Genome-wide Identification, Expression, and Sequence Analysis of CONSTANT-like Gene Family in Cannabis Reveals Role in Plant Flowering Time Regulation

Gen Pan
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Zheng Li
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Ming Yin
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Siqi Huang
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Jie Tao
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Anguo Chen
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Jianjun Li
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Huijuan Tang
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Li Chang
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Yong Deng
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Defang Li
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Lining Zhao (csbtzln@163.com)
Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205 China

---

Research article

Keywords: cannabis, genome-wide, CONSTANS-like gene (COL), expression pattern, flowering time

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-123512/v1

Version of Record: A version of this preprint was published on March 17th, 2021. See the published version at https://doi.org/10.1186/s12870-021-02913-x.

Abstract

Background: Cannabis, an important industrial crop, has a high sensitivity to photoperiods. The flowering time of cannabis is one of its important agronomic traits, and has an important effect on its yield and quality. The *CONSTANS-like* (*COL*) gene plays a key role in the regulation of flowering in this plant. However, the specific biological and functional roles of the *COL* gene family in the cannabis is still unknown.

Results: In this study, 13 *CsCOL* genes were identified in the cannabis genome. Phylogenetic analysis implied that the CsCOL proteins were divided into three subgroups, and each subgroups included conserved intron/exon structures and motifs. Chromosome distribution analysis showed that 13 *CsCOL* genes were unevenly distributed on 7 chromosomes, with chromosome 10 having the most *CsCOL* members. Colinearity analysis showed that two syntenic gene pairs of *CsCOL4* and *CsCOL11* were found in both rice and *Gossypium raimondii*. Among 13 *CsCOLs*, *CsCOL6* and *CsCOL12* were a pair of tandem duplicated genes, whereas *CsCOL8* and *CsCOL11* may have resulted from segmental duplication. Furthermore, tissue-specific expression showed that ten *CsCOL* genes were preferentially expressed in the leaves, one *CsCOL* in the stem, and two *CsCOL* in the female flower. Most *CsCOL* exhibited a diurnal oscillation pattern under different light treatment. Additionally, sequence analysis showed that *CsCOL3* and *CsCOL7* exhibited amino acid difference among the early-flowering cultivars and late flowering cultivars.

Conclusion: This study provided insight into the potential functions studies of *CsCOL* genes, and highlight its roles in the regulation of flowering time in cannabis. Our results laid a foundation for the further elucidation of the functions of *COLs* in cannabis

Background

Hemp is an ancient economic crop that is widely used in textiles, food and building materials, among other fields [1]. In recent years, the use of cannabidiols, represented by CBD, has been expanding continuously, and the cannabis industry has shown good prospects for development in the future [2]. Hemp is an annual short-day crop that is sensitive to photoperiods [3]. Cannabis cultivars natural grow in high-latitude areas. However, cannabis germplasms have been introduced to low-latitude areas for planting, resulting in early-flowering times. As a result, the growth period has been shortened, seriously reducing the yield and content of cannabinoid (CBD) and fibre [3]. Thus, the development of cannabis varieties with a wide adaptability is one of the main goals of current cannabis breeding programs. Identifying the regulatory mechanism of cannabis flowering could provide a theoretical foundation for the cultivation of cannabis varieties. However, studies on the regulatory mechanism of flowering in cannabis are currently lacking.

The flowering period of plants is a complex quantitative trait that is comprehensively regulated by many internal and external factors, including the photoperiod, temperature, hormones and self-development [4].

Among these factors, the photoperiod is an important regulatory factor of the floral transition. In agriculture, the flowering time of cultivated plants can be adjusted to meet consumer demand by changing the length of exposure to light. With rapid advances in the fields of molecular genetics and molecular biology, many genes related to photoperiod pathway have been discovered and cloned [5–6]. Studies have shown that *CONSTANT-like (COL)* is an important regulator of the plant response to photoperiods and is a core element in the regulation of plant flowering [7–9]. *COL* belongs to the zinc finger transcription factor family, which contains a B-box-type and a CCT (CO, CO-LIKE, TOC1) domain [10]. According to the number of B-box and CCT domains, *COL* family genes can be divided into five groups [11]. In previous reports, *COL* gene family have been comprehensively studied in many plants, including Arabidopsis, rice, maize, *Populus*, radish, moso bamboo and Lilium × formolongi [5–6, 10–14]. The number of *COL* genes varies among different species. For example, among dicot plants, the *COL* family has 20 members in radish, 17 members in Arabidopsis, while in monocots, 16 members have been identified in rice, 19 members in maize, and 14 members in *Populus* [12–16].

The *COL* gene functions as a transcription factor in multiple growth and development pathways, and particularly in the photoperiod-mediated flowering pathway. Some genes in this family have been found to play an important role in the light response-mediated regulation of flowering [5, 17–19], with functions that differ between short-day (SD) and long-day (LD) conditions. For example, *OsCOL10, OsCOL13* and *OsCOL16* function as negative regulators of flowering under both SD and LD condition in rice, while *Hd1*, a member of the *COL* gene family, promotes flowering under SD and suppresses flowering under LD [5, 17, 20–21]. In Arabidopsis, the overexpression of *AtCOL3*, *AtCOL7* and *AtCOL8* can delay flowering time, while, in the contrast, the overexpression of the *AtCOL5* gene promotes flowering by enhancing the expression of *FLOWERING LOCUS T* (*FT*) [18–19, 22–23]. Similar to their functions, the expression patterns also vary among the members of the *COL* gene family. In bananas, *MaCOLs* display a higher expression in light than in darkness, reaching their peak during light periods [24]. The transcript levels of *PaCOL1* and *PaCOL2*, two members of COL gene family, in Norway spruce are induced by light and increase upon transition from darkness to light [25]. Unlike *PaCOL1* and *PaCOL2*, *PttCO1* and *PttCO2* showed a distinct expression pattern with an increase in expression in the early night [26]. Meanwhile, the differences in the sequences of these genes in the CDS region were reported to associated with their functions in the photoperiod-mediated flowering pathway. For example, the deletion of 2 bp in the second exon of *Hd1* in “Kasalath” resulted in delay of flowering time in rice accessions [20]. Similarly, nucleotide polymorphisms in the *OsCOL16* coding sequence were mainly composed of three alleles (A1, A15 and A22), the flowering time of which varied [5]. These studies indicated that, due to differences in the expression patterns and CDS sequences, the *COL* gene family exerts multiple functions in the regulation of flowering time under SD and LD conditions.

Although *COL* genes play an important role in the growth and development of many plants, a comprehensive analysis of the *COL* family genes in cannabis is currently lacking. In addition, no systemic analyses of any other gene families in cannabis have been conducted due to the unavailability of cannabis genome assembly with a lack of information on gene locations at the chromosome level. The genome of cannabis was recently sequenced and made available on the cannabis genomic database

[27], allowing for a comprehensive analysis of the *COL* gene family in cannabis. In the present study, the *COL* gene family from cannabis was analysed using bioinformatics, and the temporal and spatial expression patterns of the *COL* gene were studied. Additionally, differences in the amino acid (aa) sequences of *CsCOL3* and *CsCOL7* between a early-flowering and late-ripen cultivars were explored. Thus, the results presented in this study provide a biological basis for further studies to analyze the molecular function of the *CsCOL* gene family in cannabis.

Results

Identification of 13 CsCOL gene family in cannabis

A total of 13 *CsCOL* were identified from the cannabis genome database (*CsCOL1* to *CsCOL13*). The 13 CsCOLs included both B-box and CCT conserved domains. Their physicochemical properties were analysed using ProtParam (http://web.expasy.org/protparam/) (Table 1). As shown in Table 1, the length of CsCOL proteins varied from 184 (CsCOL8) to 507 (CsCOL12) aa, molecular weights ranged from 26.02 kDa to 56.24 kDa, and pI varied from 4.99–6.36. In addition, the grand average of hydropathicity varied from –1.088 to –0.245 and the aliphatic index ranged from 38.10 to 69.90 (Table 1).

Table 1
Protein characteristics of 13 CsCOL in cannabis

Gene	Gene ID	Length (aa)	MW (Da)	pI	GRAVY	Aliphatic index
CsCOL1	LOC115722074	375	42,446.16	5.82	-0.766	61.63
CsCOL2	LOC115714019	462	51,911.74	5.42	-0.687	61.00
CsCOL3	LOC115697429	443	48,003.43	5.26	-0.563	61.87
CsCOL4	LOC115725326	387	41,576.35	5.08	-0.245	69.90
CsCOL5	LOC115714839	507	56,248.68	5.95	-0.752	60.95
CsCOL6	LOC115711648	456	52,718.08	5.52	-1.069	58.62
CsCOL7	LOC115707536	337	36,832.04	5.89	-0.555	64.04
CsCOL8	LOC115703699	184	20,459.75	4.30	-1.167	38.10
CsCOL9	LOC115700798	407	43,842.56	4.89	-0.489	60.96
CsCOL10	LOC115700744	394	42,808.34	5.30	-0.610	58.43
CsCOL11	LOC115700712	241	26,026.97	4.32	-0.741	55.85
CsCOL12	LOC115711183	420	48,419.95	5.23	-1.088	59.00
CsCOL13	LOC115700838	398	43,959.03	5.19	-0.513	58.84
MW: molecular weight; pI: isoelectric point						

Gene structure, multiple alignment, and phylogenetic relationship analysis of CsCOL genes

To estimate the evolutionary relationships between the members of the *CsCOL* gene family, we investigated the structure diversity by comparing the gene structure of the CsCOL protein. As showed in Fig. 1, the number of exons of the *CsCOL* genes and introns varied, ranging from 2 to 5 and from 1 to 4, respectively. All of the *CsCOL* genes contained 3' and 5' UTR regions. Phylogenetic tree analysis was conducted using the full-length CsCOL protein. As a result, 13 CsCOL proteins were divided into three groups: I, II, and III. Among these, group I and II included the most members (5), while group III only contained three members (Fig. 1). Furthermore, to explore the evolutionary relationships between the *COL* gene of different species, multiple protein sequence alignment was conducted with COL proteins from different plants, including Arabidopsis, cannabis and rice. These included 30 genes from dicotyledonous plants (e.g. Arabidopsis and cannabis) and 14 genes from monocotyledonous plants (e.g. rice). The results revealed that these COL proteins could be clustered into three major groups, namely groups I-III (Fig. 2). Group III was the smallest subfamily, it was comprised of the lowest number of COL proteins (Fig. 2).

Chromosomal location and synteny analysis

As shown in Fig. 3, except for chromosomes 5, 6 and 7, the 13 *CsCOL* gene members were found to be unevenly distributed across 7 chromosomes of the cannabis genome. Among them, chromosome 10 had the highest number of *CsCOL* genes (4), while chromosomes 2, 8, and 9 only contained one member. Interestingly, a pair of tandem replication genes were identified in chromosome 3 (*CsCOL6/CsCOL12*), suggesting that tandem duplication events participated in the expansion of the COL family in cannabis. As such, duplication events were investigated in the *CsCOL* genes of the cannabis genome. As a result, only one pair of duplicated genes (*CsCOL8/CsCOL11*) was identified within the cannabis genome, which may have resulted from segmental duplication or whole genome duplication (WGD) (Fig. 4). In order to further understand the evolutionary mechanism of the COL family in cannabis, collinearity diagrams of the COL family were constructed in two monocotyledonous plants (*Gossypium raimondii* and *Cannabis sativa* L.) and one monocotyledonous plant (*Oryza sativa* L). As shown in Fig. 5, 15 pairs of orthologous genes were identified between cannabis and cotton Raymond, much greater than those identified between cannabis and rice (2). Among these genes, *CsCOL4* and *CsCOL11* were identified in both rice and cotton Raymond, *CsCOL1*, *CsCOL7*, *CsCOL8*, *CsCOL5*, and *CsCOL9* were found in cotton Raymond alone, and the remaining were not present in any of the duplicated blocks (Fig. 5).

Spatial and temporal expression patterns analysis of 13 CsCOL genes

To gain have insights into the possible role of the *CsCOL* genes in the development of cannabis, the expression pattern of the *CsCOL* genes was analysed in four plant tissues: female flower, stem, leaf, and

root. The results revealed that all genes were constructively expressed in various tissues, but with different expression patterns (Fig. 6). Among the 13 *CsCOL* genes, 10 were found to be highly expressed in the leaf tissue, *CsCOL2* and *CsCOL3* were highly expressed in the female flower, and different expression patterns were found for *CsCOL13*, with its highest expression level in the stem, and lower expression levels in other tissues (Fig. 6).

Previous studies found that the *COL* gene played important role in the regulation of flowering time. To evaluate the possible functions of *CsCOL* genes, qRT-PCR was used to analysed the expression levels of *CsCOL* genes under different photoperiod treatments at 4-h intervals (Fig. 7). Under SD conditions, the diurnal expression pattern of the *CsCOL* genes varied. The expression patterns were roughly divided into three types (Fig. 7). The first categories exhibited high levels of expression at the end of darkness, including *CsCOL1-3, CsCOL5-7* and *CsCOL 10–12*. The second types showed an increased expression at 04:00 in the night (darkness), including *CsCOL4*, *CsCOL8* and *CsCOL13*. The remaining *CsCOL* genes displayed highest expression at the end of light (Fig. 7). Under LD conditions, although the transcript level of all *COL* genes was induced in light, two types of diurnal expression patterns were observed (Fig. 7). The first type included the transcript levels of most *CsCOL* genes, which peaked at 12:00 PM in the day (light) (*CsCOL1-5*, *CsCOL7*, *CsCOL8*, *CsCOL10*, *CsCOL11*, and *CsCOL13*), while the second category exhibited the highest expression levels at 16:00 in the day (light) (*CsCOL6*, *CsCOL9*, and *CsCOL12*). Collectively, these results suggest that the majority of the *CsCOL* genes exhibited a diurnal oscillation expression pattern under the SD and LD conditions.

To further explore the function of COL in cannabis, we investigated the expression patterns of 13 *CsCOL* genes in an early-flowering variety, “Qingma 1” (“Q1”), and a late-ripening variety, “Yunma 7” (“Y7”). The flowering time of“Q1”is 29 d after sowing, while that of “Y7” is 117 d under SD conditions in the field (Fig. S1). Under SD conditions, among the 13 *CsCOL* genes, three genes (*CsCOL4*, *CsCOL8*, and *CsCOL11*) showed higher expression levels in “Q1” than in “Y7” at the peak of the transcript levels of these genes, while six genes (*CsCOL5*, *CsCOL6*, *CsCOL7*, *CsCOL9*, *CsCOL10*, and *CsCOL12*) showed an opposing pattern. The remaining *COL* genes exhibited similar expression levels between “Y7” and “Q1” (Fig. 8).

Sequencing analysis of CsCOL genes

To further analyze the potential roles of the *CsCOL* genes in the regulation of flowering time, we investigated the amino acid sequence of the *CsCOL* genes in Y7 and Q1. Unfortunately, only *CsCOL3* and *CsCOL7* were successful cloned. For *CsCOL3*, two amino acid differences were found between Y7 and Q1, neither of which was located in B-Box nor CCT domain (Fig. 9A). With regards to *CsCOL7*, four amino acid differences were found between Y7 and Q1, all of which were located in the B-Box □ domain (Fig. 9B).

Discussion

The *COL* gene family plays a key role in the regulation of flowering time, and has been reported in many plants species, including Arabidopsis, rice, maize, *Populus*, Lilium × formolongi [5, 10, 12–14]. However, a genome-wide investigation of the *COL* family gene in cannabis has yet to be conducted. Due to the unavailability of a high-quality cannabis genome sequence, work on the genome-wide identification of *COL* genes in the cannabis genome has been lacking. The most recently assembled cannabis genome contained gene location information at the chromosome level [27], which allowed for a comprehensive analysis of the *COL* gene family in cannabis. In the present study, 13 *COL* members were identified in the cannabis genome. These were divided into three subgroups (Fig. 1–2), which is similar to the grouping in rice and Arabidopsis [16]. Similar numbers of *COL* genes were found in other plants, including *Populus* (14 *COL* genes), sorghum (15), and rice (16) [13, 16, 28]. The fact that the size of the genome of the four plants differed suggests that the number of *COL* genes in the COL superfamily was stable and did not vary with genome size.

Tandem replication events are associated with the occurrence of novel functions and gene expansion. In cannabis, the replication events have been found to occur in the *CBCAS*, *THCAS*, and *CBDAS* genes [27, 29]. On the other hand, no tandem duplication events have been observed in the *COL* genes of cotton [28]. However, in this study, a tandem gene pair (*CsCOL6* and *CsCOL12*) was found on chromosome 3 in cannabis, indicating that gene replications may be a important driving force of cannabis gene evolution. In addition to tandem replication events, segmental duplication has been reported as the main driving force of gene expansion in the *COL-like* gene family in cotton and maize [12, 28]. Consistent with these findings, in this study, a segmental duplications gene pair (*CsCOL8/CsCOL11*) was found in the cannabis genome (Fig. 4). However, among the genes involved in duplication, these two pair duplication genes displayed different expression patterns under the SD and LD conditions (Figs. 1, 7 and 8), which indicated that these genes experienced functional divergence during gene duplication. In addition, 15 pairs of orthologous genes between cannabis and cotton Raymond were identified, while only two pairs were found between cannabis and rice (Fig. 5). This observation implies that cannabis *COL* genes have a closer relationship with cotton Raymond than with rice, which may be consistent with the evolutionary relationship between monocotyledons and dicotyledons. Interestingly, *CsCOL4* and *CsCOL11* were found in both rice and cotton Raymond, indicating that these *COLs* genes expanded in a species-specific manner from common ancestral genes before the dicot–monocot divergence.

Although previous studies have shown that *COL* genes are widely expressed in different plant tissues, they have been found to be preferentially expressed in the leaves [5, 14, 17]. Leaves sense photoperiod signals and express *COL* to activate FT and promote flowering [9] In this study, we investigated the transcript levels of 13 *CsCOLs* in various plant organs, including the female flower, leaves, roots, and stems. As a result, 10 *COL* genes in cannabis were found to be preferentially expressed in the leaves, with an expression pattern similar to that observed in other plants, indicating their potential functions in leaves (Fig. 7).

The photoperiod is considered to be a key determining factor of flowering timing in plants, and *COL* genes have been demonstrated to be involved in the regulation of photoperiod-mediated flowering [5, 9,

17]. Therefore, we investigated the diurnal variations in the transcript levels of the *CsCOL* genes under the LD and SD conditions. Under the SD conditions, all the *COL* genes exhibited a diurnal oscillation expression pattern, with few differences between them. The transcript levels of 9 *COL* gene members were found to peak at the dawn (Fig. 7), similar to *OsCOL16*, *PtCOL1/2*, *AtCOL1*, *AtCOL2*, and *AtCO* in other plants [5, 13]. Under the LD conditions, the expression patterns of all *CsCOL* genes was roughly divided into two types. The first type included 10 *CsCOL* genes that were expressed more highly after light treatment, peaking at 12 h, consistent with *COL* genes including *PtCOL14* in *Populus* [13]. Similar to *LfCOL13-16*, *OsCOL10*, and *OsCOL16*, the remaining 3 *CsCOL* genes showed higher expression levels in light than in darkness, and peaked at 16 h [5, 14, 17].

Different expression levels of *COL* genes may be associated with the flowering time in different varieties. In this study, the “Q1” variety exhibited an earlier flowering time than “Y7” under the SD conditions in the field (Fig. S1). To further determine the potential functions of *CsCOL* genes in the regulation of the flowering time, we evaluated the transcript levels of all *CsCOL* genes in the “Q1” and “Y7” varieties under the SD conditions. As shown in Fig. 8, the expression levels of *CsCOL4*, *CsCOL8*, and *CsCOL11* were higher in “Q1” than “Y7” at the peak transcript level, while six genes (*CsCOL5*, *CsCOL6*, *CsCOL7*, *CsCOL9*, *CsCOL10* and *CsCOL12*) showed a contrasting pattern (Fig. 8). Interestingly, except for the similar expression patterns of *CsCOL6*, *CsCOL12*, and *OsCOL16* under the SD and LD conditions (Fig. 7), these three genes also belonged to the same subgroup based on their phylogenetic relationship analysis (Fig. 2). In a previous study, *CsCOL6* was found to repress flowering in rice [5]. Thus, CsCOL6/CsCOL12 may exert a similar function to OsCOL16, thus delaying flowering in cannabis. However, this requires further study.

Previous studies have suggested that differences in the amino acid sequences of *COL* genes could explain their varied functions in the photoperiod-mediated flowering pathway [5, 10, 20]. In present study, differences in the amino acid sequences of *CsCOL3* and *CsCOL7* were observed between “Q1” and “Y7”, an early-flowering variety and a late-ripening variety, respectively (Fig. 9 and Fig. S1). Moreover, differences were observed in the amino acid sequence of *CsCOL7* within in the B-box, a conserved domain known for its functions in protein–protein interactions. However, whether these changes affect this type of function will need to be studied further.

Conclusions

To summarize, this study is the first to provide a comprehensive analysis of the *COL* gene family in cannabis. Our aim was to elucidate the evolution, expression profiles, and potential functions of these genes in the regulation of flowering in cannabis. Although the possible functions of the *CsCOL* gene family require further study for validation, the systemic analysis conducted in this study provides a foundation for future studies on the biological and molecular functions of *COLs* in cannabis.

Methods

Evaluation of flowering time and photoperiod treatment

“Yunma 7” (“Y7”) and “Qingma1” (“Q1”) were collected from the Institute of Bast Fiber Crops, China Academy of Agriculture Science, Changsha, China. The two varieties were randomly planted under natural short-day conditions in Changsha (southern China, 112°58' E/28°11' N, day length < 12 h during vegetative period). Once over 50% of the plants of each cultivar had bloomed, the flowering time was scored. For different photoperiod treatments under the LD (16 h light/8 h dark) and SD (8 h light/16 h dark) conditions, the leaves of these seedlings were collected at 0:00, 04:00, 08:00, 12:00, 16:00 and 20:00 after photoperiod treatment. The resulting materials were promptly transferred into liquid nitrogen for RNA extraction, repeated independently in triplicate.

Identification and analysis of physical and chemical properties of CsCOL gene family members in cannabis

The sequences of 17 Arabidopsis CONSTANS-like proteins were downloaded from the Arabidopsis Information Resource (TAIR) (http://www.aabidopsis.org/). The cannabis genome file and genome annotation file (assembly number: GCA_900626175.2) were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/) [27]. The software TBtools was used to compare the Arabidopsis *COL* gene with the cannabis genome by blast sequence alignment (E-value < $1E^{-5}$) and to screen the *CsCOL* family candidate genes in the cannabis genome. Then, the candidate genes were submitted to the Uniprot database (https://www.uniprot.org/) for batch comparison to verify whether the candidate genes contained both CCT and B-box conserved domains. ProtParam (http://web.expasy.org/protparam/) was used to analyze various physicochemical parameters of the *CsCOL* genes.

Multisequence alignment, phylogenetic analysis of CsCOL proteins, and gene structure analysis of CsCOL

Multisequence alignment analysis of Arabidopsis, rice, and cannabis COL proteins was performed using Clustal X2.1 with the default parameters [30]. A phylogenetic tree was constructed using MEGA7.0 using the neighbour-joining (NJ) method. Bootstrap values (> 50%) were estimated using 1000 replicates. FigTree software was use to edit the phylogenetic tree. The protein structure of CsCOL was predicted using NCBI-CDD software online (https://www.ncbi.nlm.nih.gov/cdd/) with the default parameters (E-value < 0.01). The conserved motif (Motif) of the *CsCOL* genes was analysed using MEME software online (http://meme-suite.org), and the predicted number was set to 10. The coding sequence (CDS) and untranslated region (UTR) of the *CsCOL* were extracted from the cannabis genome annotation file using TBtools, which was also used to combine evolutionary tree, protein domain, gene conservative motif, CDS, and UTR to construct a diagram to compare the evolutionary relationships and structures of *CsCOL*.

Chromosome distribution and synteny analysis of CsCOL

Information on the chromosome location of the *CsCOL* genes was extracted from cannabis genome file and gene annotation file using TBtools. Then, the physical location of *CsCOL* genes on chromosome was

constructed using TBtools. TBtools, MCscanX, and Circos were used to calculate and draw the tandem repeats of *COL* on the chromosome, the collinear genes between different chromosomes of cannabis, and the collinear genes among different species of hemp, *Gossypium raimondii*, and rice.

Gene cloning

The primers pairs of the *CsCOL* genes were designed according to the CDS sequences (Supplementary Table S1). cDNA of “Y7” and “Q1” was used as template for each gene. PCR was performed as follows: an initial step at 94 °C for 5 min, followed by 30 cycles of 30 s at 98 °C, 30 s at 55 °C, 2 min at 68 °C, and a final extension of 10 min at 68 °C. After the PCR procedure was finished, the PCR product was purified, ligated to pGEM-T Easy vector, and transformed into *E. coli* DH5a. Positive clones were selected for sequencing. A list of the primers used for gene cloning is provided in Supplementary Table S1.

RNA extraction and qRT-PCR analysis

Total RNA was extracted from different tissues and leaves under different photoperiods using a RNAprep Pure Plant Kit (Tiangen, Beijing). The cDNA was synthesizing using a PrimeScript 1st Strand cDNA Synthesis Kit (TaKaRa, Japan). According to the manufacturer’s instructions, quantitative RT-PCR (qRT-PCR) was conducted using a SYBR Premix Ex TaqTM kit (TaKaRa) on a 7500 Sequence Detection System (Applied Biosystems, USA). The DHS2 gene was amplified as the internal control. The primers used for qRT-PCR analysis are listed in Supplementary Table S2. The experiment was performed in triplicate.

Abbreviations

COL
CONSTANT-like; CBD:cannabinoid; LD:long day; SD:short day; FT:FLOWERING LOCUS T; WGD:whole genome duplication; CBCAS:CBCA synthases; THCAS:THCA synthases; CBDAS:CBDA synthases; qRT-PCR:quantitative real-time PCR

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets for supporting the conclusions of this article are listed in the article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by China Agriculture Technology Research System (CARS-16-E-02), Central Public-interest Scientific Institution Basal Research Fund (1610242020003; 1610242019001). Funds were used for performing the experiment, and analysis data and in writing the manuscript.

Authors' contributions

G. P., and Z.L performed the experiments; G.P wrote the paper; M.Y., and J.T., participated in the sample collection and RNA extraction; A.-G.C., J.-J.L., H.-J.T., L.C and Y.D conducted bioinformatic analysis. S.-Q.H. revised the manuscript; L.-N.Z., and D.-F.L designed the experiment, and L.-N.Z also provided the materials. All authors read and approved the version to be published. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We thank all the staff in our lab for providing useful suggestions and technical assistance. We are very grateful to the editor and reviewers for providing constructive comments for the improvement of our manuscript.

References

1 Andre, C.M., Hausman, J.F., Guerriero, G., 2016. Cannabis sativa: The Plant of the Thousand and One Molecules. Front Plant Sci 7:19.

2 Corroon, J., and Phillips, J. A. (2018). A cross-sectional study of cannabidiol users. Cannabis and Cannabinoid Res. 2018; 3(1): 152–161

3 Lisson S N, Mendham N J, Carberry P S. Development of a hemp (*Cannabis sativa* L.) simulation model 2. The flowering response of two hemp cultivars to photoperiod[J]. Australian Journal of Experimental Agriculture, 2000, 40(3):413-417.

4 Song Y H, Ito S, Imaizumi T. Flowering time regulation: photoperiod- and temperature-sensing in leaves[J]. Trends in Plant Science, 2013, 18(10).

5 Wu, X. Zheng, D. Chen, Y. Zhang, W. Ma, H. Zhang, et al., *OsCOL16*, encoding a CONSTANS-like protein, represses flowering by up-regulating *Ghd7* expression in rice. Plant Sci. 260, 60-69

6 Liu J, Cheng Z, Li X, Xie L, Bai Y, Peng L, Li J, Gao J: Expression Analysis and Regulation Network Identification of the CONSTANS-Like Gene Family in Moso Bamboo (*Phyllostachys edulis*) Under Photoperiod Treatments. DNA CELL BIOL 2019, 38(7):607-626.

7 Imaizumi T , Schultz T , Harmon F , et al. FKF1F-BOX protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis[J]. Science, 2005, 309(5732):p. 293-297.

8 Kobayashi, D. Weigel, Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering, Genes Dev. 21 (2007) 2371–2384.

9 Turck F, Fornara F, Coupland G: Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. ANNU REV PLANT BIOL 2008, 59:573-594.

10 Robson F , Costa M M , Hepworth S R , et al. Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants.[J]. Plant Journal, 2001, 28(6):619–631.

11 Khanna R, Kronmiller B, Maszle DR, Coupland G, Holm M, Mizuno T, et al. (2009) The Arabidopsis Bbox zinc finger family. Plant Cell 21: 3416–3420.

12 Song N, Xu Z, Wang J, Qin Q, Jiang H, Si W, Li X: Genome-wide analysis of maize *CONSTANS-LIKE* gene family and expression profiling under light/dark and abscisic acid treatment. GENE 2018, 673:1-11.

13 Li J, Gao K, Yang X, Khan WU, Guo B, Guo T, An X: Identification and characterization of the *CONSTANS-like* gene family and its expression profiling under light treatment in *Populus*. INT J BIOL MACROMOL 2020, 161:999-1010.

14 Li Y, Zhao Y, Zhang M, Jia G, Zaccai M: Functional and Evolutionary Characterization of the *CONSTANS-like* Family in Liliumï¿½formolongi. PLANT CELL PHYSIOL 2018, 59(9):1874-1888.

15 Hu T, Wei Q, Wang W, Hu H, Mao W, Zhu Q, Bao C: Genome-wide identification and characterization of CONSTANS-like gene family in radish (*Raphanus sativus*). PLOS ONE 2018, 13(9):e204137.

16 Griffiths, S., Dunford, R.P., Coupland, G., Laurie, D.A., 2003. The evolution of *CONSTANS-like* gene families in barley, rice, and Arabidopsis. Plant Physiol. 131, 1855–1867.

17 Tan, J., Jin, M., Wang, J., Wu, F., Sheng, P., Cheng, Z., Wang, J., Zheng, X., Chen, L., Wang, M., Zhu, S., Guo, X., Zhang, X., Liu, X., Wang, C., Wang, H., Wu, C., & Wan, J. (2016). OsCOL10, a *CONSTANS-Like* Gene, Functions as a Flowering Time Repressor Downstream of *Ghd7* in Rice. Plant & cell physiology, 57(4), 798–812.

18 Takase T, Kakikubo Y, Nakasone A, Nishiyama Y, Yasuhara M, Tokioka-Ono Y, Kiyosue T (2011) Characterization and transgenic study of *CONSTANS-LIKE8(COL8)* gene in Arabidopsis thaliana: expression of 35S: COL8 delays flowering under long-day conditions. Plant Biotechnol.: 1110140050–1110140050.

19 Wang, H., Zhang, Z., Li, H., Zhao, X., Liu, X., Ortiz, M., Lin, C., & Liu, B. (2013). *CONSTANS-LIKE 7* regulates branching and shade avoidance response in Arabidopsis. Journal of experimental botany,

64(4), 1017–1024.

20 Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y et al: Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. PLANT CELL 2000, 12(12):2473-2484.

21 Sheng P, Wu F, Tan J, Zhang H, Ma W, Chen L, Wang J, Wang J, Zhu S, Guo X et al: A CONSTANS-like transcriptional activator, *OsCOL13*, functions as a negative regulator of flowering downstream of OsphyB and upstream of Ehd1 in rice. PLANT MOL BIOL 2016, 92(1-2):209-222.

22 Datta S, Hettiarachchi GH, Deng XW, Holm M: Arabidopsis *CONSTANS-LIKE3* is a positive regulator of red light signaling and root growth. PLANT CELL 2006, 18(1):70-84.

23 Hassidim M , Harir Y , Yakir E , et al. Over-expression of *CONSTANS-LIKE 5* can induce flowering in short-day grown Arabidopsis[J]. Planta, 2009, 230(3):481-491.

24 Chaurasia, A.K., Patil, H.B., Azeez, A., Subramaniam, V.R., Krishna, B., Sane, A.P., et al. (2016). Molecular characterization of *CONSTANS-Like* (COL) genes in banana (*Musa acuminata* L. AAA Group, cv. Grand Nain). Physiol Mol Biol Plants 22, 1–15.

25 Holefors A, Opseth L, Ree RA, Ripel L, Snipen L, Fossdal CG, Olsen JE: Identification of *PaCOL1* and *PaCOL2*, two *CONSTANS-like* genes showing decreased transcript levels preceding short day induced growth cessation in Norway spruce. Plant Physiol Biochem 2009, 47(2):105-115.

26 Ding, J., Bo¨hlenius, H., Ru¨hl, M.G., Chen, P., Sane, S., Zambrano, JA., et al. (2018). *GIGANTEA-like* genes control seasonal growth cessation in *Populus*. N Phytol 218, 1491–1503.

27 Laverty KU, Stout JM, Sullivan MJ, Shah H, Gill N, Holbrook L, Deikus G, Sebra R, Hughes TR, Page JE et al: A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. GENOME RES 2019, 29(1):146-156.

28 Qin W, Yu Y, Jin Y, Wang X, Liu J, Xi J, Li Z, Li H, Zhao G, Hu W et al: Genome-Wide Analysis Elucidates the Role of *CONSTANS-like* Genes in Stress Responses of Cotton. INT J MOL SCI 2018, 19(9):2658.

29 Weiblen GD, Wenger JP, Craft KJ, ElSohly MA, Mehmedic Z, Treiber EL, Marks MD: Gene duplication and divergence affecting drug content in Cannabis sativa. NEW PHYTOL 2015, 208(4):1241-1250.

30 Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al: Clustal W and Clustal X version 2.0. BIOINFORMATICS 2007, 23(21):2947-2948.

Figures

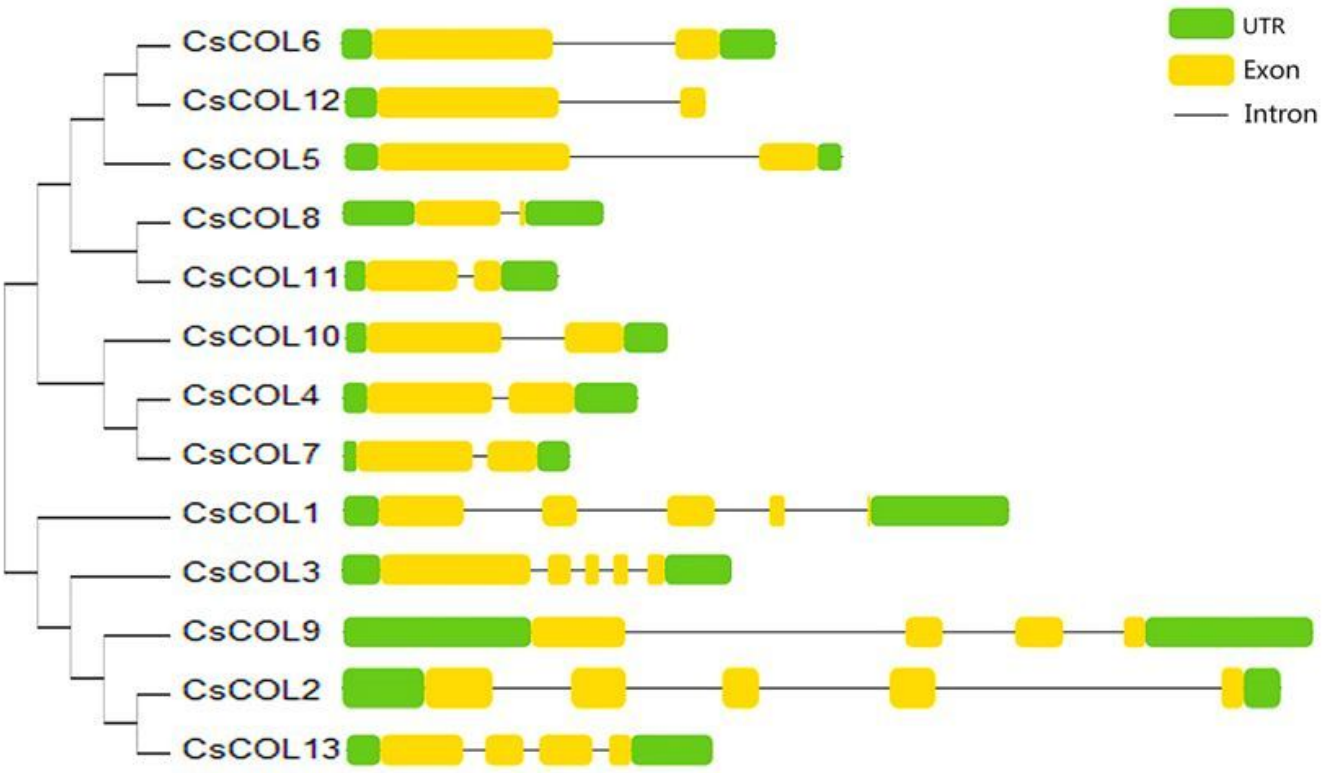

Figure 1

Phylogenetic and gene structure analyses of cannabis CsCOLs. A: Phylogenetic analysis. B: Gene structures. The exon, untranslated region (UTR), and intron are represented by the yellow and blue rectangles, and a blank line, respectively.

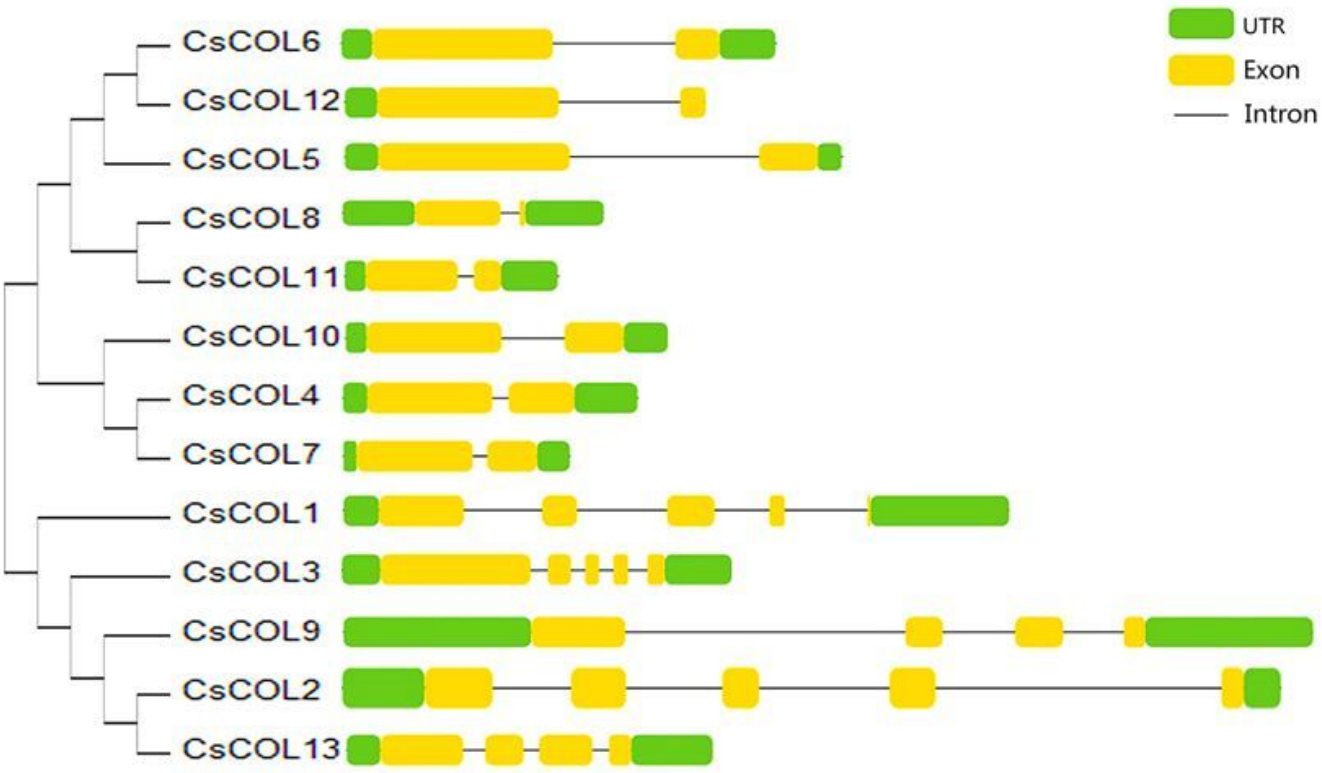

Figure 1

Phylogenetic and gene structure analyses of cannabis CsCOLs. A: Phylogenetic analysis. B: Gene structures. The exon, untranslated region (UTR), and intron are represented by the yellow and blue rectangles, and a blank line, respectively.

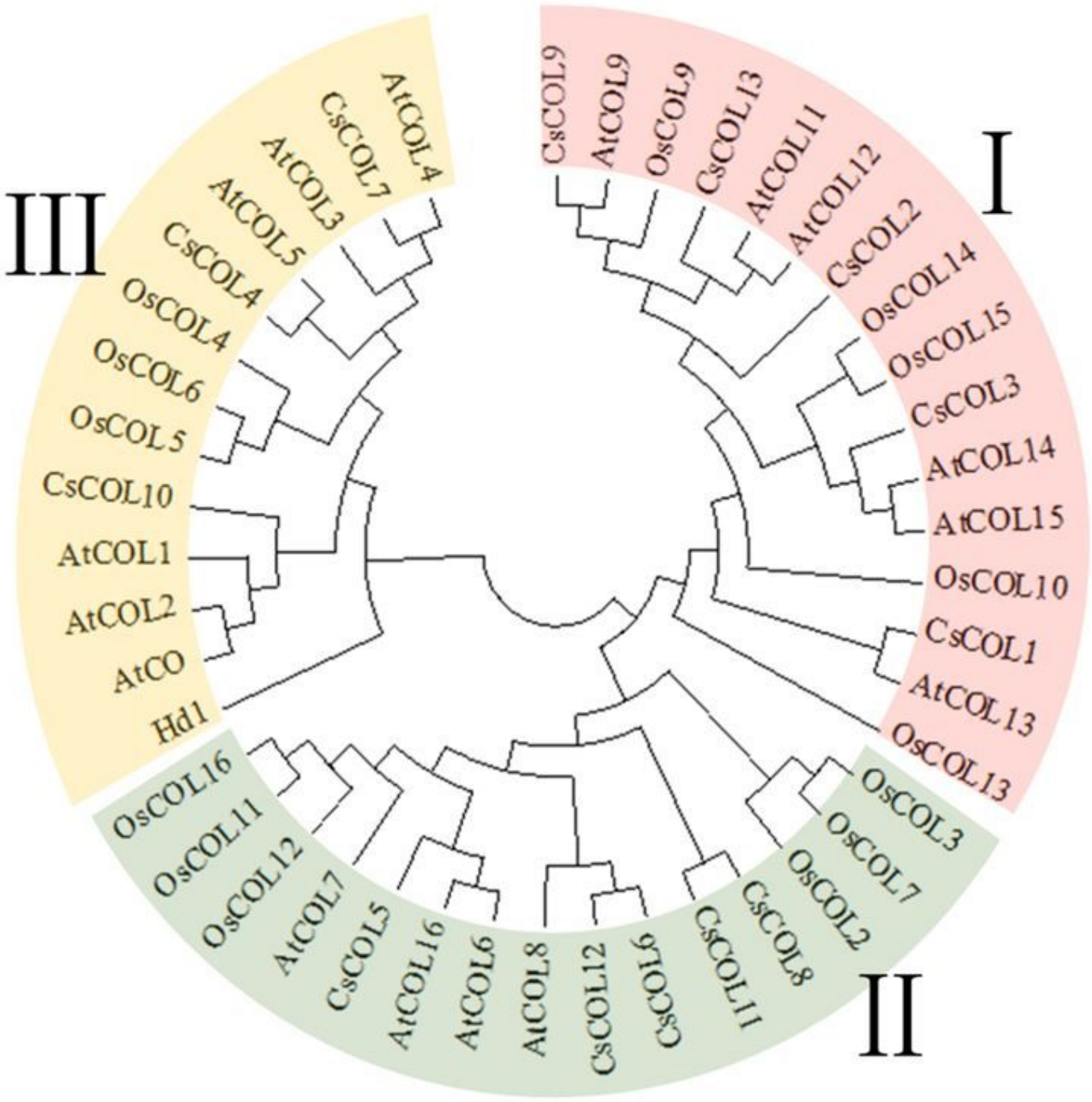

Figure 2

Phylogenetic tree of the COL proteins from three plant species. The phylogenetic tree was constructed based on the 90% shared amino acid sites using the maximum likelihood method. At, Arabidopsis thaliana; Os, Oryza sativa; Cs, Cannabis sativa.

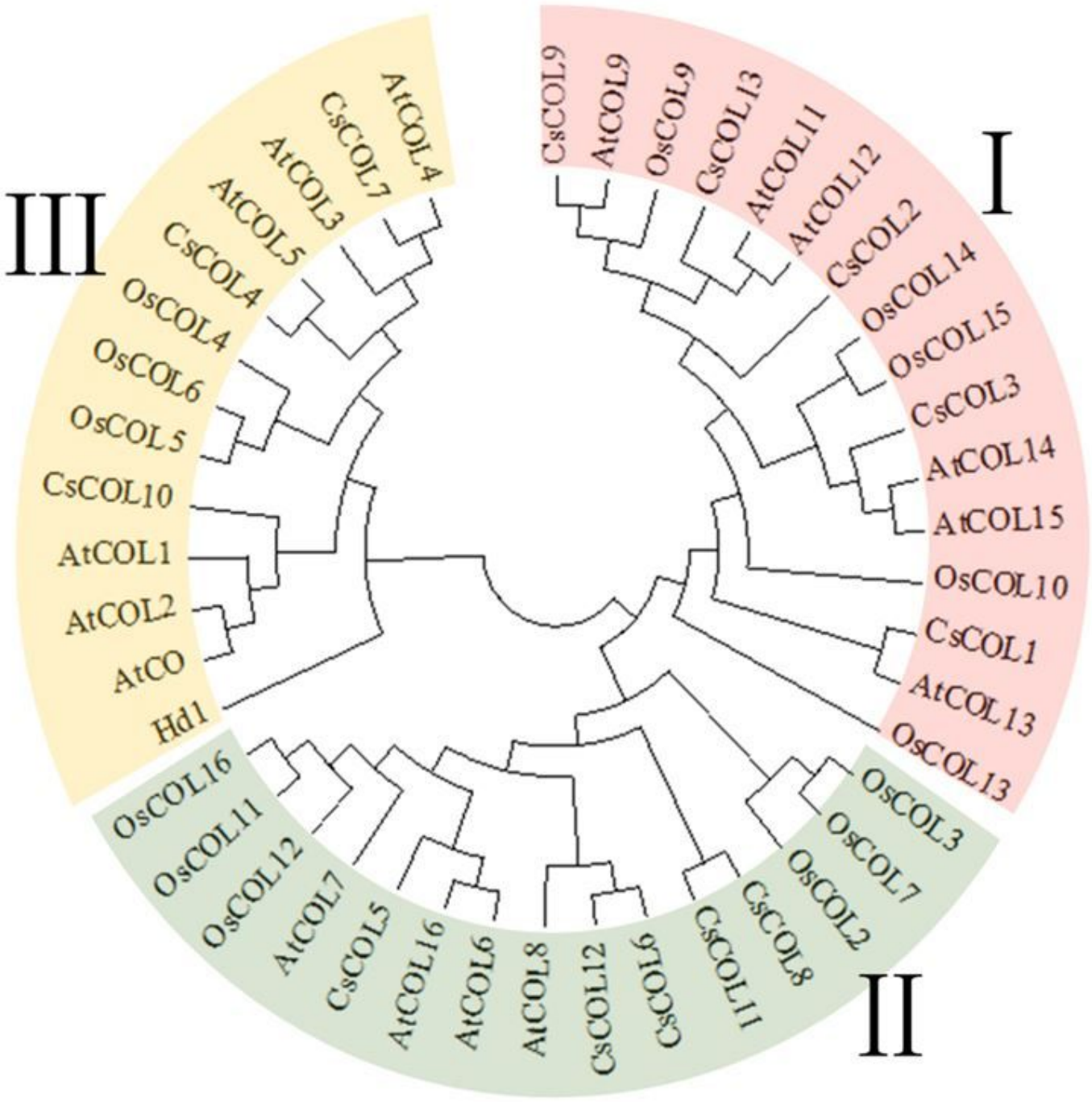

Figure 2

Phylogenetic tree of the COL proteins from three plant species. The phylogenetic tree was constructed based on the 90% shared amino acid sites using the maximum likelihood method. At, Arabidopsis thaliana; Os, Oryza sativa; Cs, Cannabis sativa.

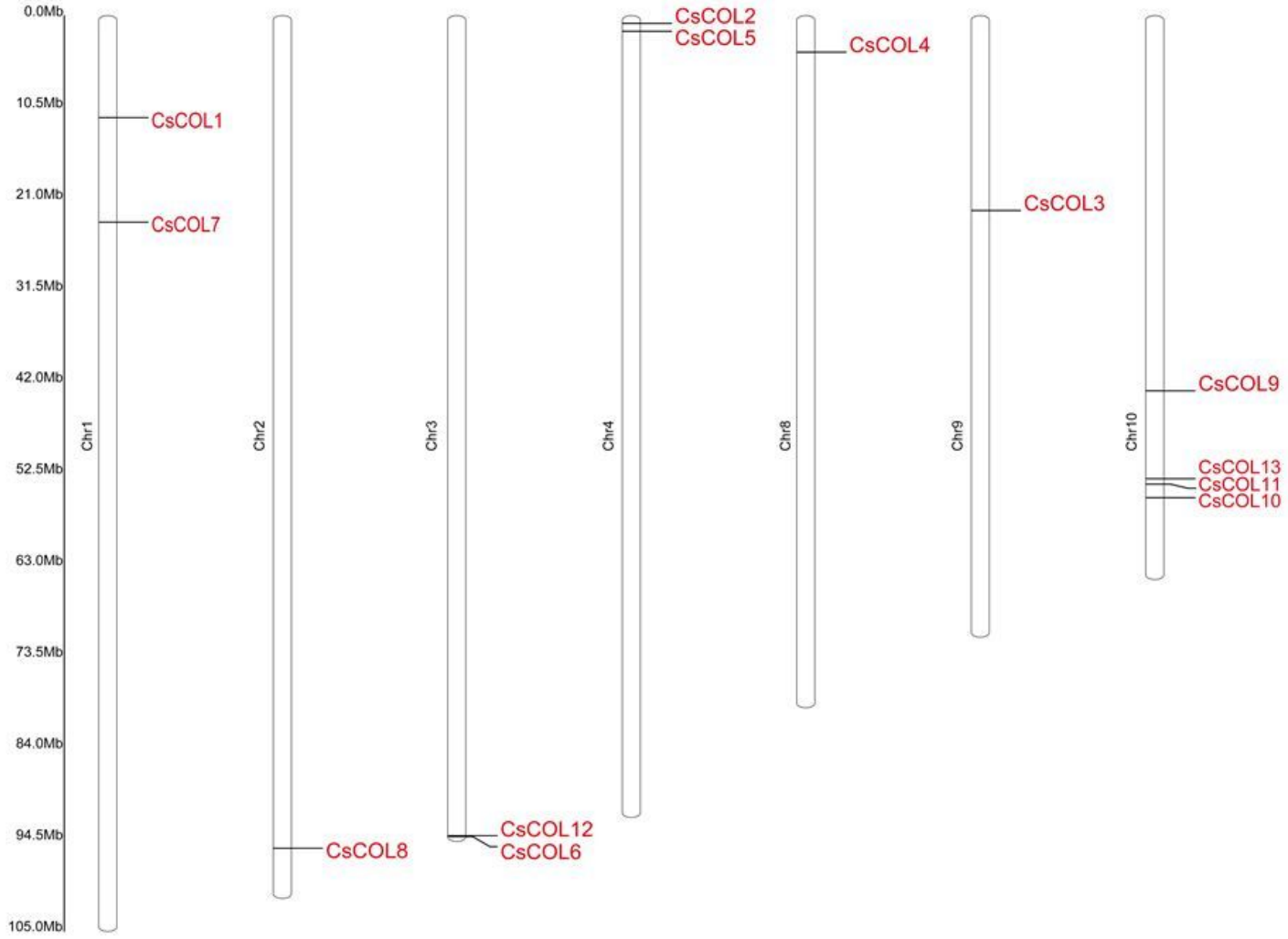

Figure 3

The physical location of 13 CsCOLs on the cannabis chromosome.

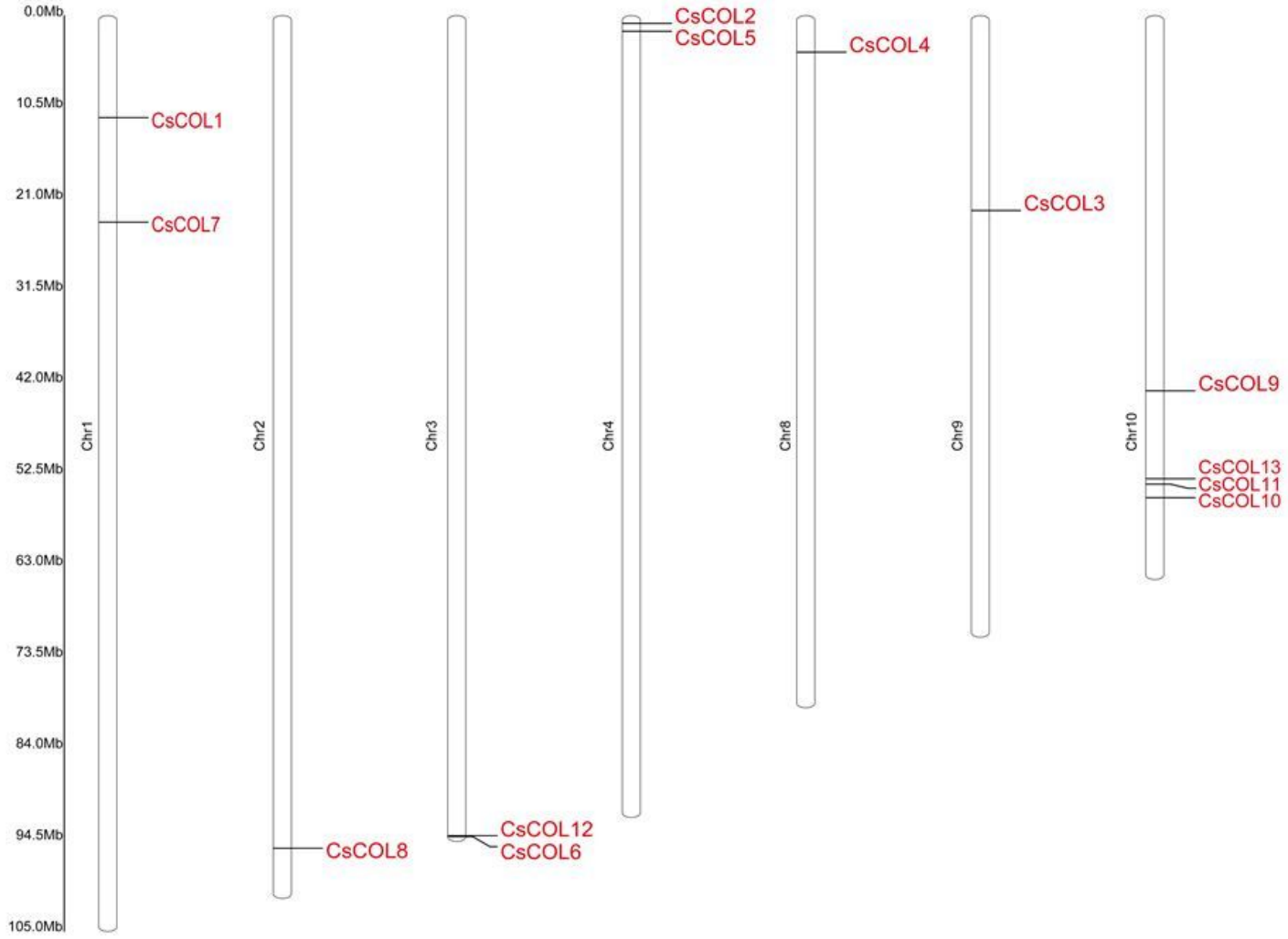

Figure 3

The physical location of 13 CsCOLs on the cannabis chromosome.

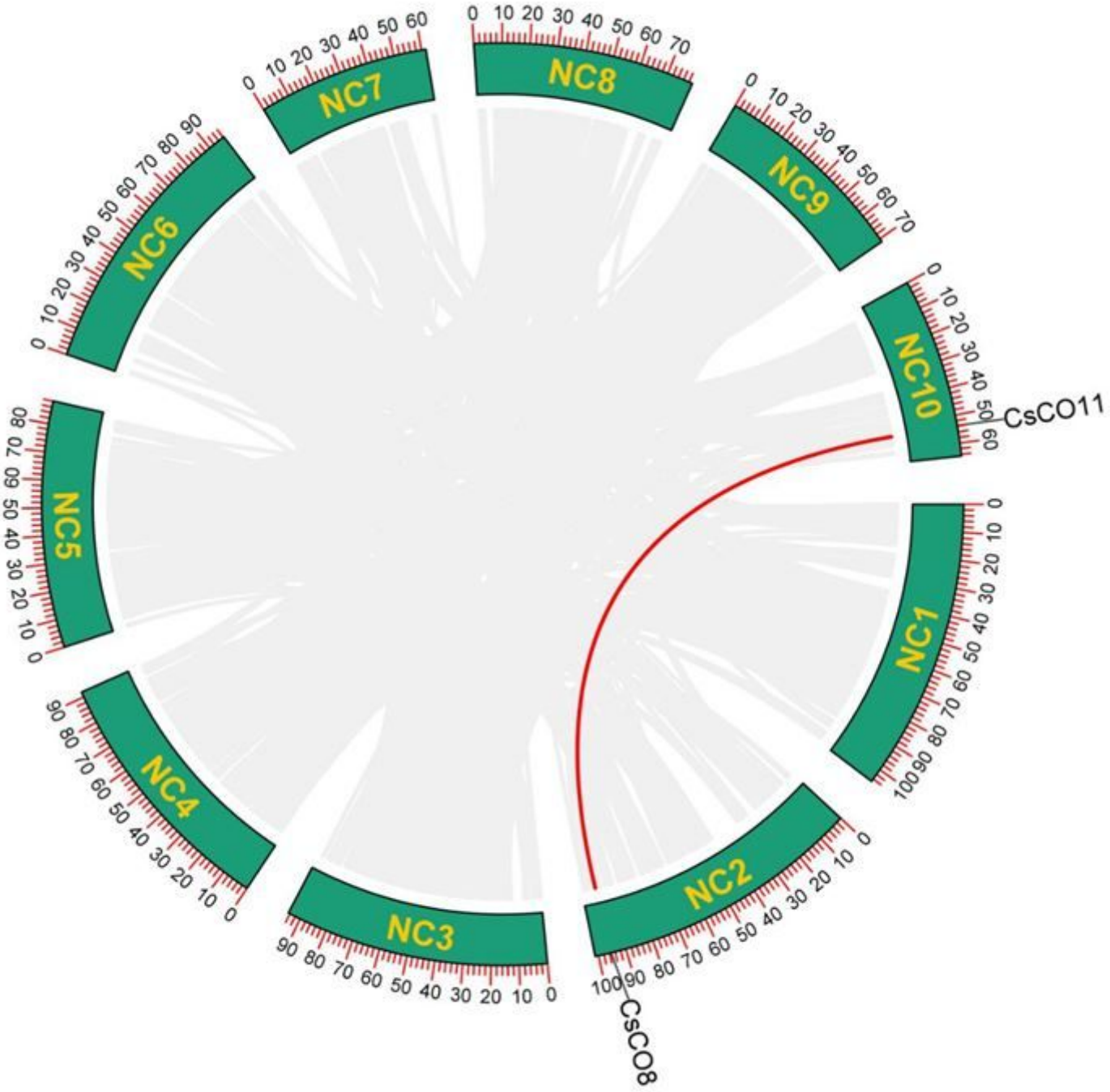

Figure 4

Schematic representation of the interchromosomal relationships between the CsCOL genes in the cannabis genome. Coloured lines indicate all syntenic blocks in the cannabis genome.

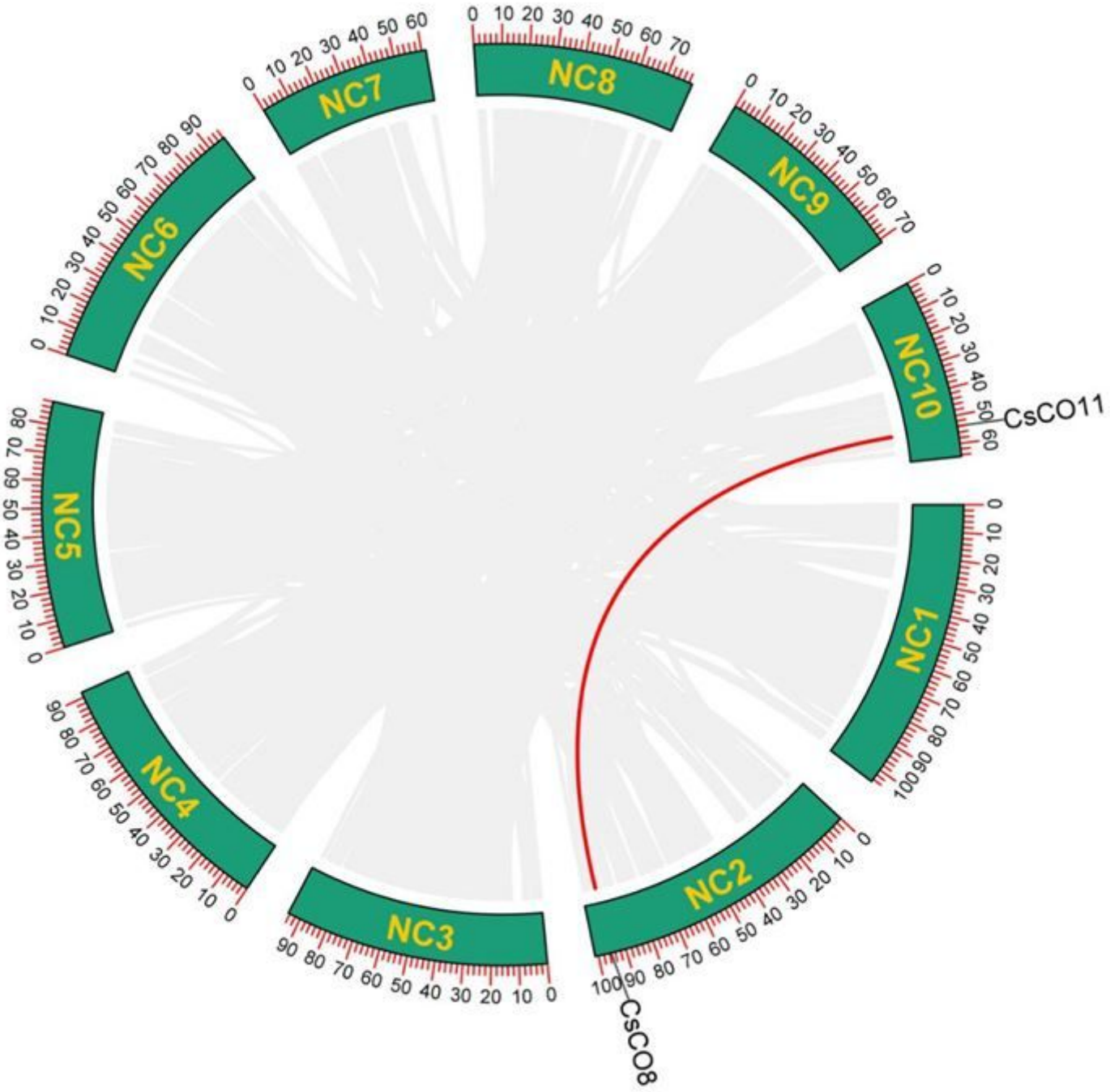

Figure 4

Schematic representation of the interchromosomal relationships between the CsCOL genes in the cannabis genome. Coloured lines indicate all syntenic blocks in the cannabis genome.

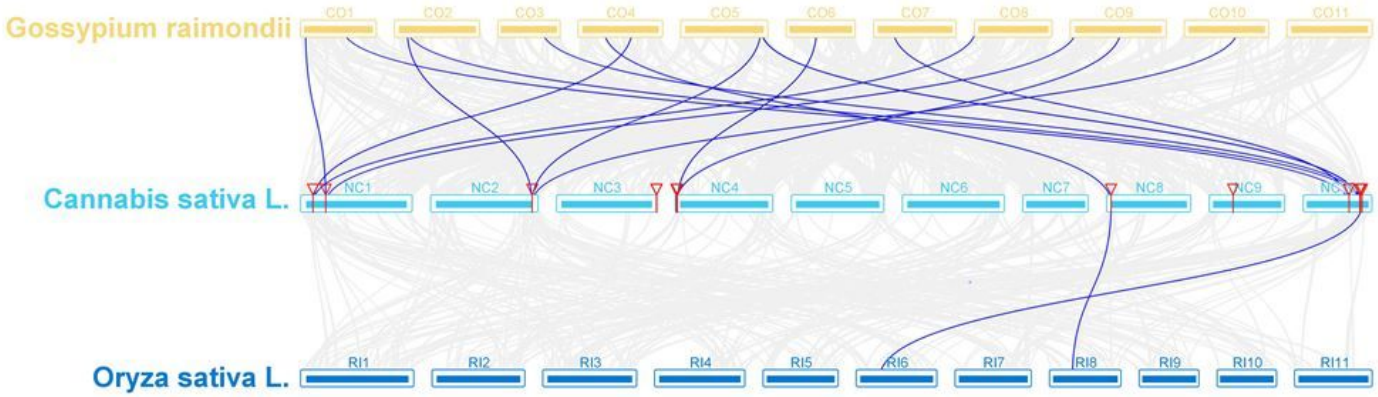

Figure 5

Synteny analysis of the COL genes between cannabis and Gossypium raimondii.

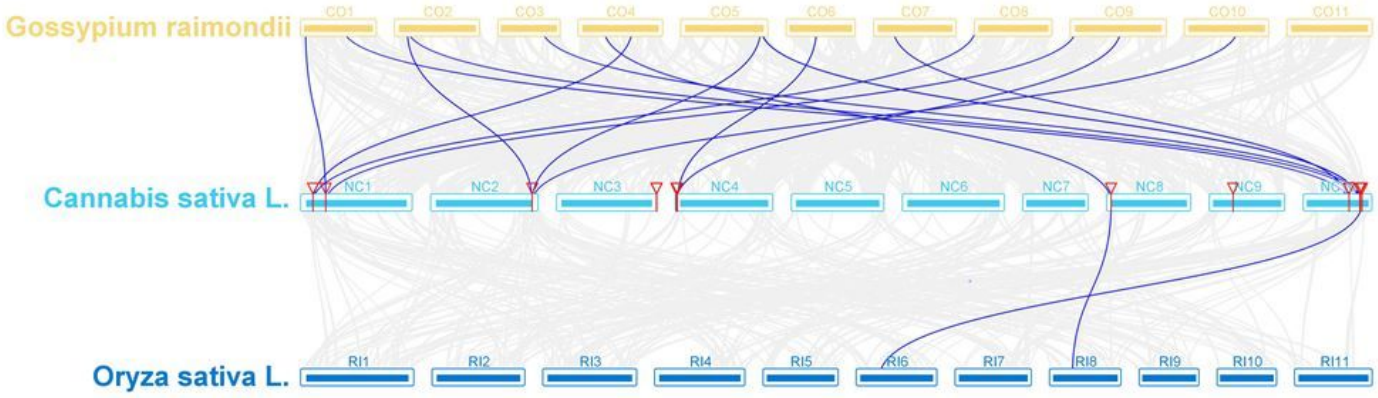

Figure 5

Synteny analysis of the COL genes between cannabis and Gossypium raimondii.

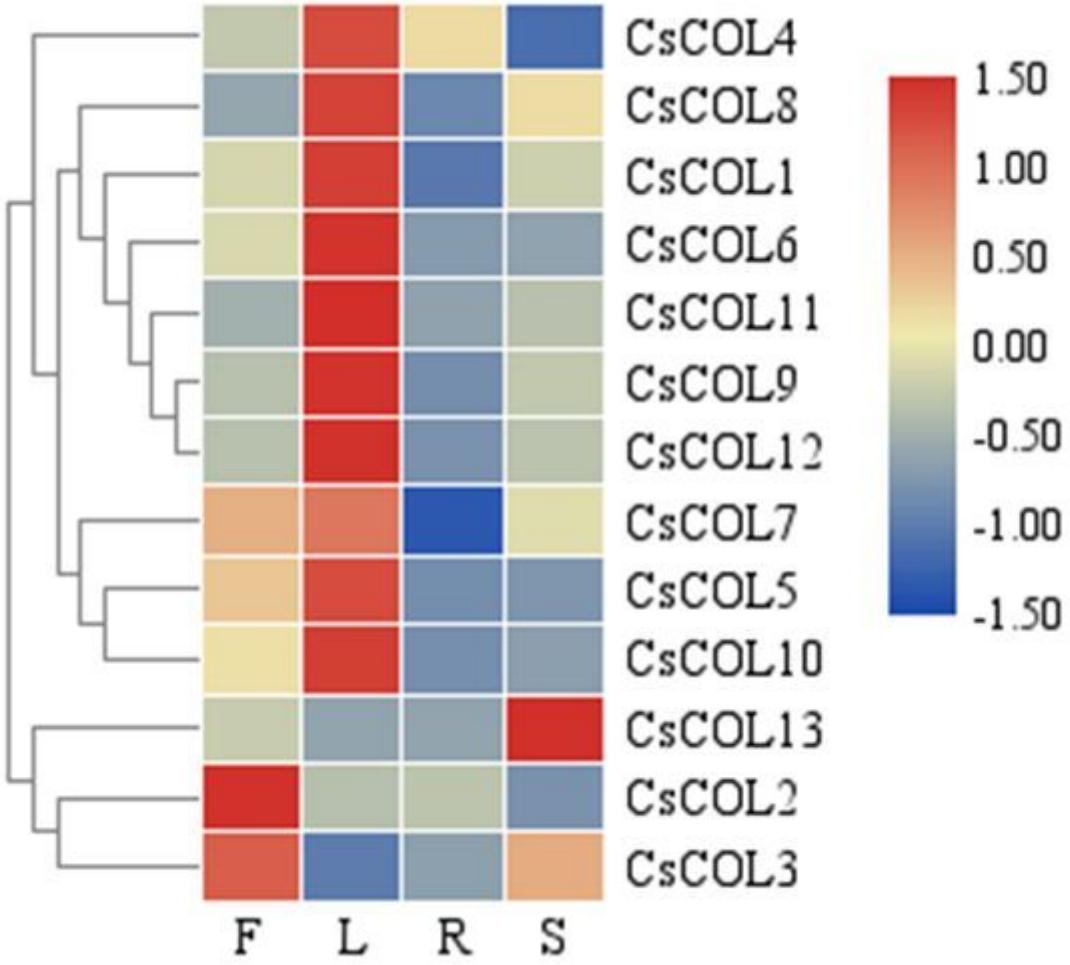

Figure 6

Tissue-specific gene expression of 13 CsCOLs in cannabis. F: female flower; L: leaf; R: root; S: stem.

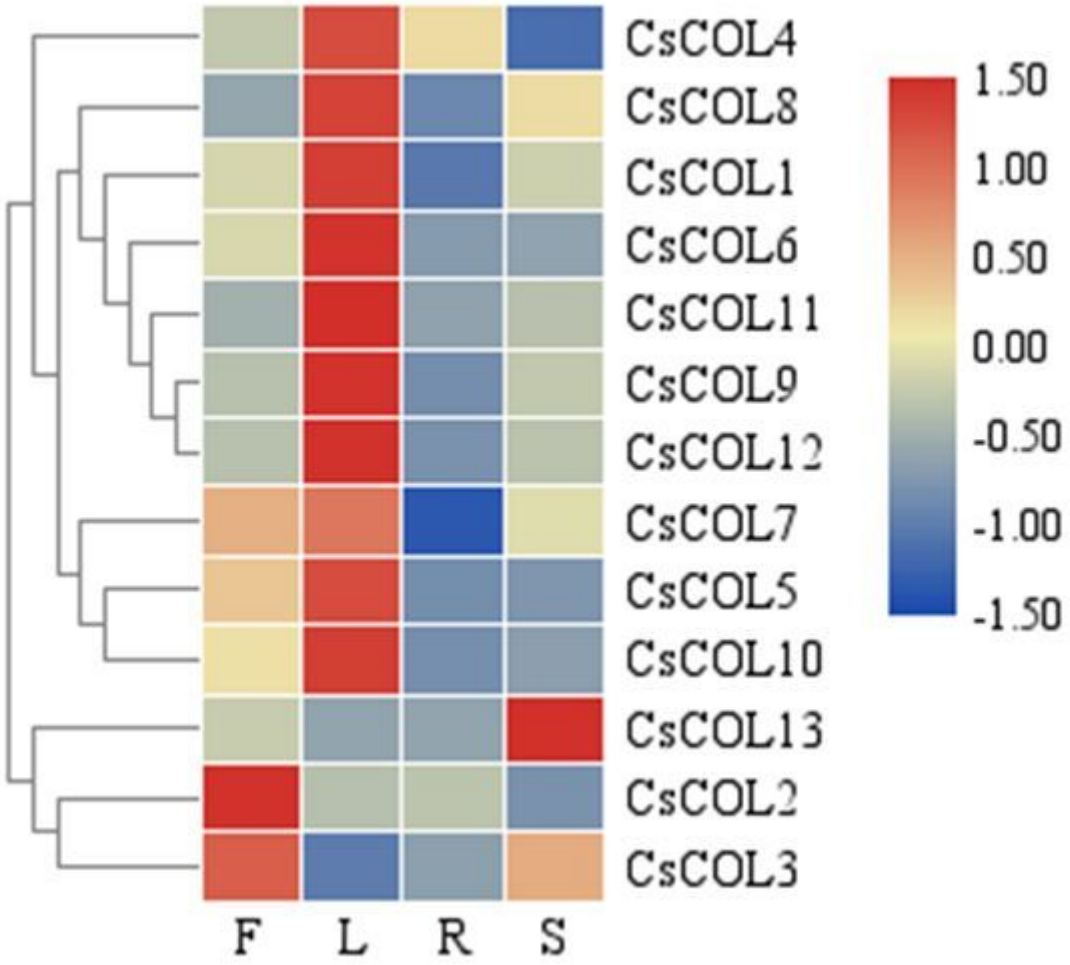

Figure 6

Tissue-specific gene expression of 13 CsCOLs in cannabis. F: female flower; L: leaf; R: root; S: stem.

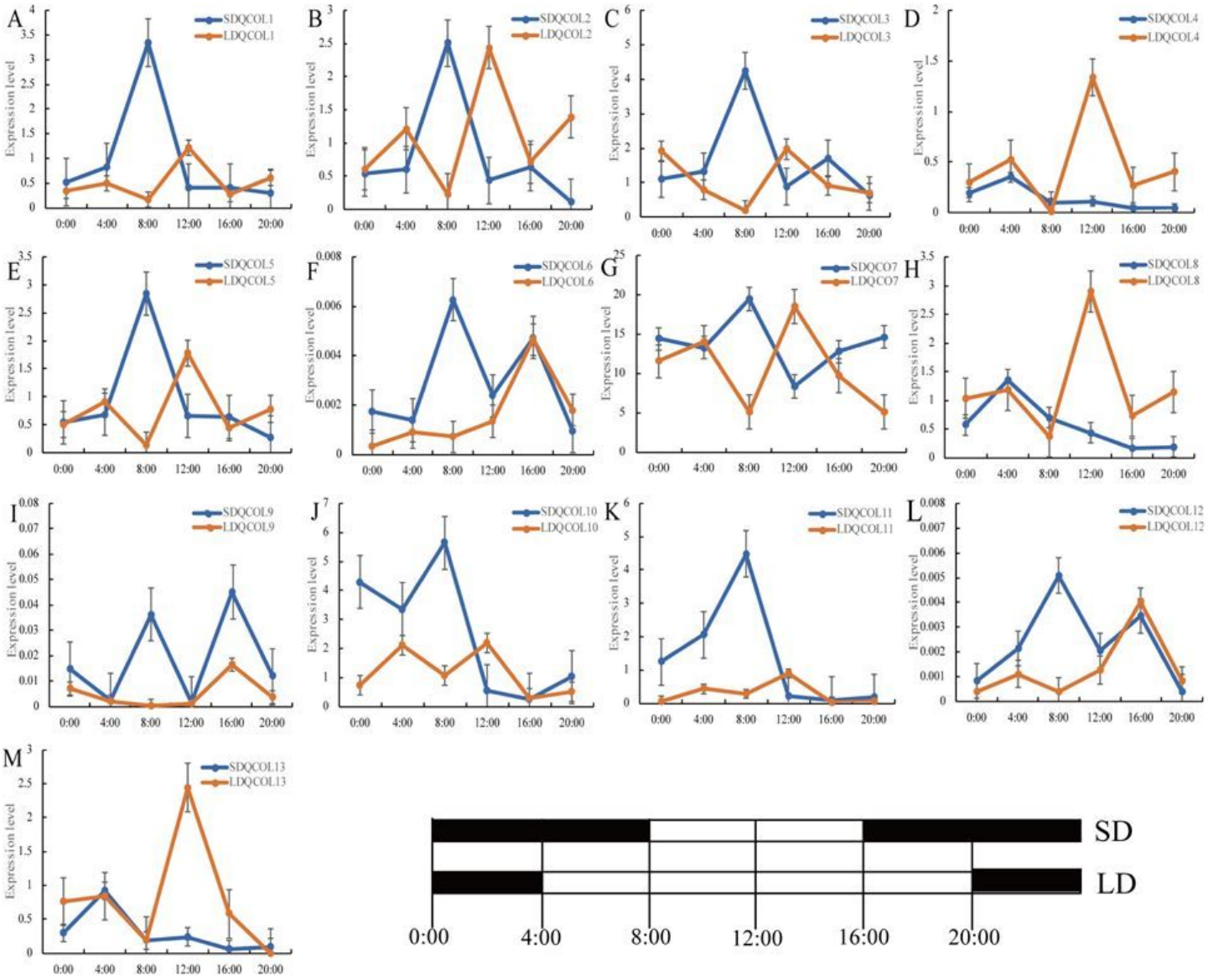

Figure 7

Expression patterns of the CsCOL gene under LD (8 h dark/16 h light) and SD (16 h dark/8 h light) conditions. SDQCOL1-13 represented the expression levels of CsCOLs of “Qingma 1” under the SD condition; LDQCOL1-13 represented the expression levels of CsCOLs of “Qingma 1” under the LD condition.

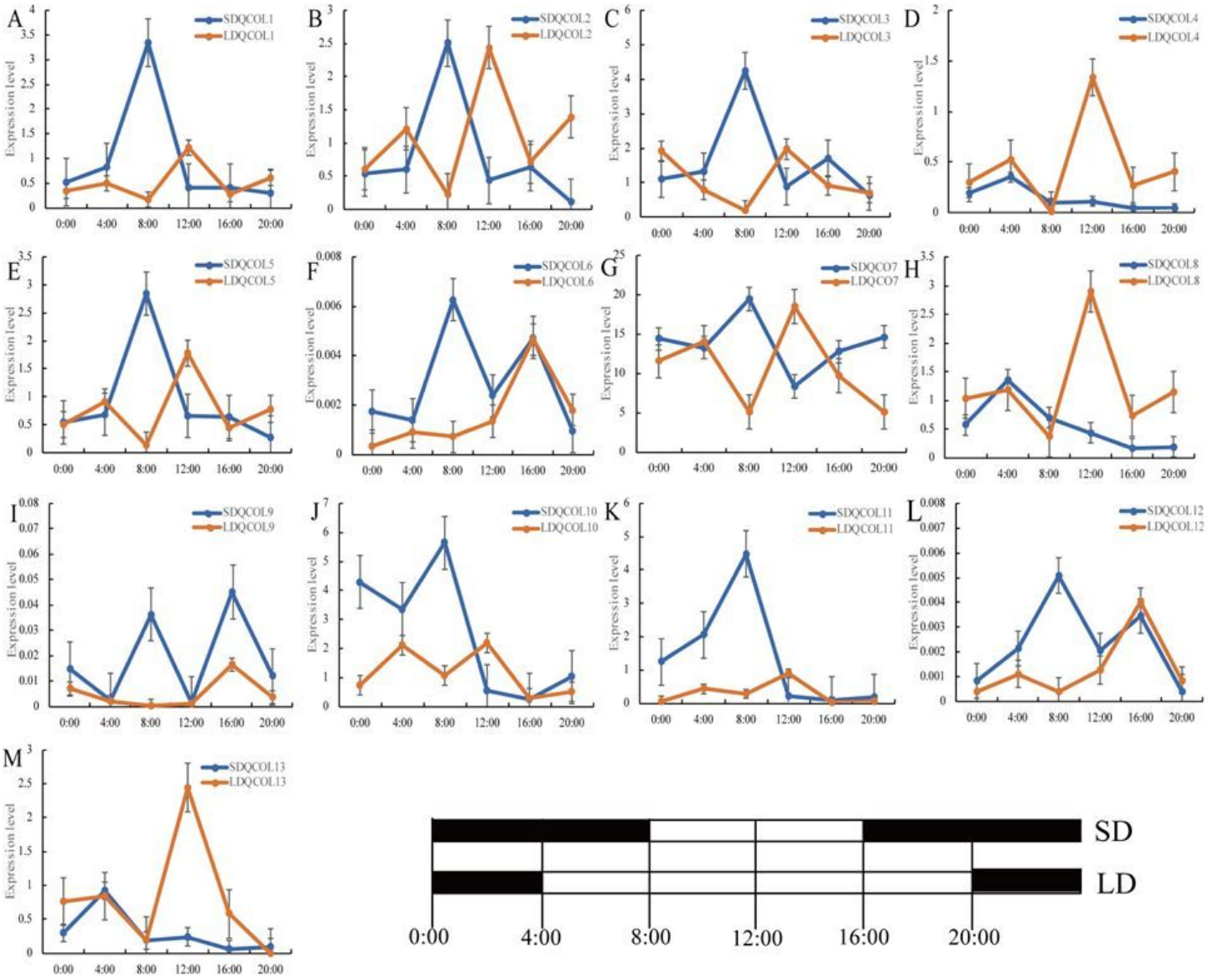

Figure 7

Expression patterns of the CsCOL gene under LD (8 h dark/16 h light) and SD (16 h dark/8 h light) conditions. SDQCOL1-13 represented the expression levels of CsCOLs of “Qingma 1” under the SD condition; LDQCOL1-13 represented the expression levels of CsCOLs of “Qingma 1” under the LD condition.

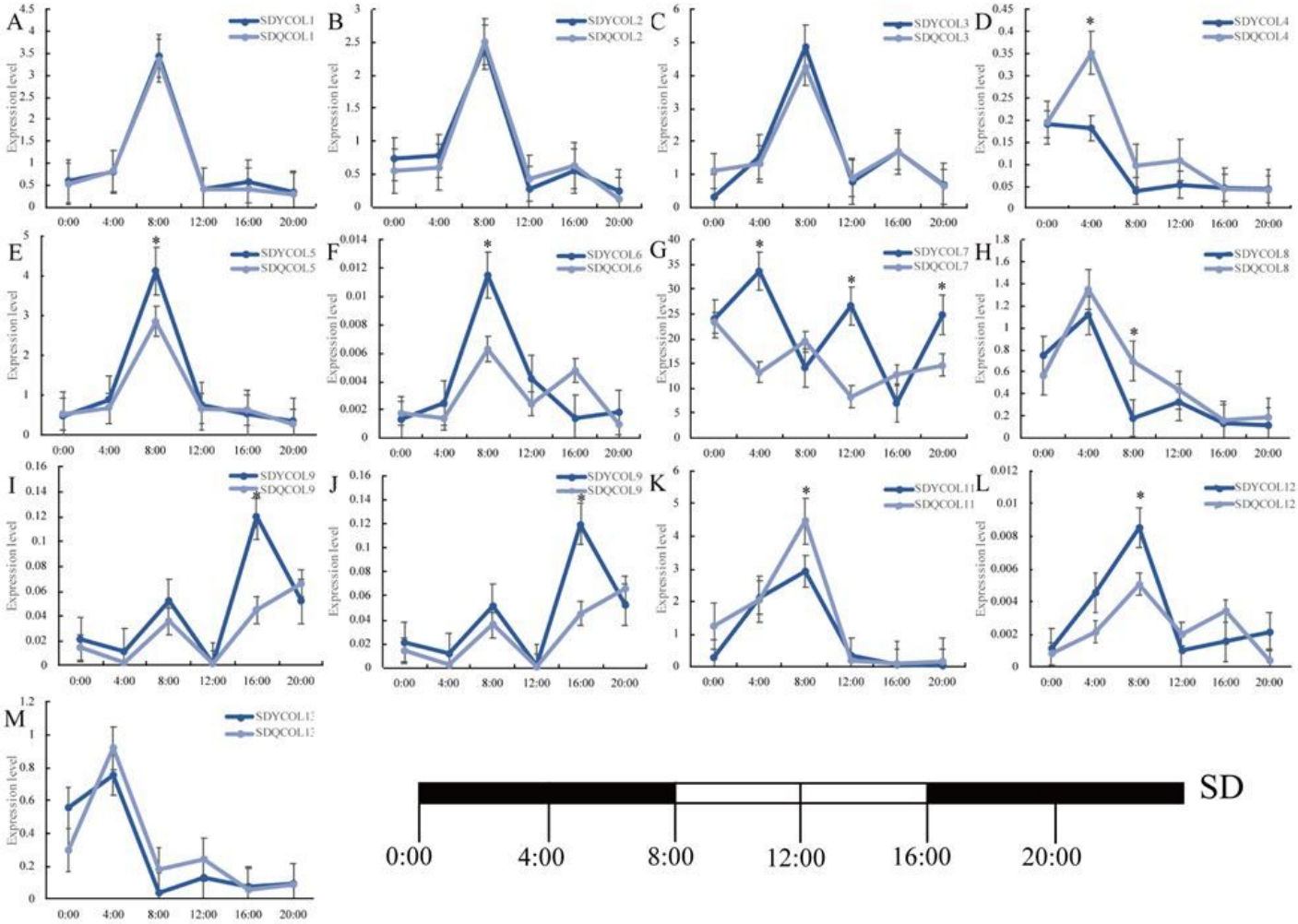

Figure 8

Expression patterns of the CsCOL gene of Yunma 7 and Qingma 1 under the SD condition (16 h dark/8 h light). SDQCOL1-13 represented the expression levels of CsCOLs of the “Qingma 1” under the SD condition; SDYCOL1-13 represented the expression levels of CsCOLs of the “Yunma 7” under the LD condition. Data are means (±SD), n=3. Significant differences were determined by one-way ANOVA: *P<0.05.

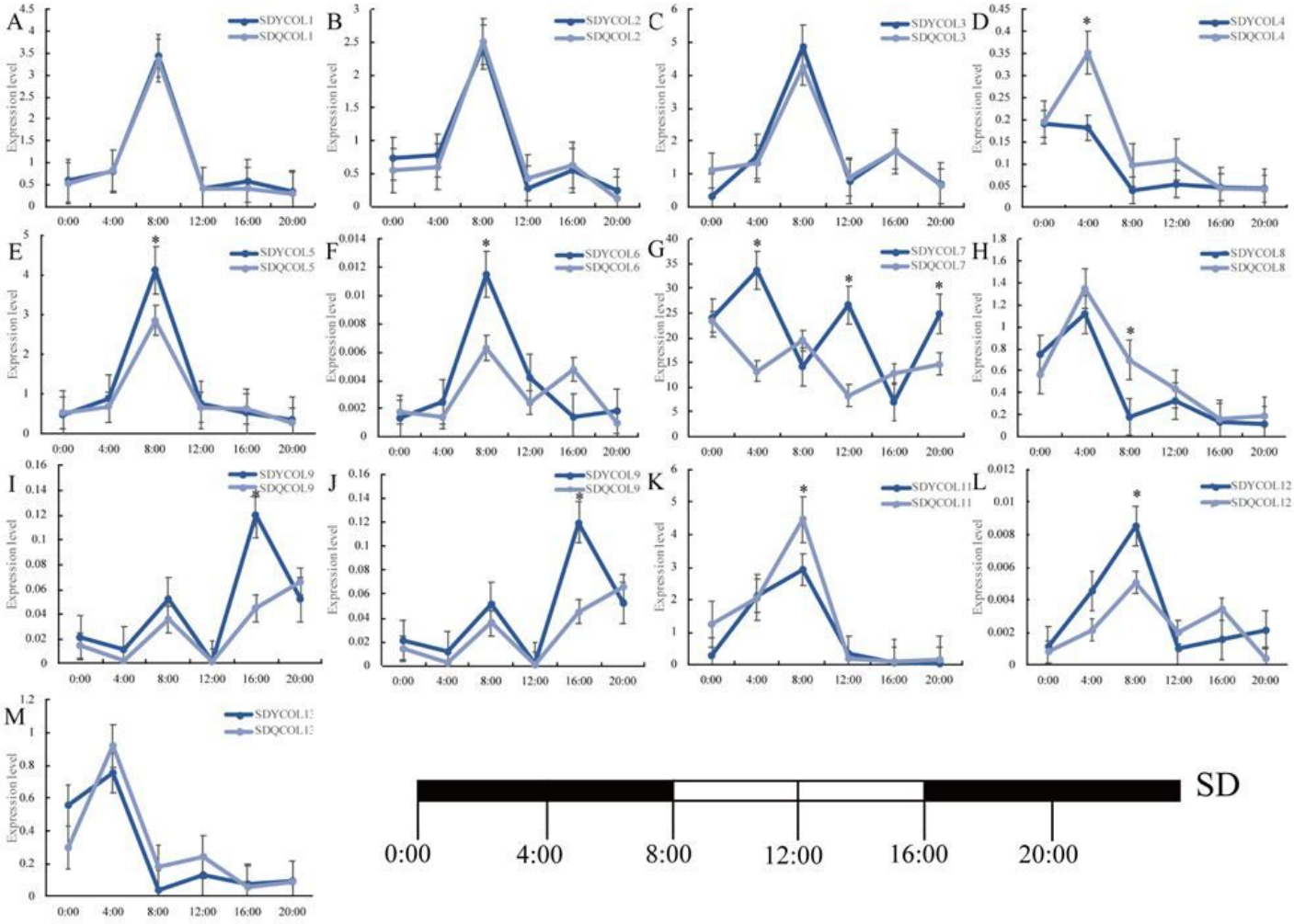

Figure 8

Expression patterns of the CsCOL gene of Yunma 7 and Qingma 1 under the SD condition (16 h dark/8 h light). SDQCOL1-13 represented the expression levels of CsCOLs of the “Qingma 1” under the SD condition; SDYCOL1-13 represented the expression levels of CsCOLs of the “Yunma 7” under the LD condition. Data are means (±SD), n=3. Significant differences were determined by one-way ANOVA: *P<0.05.

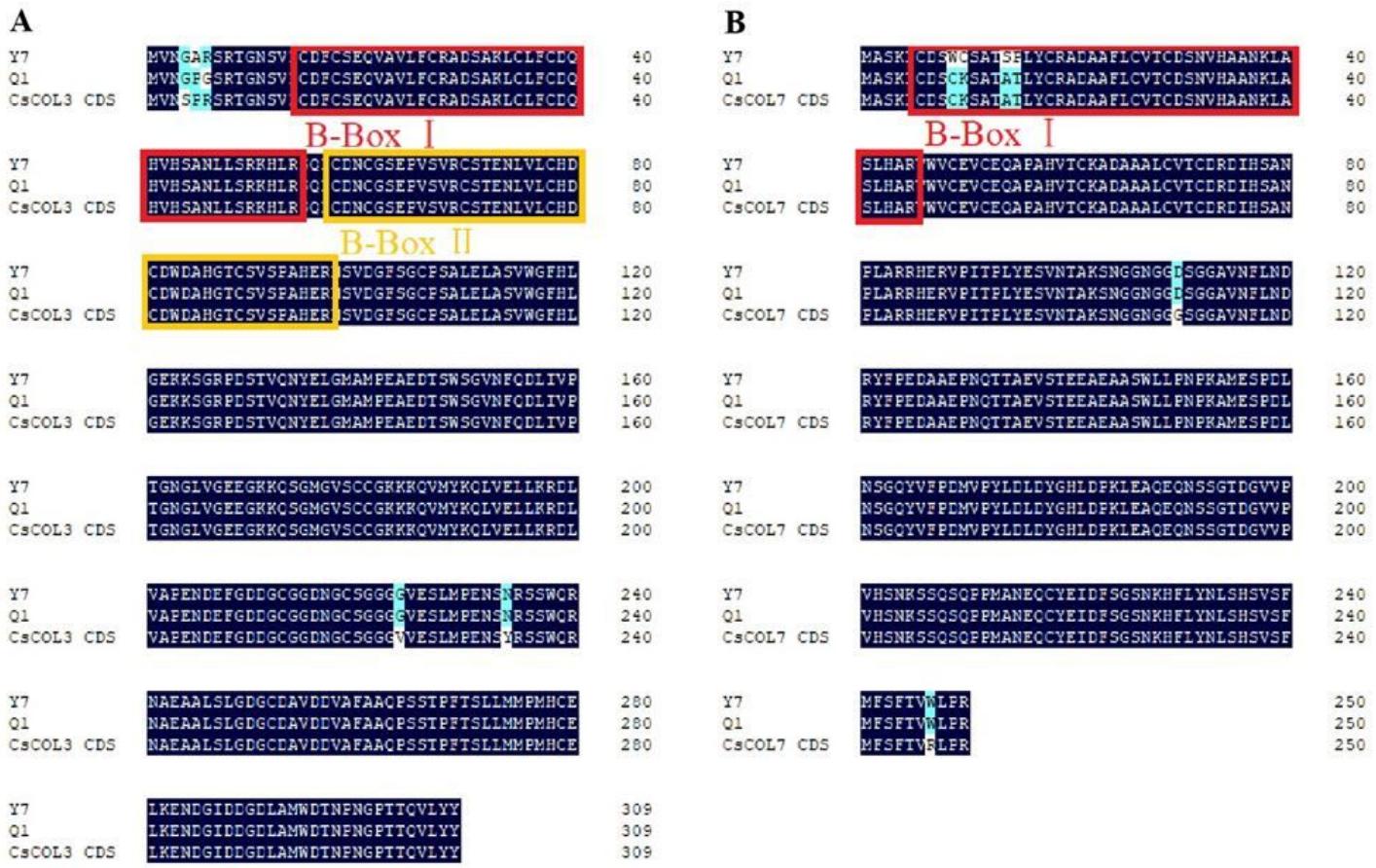

Figure 9

Comparison of the amino acid sequences of CsCOL3 (A) and CsCOL7 (B) between “Yunma 7 (Y7)” and “Qingma 1 (Q1)”.

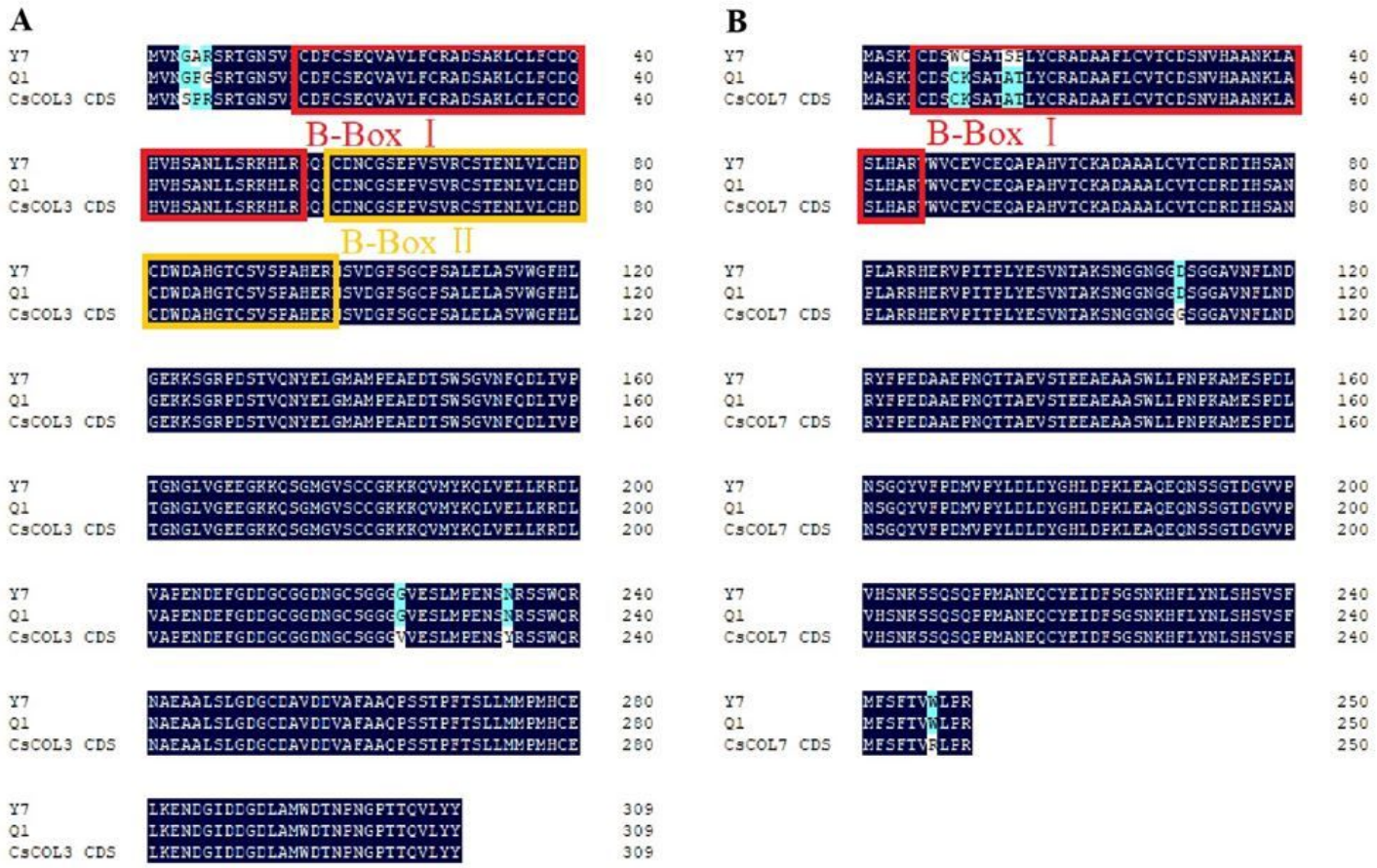

Figure 9

Comparison of the amino acid sequences of CsCOL3 (A) and CsCOL7 (B) between “Yunma 7 (Y7)” and “Qingma 1 (Q1)”.

Supplementary Files

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

- supportinformation.doc
- supportinformation.doc