gene was first identified from cotton fibers according to its sequence similarity with a bacterial *CesA* gene [3]. Subsequently, cellulose synthase genes were identified in Arabidopsis [4], rice [5], and maize [6], barley [7]. The CesA family contains a conserved motif (DDDQxxRW) and a zinc-finger domain [8]. In *Arabidopsis thaliana*, AtCesA1, AtCesA3, and AtCesA6 form a cellulose synthase complex and participate in the synthesis of the primary cell walls. Whereas,

*Correspondence: cx@jaas.ac.cn; cht@jaas.ac.cn
Jiangsu Academy of Agricultural Sciences, Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Institute of Industrial Crops, Nanjing 210014, China

Liu *et al. BMC Genomic Data* (2022) 23:9

Page 2 of 15

AtCesA4, AtCesA7, and AtCesA8 mediate in the synthesis of the secondary cell walls. It is generally accepted that despite the involvement of most CesA in the synthesis of the primary and secondary cell wall [9–11], AtCesA2, AtCesA5, AtCesA9 are considered homologous proteins of AtCesA6, and these proteins are functionally redundant with each other.

Csl (cellulose synthase-like) genes, which share a relatively high sequence similarity to *CesA* (cellulose synthase A) genes [12], is divided into 9 subfamilies, in which *CslA* and *CslC* are distantly related to the other families; *CslA*, *CslC*, *CslD* and *CslJ* subfamilies are ubiquitously present in all terrestrial plants [5], *CslF* and *CslH* are specific to monocots, *CslB* and *CslG* are thought to be unique to eudicots [13, 14]. There are many evidence supporting *CSl* gene family involving in the biosynthesis of cell wall polysaccharides. CslA catalyzes the synthesis of (1,4)-β-D-mannan [15–18], and CslC is involved in catalyzing the formation of the xyloglucan cytoskeleton [18–20]. Meanwhile, CslD also plays a vital role in xylan and galactoaldoglycan synthesis [21–24]. Additionally, CslF is a gene family unique to monocotyledons, and mediates the synthesis of β-(l,3;l,4)-D-glucan [25, 26]. However, at present, the biological functions of CslB, CslE, CslG and CsJ gene families remain unknown.

Not only a major source of protein and essential nutrients, common bean (*Phaseolus vulgaris* L.) is also an important crop to society and the global economy [27]. However, the CesA/Csl gene families in common bean have not yet been extensively explored. Molecular biology, genomics, and computational biology have transformed the field of biology, gene discovery and functional gene annotation in plant genome-wide data is a rapidly growing research area. Considering the critical role of CesA/Csl in both the integrity and function of plants, we present a comprehensive phylogenetic and functional bioinformatics analysis of the *CesA/Csl* gene family in common bean. Then, using quantitative real-time polymerase chain reaction (qRT-PCR) analysis on genes identified in our computational pipeline, we validate the main genes central for the development of legume pods. These findings shed new light on the relationship between *CesA/Csl* function and the development of common bean. Furthermore, this research presents a theoretical framework for gene cloning and expression in the future, with the application of genetically improving the common bean through breeding.

Results

Identification of cellulose synthase genes in common bean
In order to identify the cellulose synthase gene family of common bean, first, the Hidden Markov Model of 40 *Arabidopsis* cellulose synthase proteins was constructed, then the model was used as queries to search against the common bean protein databases with the BLASTP program at an e-value threshold of 10–10. Then, we searched for the cellulose synthase gene family of common bean using the constructed model and finally a total of 39 sequences can matched to *CesA/Csl* superfamily. 14 gene members contained a cellulose synthase domain (CS) and zinc finger structure (zf-UDP), 25 gene members only harbored a CS domain. The identified cellulose synthase proteins were named according to the order of their subfamilies and gene IDs.

These putative cellulose synthase genes in this analysis were predicted to range from 467 to 1374 amino acids in length and 53.34 kDa to 155.53 kDa in molecular weight. Furthermore, the protein isoelectric points (pIs) ranged from 5.62 to 9.05, the number of predicted TMHs ranged from 0 to 13. The subcellular localization of the putative cellulose synthase genes was predicated to be located in the membrane bound golgi and plasma membranes PHAVU_005G116500g, which were exist in extracellular (secreted) (Table 1).

Phylogenetic analysis of cellulose synthase gene in common bean
A phylogenetic analysis was used investigate the evolutionary relationships among cellulose synthase proteins. Constructed with cellulose synthase proteins from Arabidopsis and 39 from common bean, the phylogenetic analysis showed that 15 putative cellulose proteins from common bean belong to the CesA family, while the remaining 25 cellulose synthase proteins are members of the Csl (B, D, E, and G) family (Fig. 1). CslD is close to CesA, while CslG is distantly related to the other families.

Gene structure analysis of cellulose synthase gene in common bean
Exon-intron structures of each *CesA/Csl* gene were constructed through the sequence alignment of their corresponding genomic DNA. Based on the phylogenetic analysis, putative *CesA/Csl* genes' exon/intron structures in common bean were organized into five subgroups (Fig. 2). *CesA/Csl* genes in the same subgroup had conserved exon/intron structures, while genes in different groups exhibited distinct gene structures. And we found that *CesA* gene members had the most introns, while the *CslD* gene members had the fewest number of introns.

Conserved motif domains of CesA/Csl gene in common bean
To evaluate the structural diversity of cellulose synthase proteins, we used the online program MEME (http://meme.sdsc.edu/meme/cgi-bin/meme.cgi) to search for conserved motifs in putative cellulose synthase protein

Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 3 of 15

Table 1 Classification and characterization of the putative Cellulose Synthase genes in *Phaseolus vulgaris*

Gene_name	Family	Chromosome. No	Amino acid no.	Molecular weight (Da)	Isoelectric points	Location	Number of predicted TMHs	Domain
PHAVU_001G211000g	CslD	1	1149	127,991.62	8.49	Membrane bound Golgi	6	CS (PF03552)
PHAVU_002G040200g	CslD	2	1144	127,998.61	7.06	Membrane bound Golgi	8	CS (PF03552)
PHAVU_002G136300g	CslD	2	1117	125,646.3	6.11	Membrane bound Golgi	8	CS (PF03552)
PHAVU_002G188600g	CesA	2	1034	117,583.67	8.18	Plasma membrane	8	zf-UDP (PF14569), CS (PF03552)
PHAVU_002G240200g	CesA	2	976	110,235.69	5.89	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_002G268200g	CesA	2	1097	123,785.06	6.66	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_003G023000g	CslD	3	992	111,961.41	8.65	Membrane bound Golgi	8	CS (PF03552)
PHAVU_003G154600g	CesA	3	1031	116,535.77	6.09	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_003G290600g	CslG	3	734	84,006.68	7.23	Plasma membrane	7	CS (PF03552)
PHAVU_004G093300g	CesA	4	1089	123,074.89	6.76	Plasma membrane	8	zf-UDP (PF14569), CS (PF03552)
PHAVU_005G001000g	CslG	5	700	79,831.81	8.04	Plasma membrane	5	CS (PF03552)
PHAVU_005G010400g	CesA	5	1033	117,479.39	6.23	Plasma membrane	8	zf-UDP (PF14569), CS (PF03552)
PHAVU_005G022100g	CesA	5	1075	120,033.65	6.94	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_005G116200g	CslB	5	743	84,181.59	8.38	Membrane bound Golgi	9	CS (PF03552)
PHAVU_005G1163000g	CslB	5	528	58,607.06	7.55	Plasma membrane	6	CS (PF03552)
PHAVU_005G1164000g	CslB	5	520	57,674.75	6.18	Plasma membrane	5	CS (PF03552)
PHAVU_005G116500g	CslB	5	1374	155,531.98	9	Extracellular (Secreted)	13	CS (PF03552)
PHAVU_005G116700g	CslB	5	750	85,099.39	7.5	Membrane bound Golgi	6	CS (PF03552)
PHAVU_006G058400g	CslE	6	738	84,196.75	8.22	Membrane bound Golgi	7	CS (PF03552)
PHAVU_006G0586001g	CslE	6	528	60,056.71	8.08	Membrane bound Golgi	6	CS (PF03552)
PHAVU_006G058700g	CslE	6	752	86,312.44	7.14	Membrane bound Golgi	7	CS (PF03552)
PHAVU_007G081700g	CesA	7	1093	123,205.98	6.39	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_007G130400g	CslG	7	741	84,593.15	5.94	Plasma membrane	7	CS (PF03552)
PHAVU_007G190300g	CesA	7	884	100,646.12	6.26	Plasma membrane	8	CS (PF03552)
PHAVU_008G193000g	CslD	8	1128	125,931.82	5.89	Membrane bound Golgi	6	CS (PF03552)
PHAVU_008G279600g	CslE	8	748	85,229.48	8.68	Membrane bound Golgi	6	CS (PF03552)
PHAVU_008G279700g	CslE	8	1006	114,636.64	8.54	Membrane bound Golgi	10	CS (PF03552)

Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 4 of 15

Table 1 (continued)

Gene_name	Family	Chromosome. No	Amino acid no.	Molecular weight (Da)	Isoelectric points	Location	Number of predicted TMHs	Domain
PHAVU_008G279800g	CslE	8	744	85,290.55	8.41	Membrane bound Golgi	7	CS (PF03552)
PHAVU_009G090100g	CesA	9	968	109,222.66	6.33	Plasma membrane	8	zf-UDP (PF14569), CS (PF03552)
PHAVU_009G094200g	CesA	9	1084	122,269.89	6.35	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_009G205100g	CesA	9	1041	117,486.26	6.31	Plasma membrane	8	zf-UDP (PF14569), CS (PF03552)
PHAVU_009G205200g	CesA	9	974	109,369.83	5.97	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_009G2260000g	CslG	9	505	57,280.95	9.05	Plasma membrane	5	CS (PF03552)
PHAVU_009G242700g	CesA	9	1048	118,882.58	7.95	Plasma membrane	8	zf-UDP (PF14569), CS (PF03552)
PHAVU_011G020100g	CslD	11	1148	128,959.97	7.04	Membrane bound Golgi	8	CS (PF03552)
PHAVU_011G101500g	CslB	11	746	84,596.22	8.42	Membrane bound Golgi	6	CS (PF03552)
PHAVU_011G211500g	CesA	11	1074	120,142.75	6.68	Plasma membrane	6	zf-UDP (PF14569), CS (PF03552)
PHAVU_L008300g	CslE	Scaffold_191	648	72,876.83	6.63	Membrane bound Golgi	5	CS (PF03552)
PHAVU_L009400g	CslB	Scaffold_243	467	53,343.3	5.62	Plasma membrane	2	CS (PF03552)

CesA Cellulose synthase A, *Csl* Cellulose synthase-like

sequences (Fig. 3). We identified 15 conserved motifs, and found that members in the same subfamily shared similar conserved motifs (Fig. 3). In addition, members in the CesA and CslD groups contained more motifs than members in CslB,, CslE and CslG groups (except PHAVU_005G116500).

Promoter regions analysis of CesA/Csl genes
To identify the cis-elements in the promoters of *CesA/Csl* genes in common bean, the 2000 bp basepairs upstream of the start codon of each gene were analysed using Plant-CARE online (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The results showed that abundant cis-elements were present in the promoters of *CesA/Csl* genes (except PHAVU_L008300). CAAT-box and TATA-box were the most abundant elements. TATA-box, a core promoter element, which located in about 30 bp upstream of the transcription start site, while CAAT-box is a common *cis*-acting element in promoter and enhancer regions. Also present were MYB transcription factor binding sites (TAACCA), light response element Box 4, and stress response elements, including MYC (in response to drought stress) and ARE (related to anaerobic stress). In addition, hormone response elements, ERE and ABRE, were observed, which respond respectively to ethylene and abscisic acid. The *cis*-acting elements had diverse functions and abundant types, indicating that the cellulose synthase gene family may plays an important role in the growth and development of common bean (Fig. 4, Supplementary Material File 2).

CAAT-box: a common *cis*-acting element in promoter and enhancer regions; TATA: a core promoter element located about 30 bp upstream of the transcription start site; MYB: MYB recognition site; G-box: *cis*-acting regulatory element involved in light responsiveness; TATA-box: a sequence of DNA, consisting of nucleobases TAT AAA, located in the promoter region about 25 base pairs before the site of transcription; MYC: *cis*-acting element involved in drought and abscisic acid responsiveness; Box4: conserved DNA module involved in light responsiveness; AT-TATA-box:; ERE: ethylene-responsive element; ABRE: *cis*-acting element involved in abscisic acid responsiveness; ARE: *cis*-acting regulatory element essential for the anaerobic induction.

Chromosome location of CesA/Csl genes in common bean
Then, we investigated the chromosome distribution of the 39 cellulose synthase genes using the physical

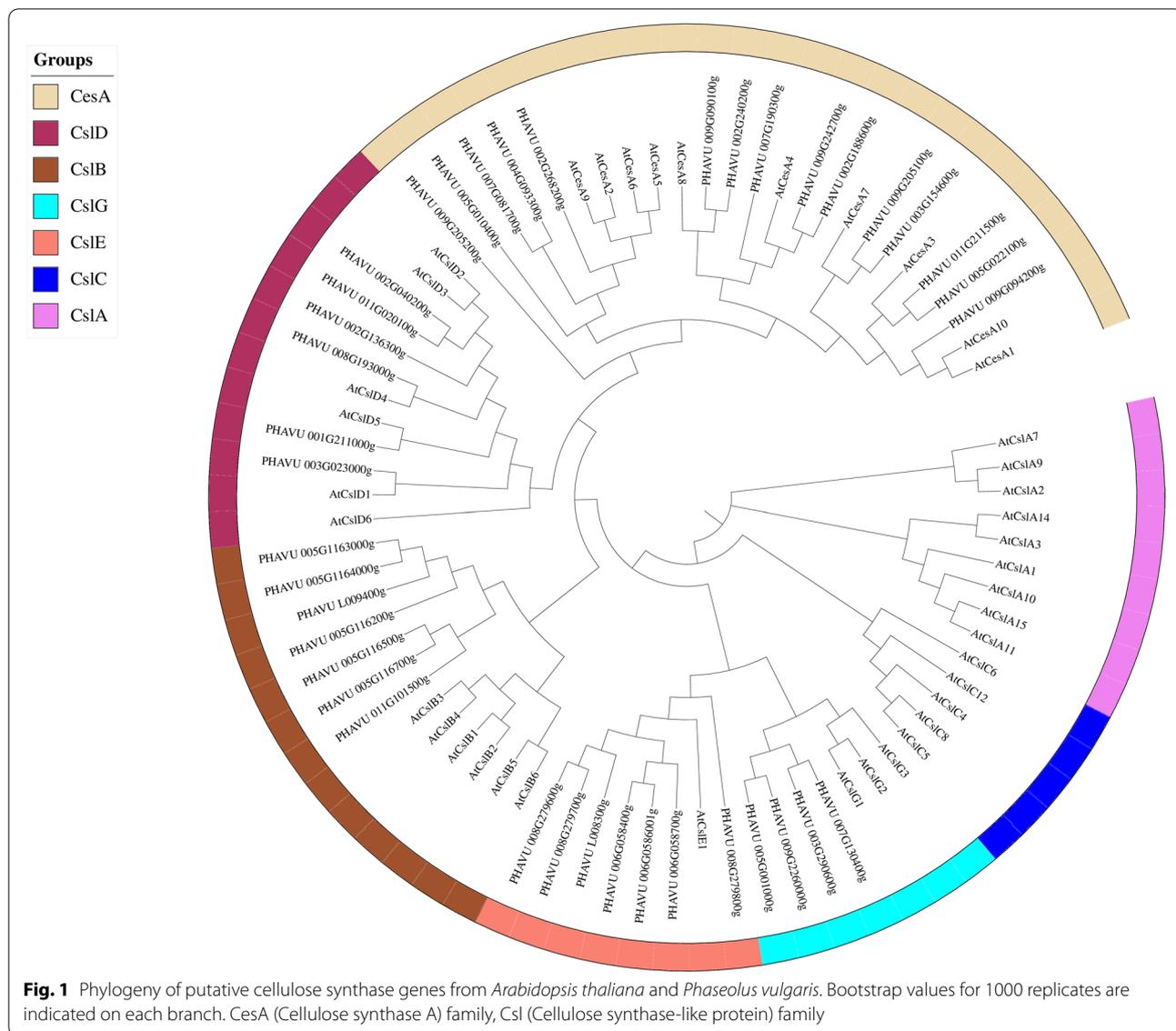Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 5 of 15

Fig. 1 Phylogeny of putative cellulose synthase genes from *Arabidopsis thaliana* and *Phaseolus vulgaris*. Bootstrap values for 1000 replicates are indicated on each branch. CesA (Cellulose synthase A) family, Csl (Cellulose synthase-like protein) family
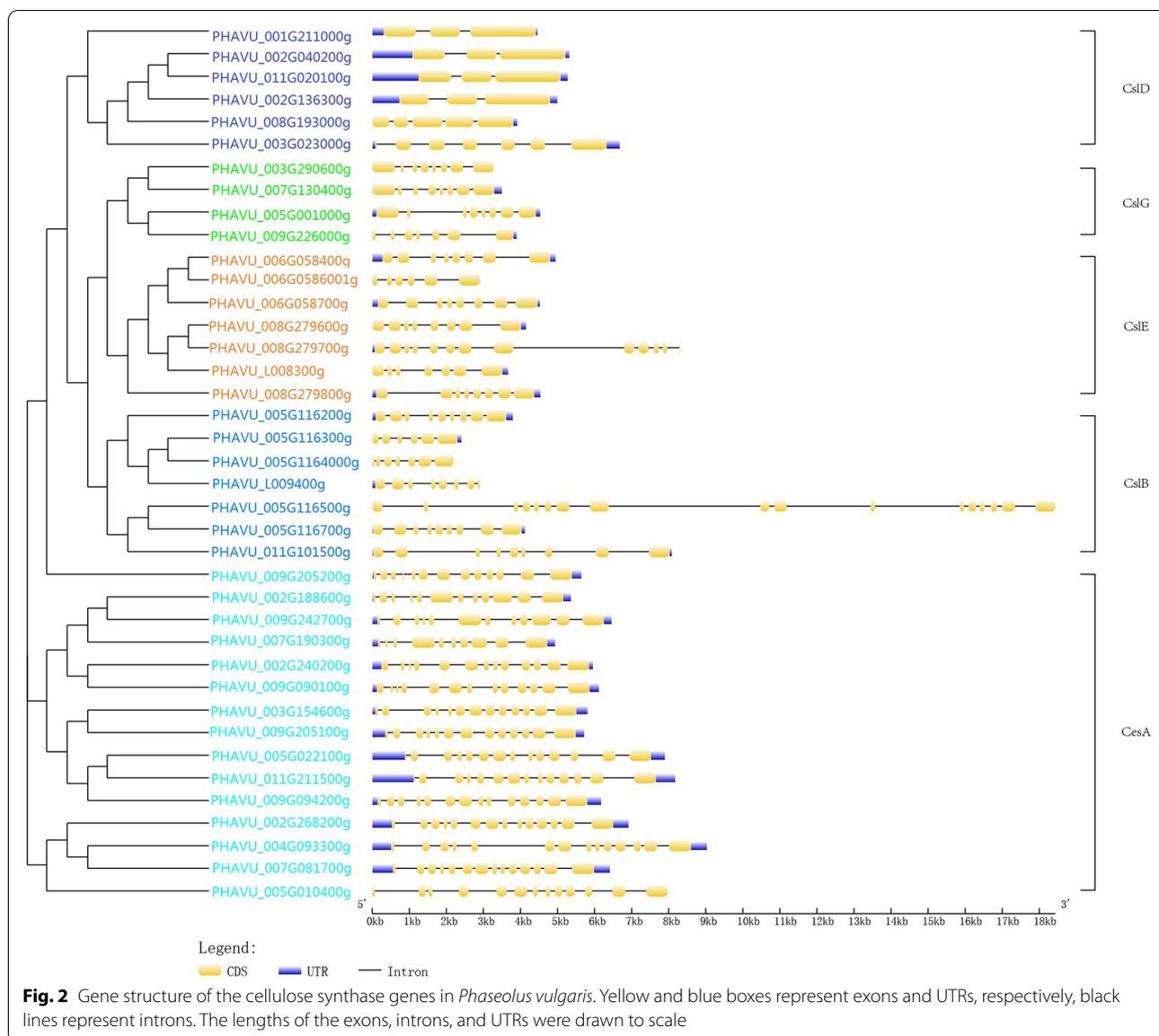
locations of the sequences on the chromsomes of common bean. As demonstrated in the location image, 39 *CesA/Csl* gene members are distributed on 10 chromosomes, and no genes mapping to chromosome 10 (Fig. 5). Chromosome 5 contained the largest number of cellulose synthase gene members (eight), followed successively by chromosome 9 (six), chromosome 2 (five), and Chromosomes 3, 6, 7, and 11 (three). Besides, we found 3 tandem duplication sets: PHAVU_005G1163000g (CslB)/ PHAVU_005G116500g (CslB)/ PHAVU_005G116200g (CslB)/ PHAVU_005G1164000g(CslB)/ P H A V U _ 0 0 5 G 1 1 6 7 0 0 g (C s l B) , P H A V U _ 0 0 6 G 0 5 8 4 0 0 g (C s l E) / P H A V U _ 0 0 6 G 0 5 8 7 0 0 g (C s l E) ,

P H A V U _ 0 0 8 G 2 7 9 7 0 0 g (C s l E) / PHAVU_008G279600g(CslE)/ PHAVU_008G279800g(CslE).

The comparative synteny relationship map of *Phaseolus vulgaris* revealed a high degree of similarity with *Glycine max* (Fig. 6A) and a low degree of similarity with *Arabidopsis thaliana (*Fig. 6B).

Expression Profiles of CesA/Csl genes in common bean pod development

To investigate the functions of *CesA/Csl* genes in common bean pod development, we used RT-qPCR with gene-specific primers (Table 2) to analyze the expression levels of *CesA/Csl* genes at three distinct pod

Fig. 2 Gene structure of the cellulose synthase genes in *Phaseolus vulgaris*. Yellow and blue boxes represent exons and UTRs, respectively, black lines represent introns. The lengths of the exons, introns, and UTRs were drawn to scale
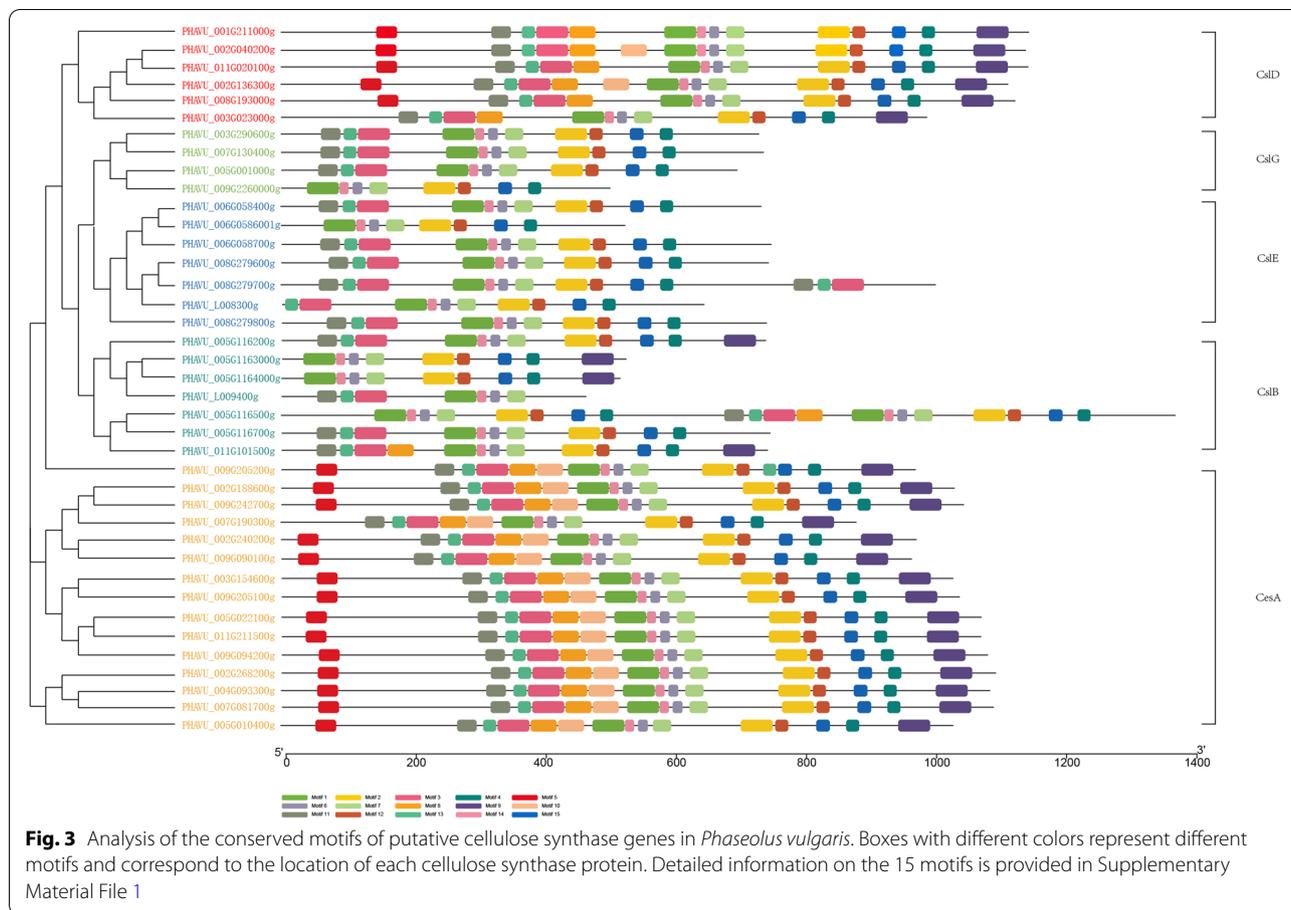
developmental stages. A total of 21 *CesA/Csl* genes included in this analysis were selected based on results from the sequence alignments, phylogenetic analysis, and gene structure analysis. All *CesA/Csl* genes (seven CesA, four CslD, four CslB, two CslG and four CslE) were expressed at all three stages of pod development, suggesting their important roles in the development of the pod in common bean (Fig. 7A, Fig. 7B).

The expression of 7 *CesA* genes were evaluated and the results showed that these genes in *CesA* subfamily diaplayed temporal variations in different pod development of common bean. The expression of 5 *CesA* genes showed a trend of first increasing in stage S2 and

then decreasing in stage S3 (*PHAVU_005G022100g, PHAVU_009G094200g, PHAVU_002G188600g, PHAVU_004G093300g* and *PHAVU_009G205200g*), while the expression of *PHAVU_003G154600g* showed an opposite trend (Fig. 7A). Moreover, the expression level of *PHAVU_007G190300g* decreased with the pod development (Fig. 7A)*.*

In *Csl* genefamily, we found that all *Csl* genes selected were expressed at all three stages of pod development. *PHAVU_001G211000g, PHAVU_011G020100g* in *CslD* subfamily, *PHAVU_005G020100g* in *CslB* subfamily showed similar expression pattern: increasing in the S2 stage and decreasing in the S3 stage, which
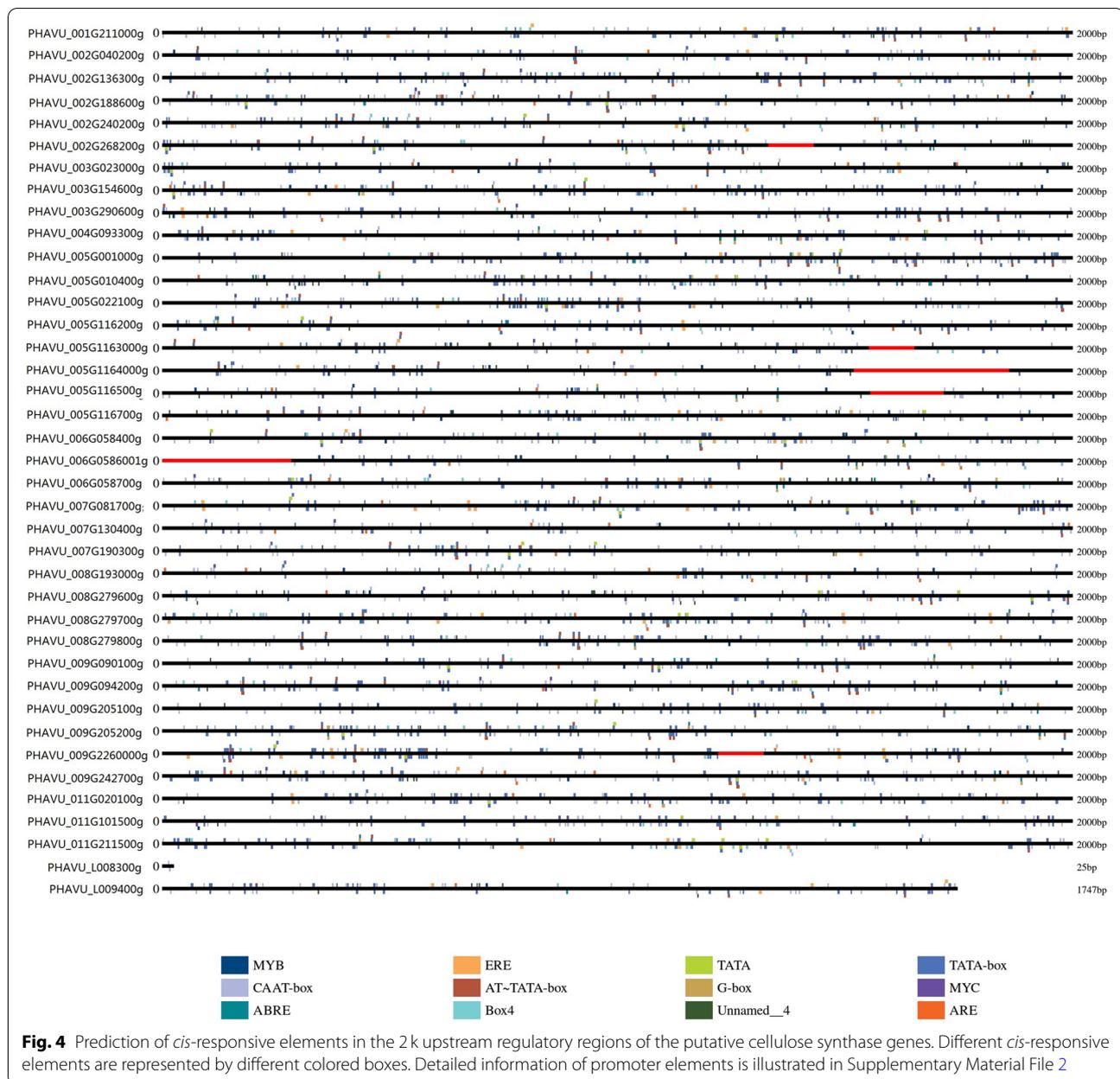
Fig. 3 Analysis of the conserved motifs of putative cellulose synthase genes in *Phaseolus vulgaris*. Boxes with different colors represent different motifs and correspond to the location of each cellulose synthase protein. Detailed information on the 15 motifs is provided in Supplementary Material File 1

is similar to the expression of *PHAVU_005G022100g, PHAVU_009G094200g, PHAVU_002G188600g, PHAVU_004G093300g and PHAVU_009G205200g* in *CesA* subfamily (Fig. 7B). *PHAVU_002G136300g* in *CslD* subfamily, *PHAVU_009G2260000g* in *CslG* subfamily and *PHAVU_L008300g* in *CslE* subfamily displayed similar expression pattern: decreasing in the S2 stage and slight increasing in the S3 stage. Whereas *PHAVU_008G279800g* and *PHAVU_006G058700g* in *CslE* subfamily showed similar expression trend, but the expression level of *PHAVU_008G279800g* and *PHAVU_006G058700g* in S3 stage significantly higher than the expression level in S1 stage. In addition, the expression level of *PHAVU_005G1163000g* in *CslB* subfamily decreased with the development of pod, the expression level of *PHAVU_011G101500g* in *CslB* subfamily and *PHAVU_003G290600g* in *CslG* subfamily increased only in the S3 stage, the expression level of *PHAVU_003G023000g* in *CslD* subfamily significantly increased in the S2 and S3 stages (Fig. 7B).

Discussion

Until now, the *CesA/Csl* gene family has been extensively characterized in many plant species, including Arabidopsis, barley, cotton, rice, sorghum, soybean [11, 28–32]. However, this gene family remain unidentified and uncharacterized in common bean. In this study, we conducted a genome-wide survey and identified 39 putative *CesA/Csl* genes in common bean genome (Fig. 1 and Table 1). This results coupled with the sequence alignment, phylogenetic analysis, gene structure construction, chromosome location and expression analysis, could provide important clues in understanding the roles of the *CesA/Csl* superfamily in in pod development in higher plants.

The *CesA/Csl* gene family found across plant species may be subcategorized into nine groups: CslA-CslH and CslJ [2, 33]. All land plants contain CslA, CslC, and CslD, while CslF and CslH are found only in grasses, and cereals do not usually contain CslB or CslG [34, 35]. Using phylogenetic analysis, the 41 *CesA/Csl* genes in *Arabidopsis* were categorized into one CesA group and six Csl groups (Csl A-E and G) [11]. In this study, the

Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 8 of 15

Fig. 4 Prediction of *cis*-responsive elements in the 2 k upstream regulatory regions of the putative cellulose synthase genes. Different *cis*-responsive elements are represented by different colored boxes. Detailed information of promoter elements is illustrated in Supplementary Material File 2

phylogenetic analysis showed that the 39 putative *CesA/Csl* genes in common bean could be classified into 5 subfamilies: CesA, CslD, CslB, CslG, and CslE (Fig. 1), consistent with studies of plants and algae [1, 35, 36]. From the phylogeny we can found that CslD is close to CesA, which is consistent with the earlier reports suggesting a common origin and conserved domains of this two families [37].

Among them, 15 putative Cellulose Synthase genes clustered into the CesA gene family, which was the most abundant genes among the 40 *CesA/Csl* genes, while the

remaining 24 genes clustered into the other 4 CesA/Csl subfamilies (Fig. 1), suggesting that they have experieced extensive expansion and diversification [33].

Investigation of gene structure and function lends a better understanding of the evolution of a gene family, revealing the divergence, conservation, or expansion of a given gene family [32, 38–40]. Similar to other plants, such as soybean [32] and tomato [31], most *CesA/Csl* genes (*CesA* and *CslD* members) share a similar gene structure in each subfamily (Fig. 2), suggesting that they are highly conserved. In contrast, members in
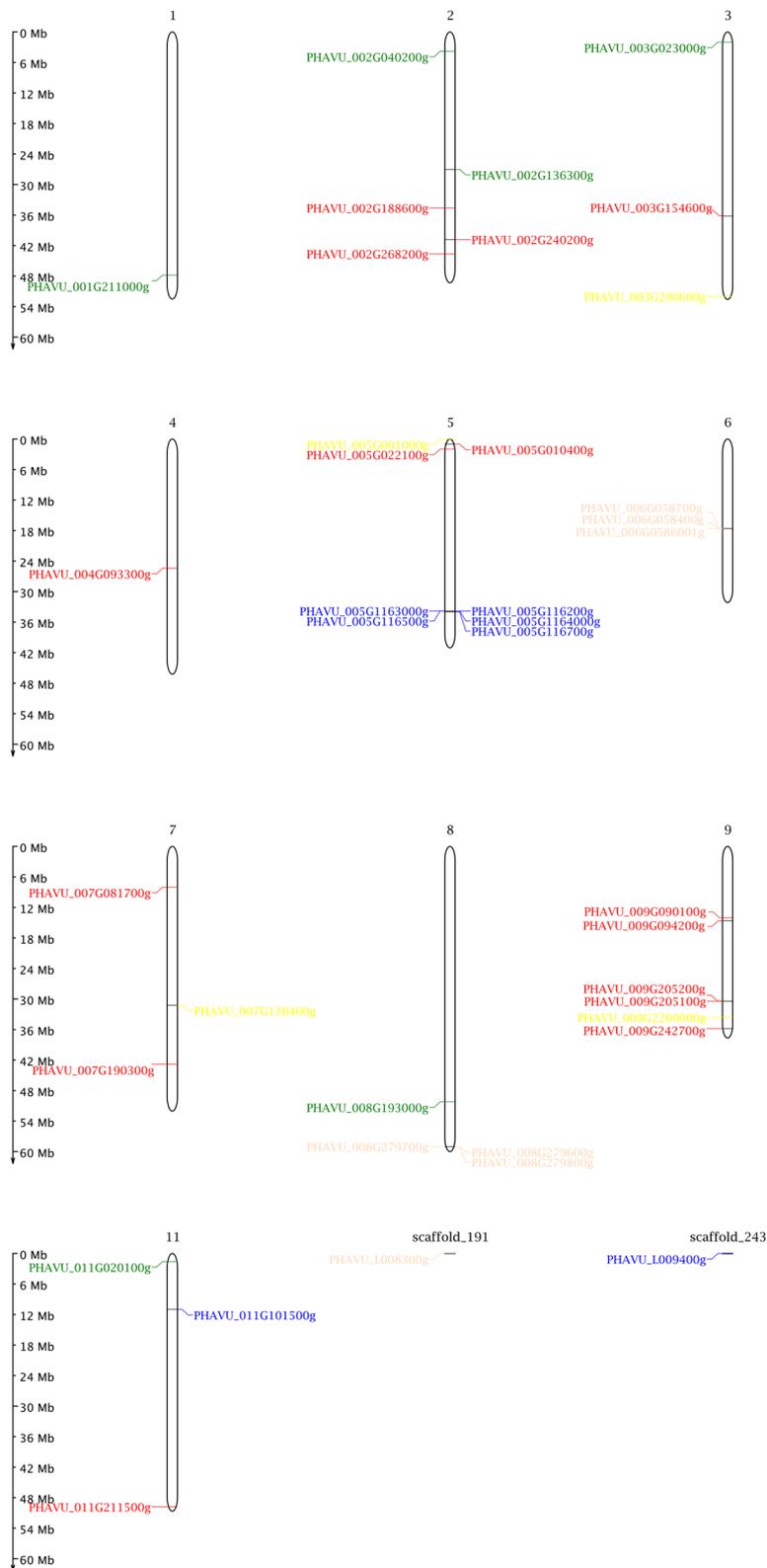
Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 9 of 15

Fig. 5 Chromosome map of the cellulose synthase gene family. Gene name color corresponds to gene family/sub family. i.e. green (CslD), red (CesA), yellow (CslG), blue (CslB), light pink (CslE). A scale on the left represents the length of chromosome in megabases (Mb)

Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 10 of 15

CslB, *CslG* and *CslE* subfamily exhibit variable gene structures possibly due to chromosome fusion and/or rearrangement [40]. Therefore, tandem or segmental duplication events in the *CesA/Csl* gene family have resulted in shared exon/intron structures and similar structural organization in each gene subfamily. Phylogenetic and domain analyses confirm these results. Chromosome mapping in this study further revealed that the tandem duplications also existed in *CesA/Csl* gene families (Fig. 5).

The *cis*-elements analysis detected a larger amount of *cis*-elements in the putative promoter regions of the *CesA/Csl* genes in common bean, which suggested that *CesA/Csl* genes might have potential roles in many signaling pathways.

CesA/Csl genes have been found to play an important role in plant cell walls in the biosynthesis of cellulose and hemicellulose [7, 16, 41]. During the pod development of common bean, the expression profiles of 21 CesA/Csl genes were revealed by RT-qPCR. The results showed that all 21 CesA/Csl genes were expressed in all three pod development stages, suggesting that all these genes are necessary for the pod growth. Most *CesA* genes in this study expressed highly in the young pod (S2 stage), which is in accordance with the results found in soybean [32], suggesting that these *CesA* genes may be involved in cellulose synthesis during the early pod development stage in common bean. We also found that 3 genes in *CslD* subfamily showed high expression level in the early pod development stage. And the expression level of genes in *CesA* and *CslD* subfamilies is higher than that of other *Csl* genes, which implies that gene members in *CesA* and *CslD* subfamilies are more actively involved in seed development than other *Csl* genes. Therefore, future investigation should aim to identify each *CesA/Csl* gene's function in common bean.

Conclusions

Based on the genomic data, 39 cellulose synthase genes were identified from common bean. The genes encoding these proteins were distributed unevenly on 10 chromosomes, and there were 3 tandem duplications. These 39 cellulose synthase proteins could be divided into five subfamilies according to their structure and phylogenetic relationship, members in the same subfamily share conserved exon-intron distribution and motif compositions. Based on the analysis of *cis*-element in the promoter

region, we found abundant and distinct *cis*-acting elements, which indicate that the cellulose synthase gene family plays a vital role in the growth and development of common bean. Additionally transcriptional analysis showed that 21 *CesA/Csl* genes selected were constitutively expressed in the pod development, *CesA/Csl* gene members in different groups showed different expression trend at three stages of pod development. In general, this study revealed a putative biochemical and physiological functions of cellulose synthase genes in common bean, which provides a foundation for further function identification of *CesA/Csl* gene family.

Methods

Identification of *CesA/Csl* gene family in common bean

The Hidden Markov Model (HMM), which established by 39 CesA/Csl protein sequences of Arabidopsis, was used to search the CesA/Csl gene family in common bean genomes at an e-value cutoff of $1e^{-10}$. To ensure genes identified with HMM model were accurate, further filtering of unique sequences was performed according to typical structural features of plant CesA/Csl proteins. The Phytozome 11.0 (https://phytozome.gji.doe.gov/) and ExPASy databases (https://web.expasy.org/compute_pi/) were used to obtain gene ID / name, chromosome location, peptide length and isoelectric point/molecular weight, and functial annotation information [42]. TMHMM v. 2.0 was used to predict the TMDs for each putative peptide (http://www.cbs.dtu.dk/services/TMHMM/).

Sequence and phylogenetic analyses

ClustalW was used to perform alignments of both CesA/Csl nucleotide and amino acid sequences. Including amino acid sequences of cellulose synthase proteins from Arabidopsis and common bean, the phylogenetic analysis was performed using a neighbor-joining tree method with 1000 bootstrap replicates in the software program MEGA5, which was also used to visualize the phylogenetic analysis. Protein subcellular location were analyzed by WoLF PSORT (http://psort.nibb.ac.jp).

(See figure on next page.)

Fig. 6 Synteny analysis of cellulose synthase gene family in *Phaseolus vulgaris* with the cellulose synthase gene family in *Glycine max* (**A**) and *Arabidopsis thaliana* (**B**). Synteny map was constructed by using online tool Circoletto: visualizing sequence similarity with Circos. Color variations represent the extent of similarity and homology between genes based on bit score. Detailed information of the sequences is provided in Supplementary Material File 3, 4, 5
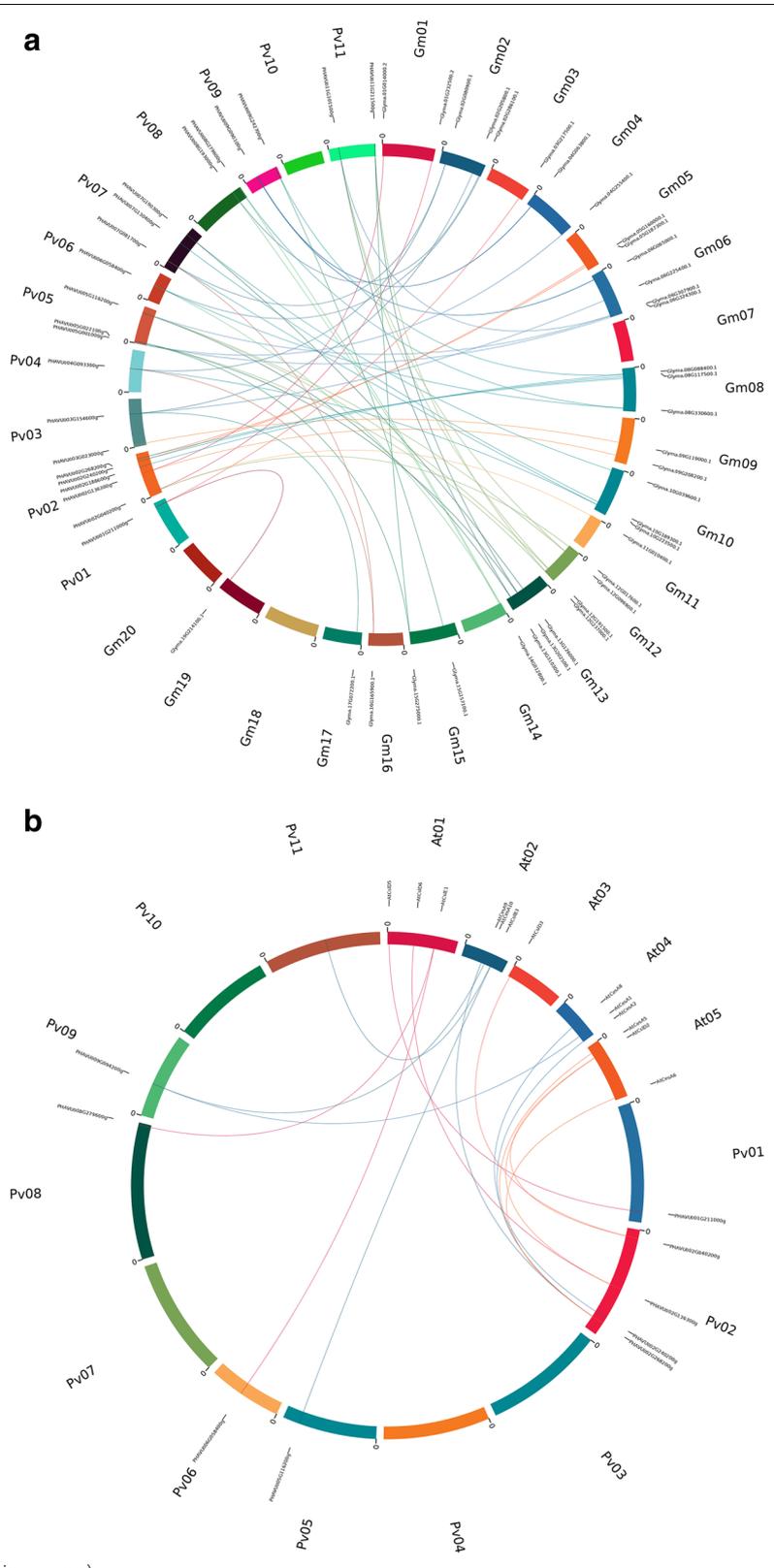
Fig. 6 (See legend on previous page.)

Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 12 of 15

Table 2 Gene-specific DNA primers for qPCR

Gene_name	Primer sequence (5′ –3′)
PHAVU_005G022100g	F-TGAGGTGGAGTGGTGTTGGA; R-GGAGGGATGAGGAGGGTTGT
PHAVU_009G094200g	F-AAGAAGAGGGCGATGAACAGA; R-TCCATGAAAGTGGCAGCAATA
PHAVU_003G154600g	F-CTGATGACGGAGCTTCAATGTG; R-TGCACGAGGTTCGATAGAAAA
PHAVU_007G190300g	F-GTGGGAATTGTGGCTGGAA; R-TGCTGGACCTGTCTGCTTGG
PHAVU_004G093300g	F-TCCTTGTTGATCCCTCCCTT; R-CCTGTTTACCCATGACACCCT
PHAVU_009G205200g	F-AGGCAGAACAGAACACCAACTC; R-TTCCACATTGTTTGGCATCAG
PHAVU_003G023000g	F-GGCAGGAGGATCAGAACACTT; R-CGGACAACCACCAACAACC
PHAVU_001G211000g	F-GACGAGGAAGCAATGAAAGGC; R-ATGGAAGGCAGAGGCAGAGG
PHAVU_002G136300g	F-CCCTTGTCATCCTTGCTGTTT; R-TCAATGCCAGCCATCACCT
PHAVU_011G020100g	F-ATTCGGTTGGTTGTCCTGGTA; R-CAGTAGGATTGTTGGGACTTCG
PHAVU_005G1163000g	F-GGAAATGGGAGTGGAATAGGA; R-GATGACATGCAAATGGTGGTTA
PHAVU_L009400g	F-CCGTAACTCACCCAGATCGTC; R-GGCTTCCACAAGGGCATAGA
PHAVU_011G101500g	F-CAATTTCTTGCCACAGGAGC; R-AACCAGCATTCATGGGTGTTAT
PHAVU_005G116500g	F-ATCAAGACGGGACGTGACAGA; R-TGACATCAAGCGGGTTATCG
PHAVU_003G290600g	F-TTGCCGGAGGATCACAAAC; R-CCATGCAGGGTCAAAGGAGA
PHAVU_009G2260000g	F-TGCCAATGGCGGTTTATG; R-GCAACATCTTTGGAGGTTTCAG
PHAVU_008G279800g	F-TTGCACCTATGAGGAGGGC; R-CCCATAAATGAAAGGGCAATAC
PHAVU_006G058400g	F-AACACCTTGCCACAAGCACTA; R-GCCAGGAAAGAAACACCCATA
PHAVU_006G058700g	F-TCAAAGGGTGAAGATGGCAAAT; R-TGGTGGCTCAATATCAGGGTC
PHAVU_L008300g	F-CAACTTCTTTGCCATCACTTCC; R-CCTCCCATACTCAAACCCTCAA
PHAVU_006G029200g	F-GGAACGAAGTGGTGGAATGG; R-AAGAATGACAAAGTGGGAGGC
IDE	F- GCAACCAACCTTTCATCAGC; R -AGAAATGCCTCAACCCTTTG

Gene structure, motif identification and chromosome localization

Gene Structure Display Server v2.0 (http://gsds.cbi.pku.edu.cn/index.php) was used to analyze the exon-intron structure of these genes [43]. MEME program (http:// meme.sdsc.edu/meme/cgi-bin/meme.cgi) was employed to analyze the protein sequences for the confirmation of the motifs. InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/) was used to annotated the motifs.

The chromosome distribution of all cellulose synthase genes of common bean was identified, and the location of CesA/Csl genes was drafted with MapChart v2.0 [44]. Phytozome 11.0 Network database (https://phytozome.jgi.doe.gov/) was also used to obtain genomic DNA and complementary DNA (cDNA) sequences of the putative cellulase synthase genes used in this analysis.

Cis-Element analysis of putative promoter regions and synteny analysis

Using the Phytozome 11.0 network database (https://phytozome.jgi.doe.gov/), 2 Kbp regulatory regions upstream from the start site of translation of CesA/Csl genes were retrieved. Then, the PlantCARE online was used to investigate the putative *cis*-regulatory elements in these promoter region sequences. The location of *cis* elements was annotated and displayed in a figure by building a physical gene map using a Perl and Scalable Vector Graphics (SVG) script. Finally, syntenic relationships of all *CesA/Csl* were analyzed using Circoletto [45].

Expression analysis of *CesA/Csl* gene family in *Phaseolus vulgaris*

Bean seeds were grown in pots in the open-air soils. Pods were harvested at different stages: 7 (S1), 14 (S2), and 21 (S3) days after flowering. Then, the total RNA of these pods was isolated using the Promega Plant RNA Kit (Promega, Beijing, China) according to the manufacture's instructions. Single-stranded cDNA was synthesized using 2 μg of total RNA and Oligd(T)18 primer with the Takara RT-PCR system in a total volume of 25 μl according to the protocol. Three independent PCR reactions were carried out for the 63 putative genes using SYBR Green Supermix (Takara) according to the manufacturer's protocol in an ABI 7500 Real-time system (ABI, CA, USA). IDE (insulin degrading enzyme) was used as an internal control to normalize the expression of CesA/
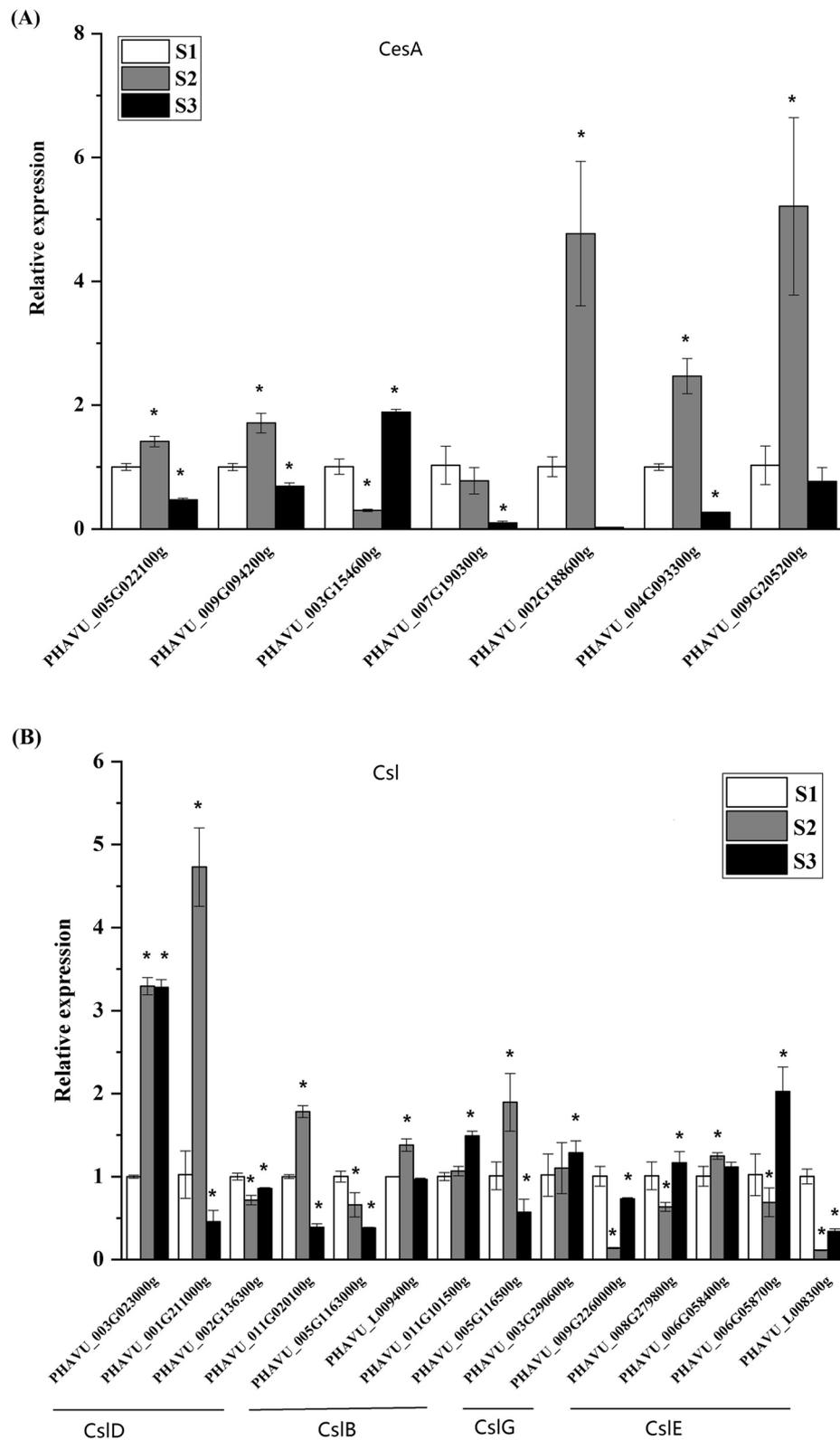
Fig. 7 Expression analysis of 6 *CesA* (**A**), 4 *CslD*, 4 *CslB*, 2 *CslG*, 4 *CslE*, (**B**) genes at three seed developmental stages of *Phaseolus vulgaris*. To visualize the relative expression levels data, S1 stage was normalized as "1", and data are means ± SD calculated from three biological replicates. * indicate significant differences in comparison with the control at $p < 0.05$

Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 14 of 15

Csl genes according to Borges [46]. Gene-specific DNA primers for qPCR are listed in Table 2.

Statistical Analysis

Statistical analyses were performed in Excel and SPSS v17.0 (link/cite SPSS.) The significance threshold between samples was $p < 0.05$, and all results of expression data were indicated as averages ±standard deviations (SDs).

Abbreviations

CesA: Cellulose synthase A; Csl: Cellulose synthase-like.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-022-01026-0.

Additional file 1. Supplementary data associated with this article can be found in the appendix. The information of 20 conserved motifs of Cellulose synthase genes in Phaseolus vulgaris detected by the online tool MEME.

Additional file 2 The information of cis-elements in promoter sequences of Cellulose synthase genes in *Phaseolus vulgaris*.

Additional file 3 The protein sequences of Cellulose synthase genes in *Phaseolus vulgaris*.

Additional file 4 The protein sequences of Cellulose synthase genes in *Glycine max*.

Additional file 5 The protein sequences of Cellulose synthase genes in *Arabidopsis thaliana*.

Additional file 6 The raw data of characterization of the putative Cellulose Synthase genes in *Phaseolus vulgaris.*

Acknowledgements

Not applicable.

Authors' contributions

X.L. designed and executed the experimental approach. H.Z., W.Z., W.X. and S.L. conducted the data curation and formal analysis. X.L. drafted the manuscript that was subsequently edited by H.C. and X.C. The author(s) read and approved the final manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study provided in Supplementary Material File 6.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

References

1. Richmond TA, Somerville CR. The cellulose synthase superfamily. Plant Physiol. 2000;124(2):495–8. https://doi.org/10.1104/pp.124.2.495.
2. Hazen SP, Scott-Craig JS, Walton JD. Cellulose synthase-like genes of rice. Plant Physiol. 2002;128(2):336–40. https://doi.org/10.1104/pp.010875.
3. Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM. Higher plants contain homologs of the bacterial CsIA genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci. 1996;93(22):12637–42. https://doi.org/10.1073/pnas.93.22.12637.
4. Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, et al. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science. 1998;279:717–20. https://doi.org/10.1126/science.279.5351.717.
5. Yin Y, Huang J, Xu Y. The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biol. 2009;9:99–113. https://doi.org/10.1186/1471-2229-9-99.
6. Appenzeller L, Doblin M, Barreiro R, Wang H, Ni X, Kollipara K, et al. Cellulose synthesis in maize isolation and expression analysis of the cellulose synthase (CesA) gene family. Cellulose. 2004;11(3–4):287–99. https://doi.org/10.1023/B:CELL.0000046417.84715.27.
7. Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB. The CESA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 2004;134:224–36. https://doi.org/10.1104/pp.103.032904.
8. Kaur S, Dhugga KS, Gill K, Singh J. Novel structural and functional motifs in cellulose synthase (CesA) genes of bread wheat (*Triticum aestivum* L.). PLoS One. 2016;11:e0147046. https://doi.org/10.1371/journal.pone.0147046.
9. Li S, Lei L, Gu Y. Functional analysis of complexes with mixed primary and secondary cellulose synthases. Plant Signal Behav. 2013;8(3):e23179. https://doi.org/10.4161/psb.23179.
10. Taylor NG, Laurie S, Turner SR. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell. 2000;12(12):2529–40. https://doi.org/10.1105/tpc.12.12.2529.
11. Persson S, Paredes A, Carroll A, Palsdottir H, Doblin M, Poindexter P, et al. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. Proc. Natl. Acad. Sci. 2007;104(39):15566–71. https://doi.org/10.1073/pnas.0706592104.
12. Zou XY, Zhen Z, Ge Q, Fan SM, Liu AY, Gong WK, et al. Genome-wide identification and analysis of the evolution and expression patterns of the cellulose synthase gene superfamily in *Gossypium* species. Gene. 2018;646:28–38. https://doi.org/10.1016/j.gene.2017.12.043.
13. Fincher GB. Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. Plant Physiol. 2009;149(1):27–37. https://doi.org/10.1104/pp.108.130096.
14. Doblin MS, Pettolino F, Bacic A. Plant cell walls: the skeleton of the plant world. Funct Plant Biol. 2010;37(5):357–81. https://doi.org/10.1071/FP09279.
15. Goubet F, Barton CJ, Mortimer JC, Yu X, Dupree P. Cell wall glucomannan in *Arabidopsis* is synthesized by CSLA glycosyltransferases, and influences the progression of embryogenesis. Plant J. 2009;60(3):527–38. https://doi.org/10.1111/j.1365-313X.2009.03977.x.
16. Liepman AH, Wilkerson CG, Keegstra K. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CsIA family members encode mannan synthases. Proc Natl Acad Sci. 2005;102(6):2221–6. https://doi.org/10.1073/pnas.0409179102.
17. Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, et al. Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. Science. 2004;303(5656):363–6. https://doi.org/10.1126/science.1090908.
18. Liepman AH, Cavalier DM. The Cellulose synthase-like A and cellulose synthase-like C families: recent advances and future perspectives. Front Plant Sci. 2012;3:109. https://doi.org/10.3389/fpls.2012.00109.
19. Dwivany FM, Yulia D, Burton RA, Shirley NJ, Wilson SM, Fincher GB, et al. The cellulose-synthase like c (CSLC) family of barley includes members that are integral membrane proteins targeted to plasma membrane. Mol Plant. 2009;2(5):1025–39. https://doi.org/10.1093/mp/ssp064.

Liu *et al. BMC Genomic Data*   (2022) 23:9

Page 15 of 15

20. Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, et al. A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. Proc Natl Acad Sci. 2007;104(20):8550–5. https://doi.org/10.1073/pnas.0703133104.

21. Bernal AJ, Jensen JK, Harholt J, Sørensen S, Moller I, Blaukopf C, et al. Disruption of AtCslD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in *Arabidopsis*. Plant J. 2007;52(5):791–802. https://doi.org/10.1111/j.1365-313X.2007.03281.x.

22. Bernal AJ, Yoo CM, Mutwil M, Jensen JK, Hou GC, Blaukopf C, et al. Functional analysis of the cellulose synthase-like genes CSLD1, CSLD2, and CSLD4 in tip-growing *Arabidopsis* cells. Plant Physiol. 2008;148(3):1238–53. https://doi.org/10.1104/pp.108.121939.

23. Li M, Xiong G, Li R, Cui J, Ding T, Zhang B, et al. Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. Plant J. 2009;60(6):1055–69. https://doi.org/10.1111/j.1365-313X.2009.04022.x.

24. Verhertbruggen Y, Yin L, Oikawa A, Scheller HV. Mannan synthase activity in the CslD family. Plant Signal Behav. 2011;6(10):1620–3. https://doi.org/10.4161/psb.6.10.17989.

25. Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, et al. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. Science. 2006;311(5769):1940–2. https://doi.org/10.1126/science.1122975.

26. Burton RA, Jobling SA, Harvey AJ, Shirley NJ, Mather DE, Bacic A, et al. The genetics and transcriptional profiles of cellulose synthase-like HvCslF gene family in barley. Plant Physiol. 2008;146(4):1821–33. https://doi.org/10.1104/pp.107.114694.

27. Schmutz J, McClean PE, Mamidi S, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46(7). https://doi.org/10.1038/ng.3008.

28. Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, Platzer M, et al. A physical, genetic and functional sequence assembly of the barley genome. Nature. 2012;491:711–6. https://doi.org/10.1038/nature11543.

29. Li Y, Yang T, Dai D, Hu Y, Guo X, Guo H. Evolution, gene expression profiling and 3D modeling of CSLD proteins in cotton. BMC Plant Biol. 2017;17:119. https://doi.org/10.1186/s12870-017-1063-x.

30. Shu O, Wei Z, John H, Lin H, Matthew C, Kevin C, et al. The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 2007;35:D883–7. https://doi.org/10.1093/nar/gkl976.

31. Song XM, Xu L, Yu JW, Tian P, Hu X. Genome-wide characterization of the cellulose synthase gene superfamily in *Solanum lycopersicum*. Gene. 2019;688:71–83. https://doi.org/10.1016/j.gene.2018.11.039.

32. Nawaz MA, Rehman HM, Baloch FS, Ijaz B, Ali MA, Khan IA, et al. Genome and transcriptome-wide analyses of cellulose synthase gene superfamily in soybean. J Plant Physiol. 2017;215:163–75. https://doi.org/10.1016/j.jplph.2017.04.009.

33. Schwerdt JG, MacKenzie K, Wright F, Oehme D, Wagner JM, Harvey AJ, et al. Evolutionary dynamics of the cellulose synthase gene superfamily in grasses. Plant Physiol. 2015;168:968–83. https://doi.org/10.1104/pp.15.00140.

34. Farrokhi N, Burton RA, Brownfield L, Hrmova M, Wilson SM, Bacic A, et al. Plant cell wall biosynthesis: Genetic, biochemical and functional genomics approaches to the identification of key genes. Plant Biotechnol J. 2006;4(2):145–67. https://doi.org/10.1111/j.1467-7652.2005.00169.x.

35. Yin Y, Johns MA, Cao H, Rupani M. A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily. BMC Genomics. 2014;15:260. https://doi.org/10.1186/1471-2164-15-260.

36. Suzuki S, Li L, Sun YH, Chiang VL. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. Plant Physiol. 2006;142(3):1233–45. https://doi.org/10.1104/pp.106.086678.

37. Ermawar RA, Collins HM, Byrt CS, Henderson M, Donovan LAO, Shirley NJ, et al. Genetics and physiology of cell wall polysaccharides in the model C4 grass, Setaria viridis spp. BMC Plant Biol. 2015;15:236. https://doi.org/10.1186/s12870-015-0624-0.

38. Betts MJ, Guigo R, Agarwal P, Russell RB. Exon structure conservation despite low sequence similarity: a relic of dramatic events in evolution? EMBO J. 2001;20:5354–60. https://doi.org/10.1093/emboj/20.19.5354.

39. Worberg A, Quandt D, Barniske AM, Löhne C, Hilu KW, Borsch T. Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. Org Divers Evol. 2008;7:55–77. https://doi.org/10.1016/j.ode.2006.08.001.

40. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci. 2012;109:1187–92. https://doi.org/10.1073/pnas.1109047109.

41. Liepman AH, Wightman R, Geshi N, Turner SR, Scheller HV. *Arabidopsis* - a powerful model system for plant cell wall research. Plant J. 2010;61:1107–21. https://doi.org/10.1111/j.1365-313X.2010.04161.x.

42. Artimo P. ExPASy: SIB bioinformatics resource portal. Nucl. Acids Res. 2012;40(Web Server issue):W597–603. https://doi.org/10.1093/nar/gks400.

43. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: An upgraded gene feature visualization server. Bioinformatics. 2015;31:1296–7. https://doi.org/10.1093/bioinformatics/btu817.

44. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002;93:77–8. https://doi.org/10.1093/jhered/93.1.77.

45. Darzentas N. Circoletto: visualizing sequence similarity with circos. Bioinformatics. 2010;2010(26):2620–1. https://doi.org/10.1093/bioinformatics/btq484.

46. Borges A, Tsai SM, Caldas DGG. Validation of reference genes for RT-qPCR normalization in common bean during biotic and abiotic stresses. Plant Cell Rep. 2012;31:827–38. https://doi.org/10.1007/s00299-011-1204-x.

Publisher's Note