Genome-Wide Identification and Expression Analyses of CONSTANS-Like Family Genes in Cucumber (Cucumis sativus L.)

Zhen Tian · Xiaodong Qin · Hui Wang · Ji Li · Jinfeng Chen

Received: 14 June 2020 / Accepted: 5 June 2021 / Published online: 15 June 2021
© The Author(s) 2021

Abstract
The CONSTANS-like (COL) gene family is one of the plant-specific transcription factor families that play important roles in plant growth and development. However, the knowledge of COLs related in cucumber is limited, and their biological functions, especially in the photoperiod-dependent flowering process, are still unclear. In this study, twelve CsaCOL genes were identified in the cucumber genome. Phylogenetic and conserved motif analyses provided insights into the evolutionary relationship between the CsaCOLs. Further, the comparative genome analysis revealed that COL genes are conserved in different plant species, especially collinearity gene pairs related to CsaCOL5. Ten kinds of cis-acting elements were vividly detected in CsaCOLs promoter regions, including five light-responsive elements, which echo the diurnal rhythm expression patterns of seven CsaCOL genes under SD and LD photoperiod regimes. Combined with the expression data of developmental stage, three CsaCOL genes are involved in the flowering network and play pivotal roles for the floral induction process. Our results provide useful information for further elucidating the structural characteristics, expression patterns, and biological functions of COL family genes in many plants.

Keywords COL genes · Cucumber · Expression pattern · Photoperiod-dependent flowering process

Abbreviations
CO CONSTANS
COL CONSTANS-like
DAS Days after sowing
FT FLOWERING LOCUS T
qRT-PCR Quantitative real-time PCR
SAM Shoot apical meristem

Introduction
Successful transition from vegetative to reproductive growth is important in a plant life cycle (Srikanth and Schmid 2011). It is affected by both external and internal factors, and among them, photoperiod (day length) is a pivotal environmental signal associated with inception and process of flowering. The regulation of flowering by photoperiod has been reported in many plants, such as Arabidopsis, rice and maize (Putterill et al. 1995; Yano et al. 2000; Jin et al. 2018). Despite the different responses to the day length in plant species, the specific molecular components are conserved in the photoperiod-dependent flowering pathway (Fu et al. 2015). For example, CONSTANS (CO) and FLOWERING LOCUS T (FT) genes were considered as the main regulators (Putterill et al. 1995; Corbesier et al. 2007), and the CO/FT module is conserved in many plants (Song et al. 2010). It is apparent that CO acts as a hub gene of the photoperiodic flowering network (Shim et al. 2017). At the transcriptional level, the expression abundance of CO gene is mainly regulated by FKF1-GI complex degrading CDFs gene (Imaizumi et al. 2005; Sawa et al. 2007). Later, multi-photoreceptor-mediated mechanisms have been revealed to be involved in the post-translational regulation level. The blue-light
photoreceptors cry1 and cry2, and phytochromes phyA and phyB, jointly regulate the stability of CO protein (Valverde et al. 2004; Shim et al. 2017). Through a complex regulatory network, CO proteins are accumulated and modulate the transcription of the downstream FT (Corbesier et al. 2007; Tamaki et al. 2007). The mobile florigen gene FT synthesized in the leaves is transmitted to the shoot apical meristem to initiate floral transition (Abe et al. 2005).

The CO and CONSTANS-like (COL) genes belong to plant-specific transcription factors, and they play diverse roles during the plant life cycle (Almada et al. 2009). Notably, many COL members participate in the photoperiodic flowering process; however, significant variation of the function has been reported. In Arabidopsis, CO plays a positive role and promotes flowering under LD condition (Puttermill et al. 1995); overexpression of COL5 can induce flowering in SD grown Arabidopsis (Hassidim et al. 2009). The study of COL1 and COL2 revealed that they have little effects on the flowering time (Ledger et al. 2001). Meanwhile, overexpression COL8 and COL9 result in late-flowering phenotype in Arabidopsis (Cheng and Wang 2005; Takase et al. 2011). The homologs of CO in other plant species are also thought to be involved in the photoperiod-associated flowering pathway. In rice, the Heading date 1 (Hd1) gene, the homologous of Arabidopsis CO gene, induces flowering under SD condition and exhibits the opposite response under LD treatment (Yano et al. 2000). Another homologous CO gene, HvCO9, contributes to delay flowering in barley (Kikuchi et al. 2012). In Lilium ×formolongi, three COL genes (LfCOL5, LfCOL6, and LfCOL9) are involved in initiating flowering induction under LD treatment (Li et al. 2018). In addition, the functional COL genes with flowering-inducing effects tend to belong to the group I members (Zhang et al. 2015; Chaurasia et al. 2016).

Cucumber is one of the most popular vegetables, and its fruits are rich in health-promoting properties and are consumed worldwide. It is generally considered to be a day-neutral flowering plant. However, the special Xishuangbanna cucumber (XIS, Cucumis sativus L. var. xishuangbannensis Qi et Yuan) is a typical SD flowering plant, and their flowering time is delayed in the temperate regions (Qi et al. 1983). Previous studies by Bo et al. (2015) and Pan et al. (2017) have identified QTLs related to photoperiod-dependent flowering time in XIS cucumber. Wang et al. (2020) suggested that the florigen gene CsaFT was presumably an important genetic determinant of flowering time variation when measured in four cucumber sub-groups. The key genes involved in the photoperiod-mediated flowering network are worth exploring in the photoperiod-sensitive XIS cucumber.

As the core gene of photoperiodic flowering pathway, CO and COL family genes need further investigation. In this study, we identified the CsaCOL family genes in the cucumber genome. The phylogenetic tree and conserved domain analysis were used to show the system classification and structure information of CsaCOL members. Collinearity analysis was performed to investigate the evolutionary relationship. Related cis-acting elements were found to predict the possible expression network of COL genes. The expression analysis and interaction network prediction were performed to understand the specific function of the CsaCOL genes. All these analyses provided insight into the regulation network of COL genes in the photoperiod-dependent flowering pathway.

Materials and Methods

Identification of CsaCOL Genes in the Cucumber Genome

All COL proteins were defined as genes containing both B-box and CCT domains (Robson et al. 2001). The hidden Markov model (HMM) program and related Pfam accession (B-box and CCT domains corresponding to PF00643.19 and PF06203.9) were used to find out all the CsaCOL genes in the cucumber genome database (http://cucurbitgenomics.org/organism/2, Cucumber genome sequence, Chinese Long, Version 2). All the selected CsaCOL proteins were further identified by Pfam database (http://pfam.xfam.org/) and Blastp in NCBI (https://www.ncbi.nlm.nih.gov/) to confirm the conserved domains B-box and CCT. In order to distinguish the CsaCOL genes, we named them based on the physical location on chromosomes in the cucumber genome. The ProtParam tool (http://web.expasy.org/protparam/) was used to provide basic information of the number of amino acids, molecular weight, theoretical iso-electric point (pI), and instability index (with a value < 40 considered as stable). Then the online tool PSORT (http://www.ps ort.com/psort.html) was performed to reveal the subcellular location information. All the basic information of CsaCOL genes is shown in Table 1.

Phylogenetic Tree Construction

The homologs of COL genes were obtained from Arabidopsis, watermelon, rice, tomato, and maize, and the corresponding websites are as follows: TAIR database (http://www.arabidopsis.org/), Cucurbit Genomics database (http://cucurbitgenomics.org/), Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/), Sol Genomics Network (https://solgenomics.net/), and Maize Genetics and Genomics Database (https://maizegdb.org/). The COL proteins from six plant species (Table S1) were used to construct the phylogenetic tree. All the amino acid sequences were aligned by Clustal W program and constructed by MEGA6 software.
with neighbor-joining method and 1000 times bootstrap replications.

**Conserved Motifs of CsaCOL Genes**

The basic information of CsaCOL genes in cucumber was obtained from Cucurbit Genomics database, such as physical location, sequences of amino acid, and nucleotide. The MEME online website (http://meme-suite.org/tools/meme) was used to identify the conserved motifs, with parameters’ arrangements as follows: maximum number of motifs, 10; minimum and maximum width, 6 and 200. Basic sequence information of motifs is listed in Table S2. Three conserved domains (B-box1, B-box2, and CCT motif) of CsaCOL proteins were aligned and presented by the WebLogo 3 online system (http://weblogo.threeplusone.com/) with default parameters (Crooks et al. 2004).

**Comparative Genome Collinearity Analysis**

To exhibit the collinearity relationship of the COL genes in cucumber and other five plant species (melon, watermelon, Arabidopsis, maize and rice), corresponding COL genes were mapped to chromosomes based on physical location from the database of Cucurbit Genomics Database, TAIR database, Maize Genetics and Genomics Database and Rice Genome Annotation Project. The collinearity analysis was realized through Perl and Python language in linux system.

**Analysis of the Cis-Acting Elements**

The upstream sequences (1500 bp) of CsaCOL genes’ were collected for analysis of cis-acting elements in their promoter region. Corresponding analysis was performed by PlantCARE program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; Lescot et al. 2002) and then exports the results with online tools Gene Structure Display Server program (GSDS2.0, http://gds.cbi.pku.edu.cn/index.php).

**Growth Conditions of Cucumber Plants and Sample Collection**

The cucumber inbred line ‘SWCC8,’ which belongs to XIS cucumber with property of short-day flowering, was used for the expression analysis. All seeds were sowed in the matrix (peat: vermiculite, 3:1) and then put in the incubator with 12 h/12 h (day/night) photoperiod regime, 28/18 °C temperature (day/night), a relative humidity of 80%, and 800 μmolm⁻²s⁻¹ photosynthetic photo flux density at Nanjing Agricultural University.

For studying the diurnal expression patterns of CsaCOL genes, the cucumber seedlings were transferred to two photoperiod conditions 8 h/16 h and 16 h/8 h (day/night) regimes when the first true leaf just appeared. One week later, the first true leaves were sampled every 4 h in 2 days, with three biological replicates at each time point. All samples were frozen in liquid nitrogen immediately and stored at −80 °C for RNA extraction.

To study the expression changes of CsaCOL members at different developmental stages, the second leaves counting from top were sampled from three biological replicates at each time point. All samples were frozen in liquid nitrogen immediately and stored at −80 °C for RNA extraction.

To study the expression changes of CsaCOL members at different developmental stages, the second leaves counting from top were sampled from three biological replicates every 10 days within 80 days after sowing (DAS). Here, the photoperiod regime 8 h/16 h (day/night), which was verified by Bo et al. (2010) for the proper flowering of XIS cucumber ‘SWCC8,’ was used and other conditions were the same as...
stated above. The slices of shoot apical meristem (SAM) at key developmental phases were selected and presented as the phase markers.

RNA Extraction, cDNA Synthesis and qRT-PCR Analysis

The above leave samples were ground into powder. Total RNA of these samples was extracted using Trizol Reagent (invitrogen), and then 1 μg RNA was used to synthesize a 20 μL cDNA system following the instructions in PrimeScript™ RT reagent Kit with gDNA Eraser (TAKARA). Quantitative real-time PCR (qRT-PCR) was carried out on Bio-Rad iCycler Real-Time PCR Detection System (USA) by TaKaRa SYBR Premix Ex Taq™ (Tli RNaseH Plus) with three biological replications. Total reaction system is 20 μL, containing 10 μL SYBR Premix (2×), 1 μL cDNA, 1 μL sense and anti-sense primer separately (10 μM), and 7 μL ddH2O. The qRT-PCR program was listed below, pre-denaturation at 95 °C for 1 min, followed by 40 cycles of denaturation at 95 °C for 10 s, annealing at 56 °C for 30 s, and extension at 72 °C for 30 s. Primer pairs were designed using Primer Premier 5.0, and NCBI blast program was used to identify the specificity of all primers (Table S3).

Prediction of Protein–Protein Interaction Network

Based on the interolog in Arabidopsis, the interactions between CsaCOL members and photoperiodic flowering-related genes were predicted by STRING database (http://string-db.org). Then the interaction network was presented by Cytoscape_v3.7.2 software (National Institute of General Medical Sciences, MD, USA). Combined with the expression results, the schematic diagram, including CsaCOL, and other critical genes and elements were presented.

Results

Basic Characterization of CsaCOL Genes in Cucumber

Twelve putative CsaCOL genes were finally identified in cucumber by the HMM program and then verified by Pfam and blastp database, with all the CsaCOL genes both harboring B-box and CCT domains. In order to distinguish the twelve genes, we named them CsaCOL1 to CsaCOL12 according to their physical location on chromosomes. Detailed information is presented in Table 1. The CsaCOL genes were distributed on six chromosomes of cucumber genome, with three in chromosome 1 (CsaCOL1-CsaCOL3) and 2 (CsaCOL4-CsaCOL6), two in chromosome 4 (CsaCOL7 and CsaCOL8) and 6 (CsaCOL10 and CsaCOL11), and one in chromosome 5 (CsaCOL9) and 7 (CsaCOL12), respectively. The amino acid sequences of CsaCOL proteins are between 319 (CsaCOL4) and 542 (CsaCOL11) in length, and the molecular weight is ranging from 35.16 kDa (CsaCOL4) to 60.05 kDa (CsaCOL11). The iso-electric points are various between CsaCOL genes, and the minimum and maximum iso-electric points are 5.24 (CsaCOL8) and 8.33 (CsaCOL10) separately. Protein instability index analysis showed that all CsaCOL members belong to instable proteins (instability index > 40). The predicted results of subcellular location presented that CsaCOL genes localized on nuclear, mitochondrial, and multiple other locations.

Phylogenetic Analysis of COL Genes

To understand the evolutionary relationship of the COL family genes, we constructed an unrooted neighbor-joining tree (Fig. 1) using 84 COL proteins from cucumber (12), watermelon (11), Arabidopsis (17), tomato (13), maize (16), and rice (15). All COL proteins have been verified by Pfam and Blastp database to contain both B-box and CCT domains (Table S1). In Arabidopsis, through sequence alignment, the COL members can be divided into three groups according to the divergence of B-box domain, with group I COLs containing two B-boxes; one normal B-box and another diverged B-box in group II; and group III COL members only including one B-box domain (Robson et al. 2001; Griffiths et al. 2003). Phylogenetic analysis showed that 84 COL proteins were indeed classified into three groups, with each group containing at least one COL protein from six different plant species. The distribution of cucumber CsaCOL proteins was five (CsaCOL2-CsaCOL5, CsaCOL8) in group I, four (CsaCOL6, CsaCOL7, CsaCOL11, CsaCOL12) in group II, and three (CsaCOL1, CsaCOL9, CsaCOL10) in group III. However, not all the COL proteins clustered into group I have two B-box domains. For example, CsaCOL8 was classified in group I but only contains one B-box domain. In this study, classification results based on phylogenetic tree are not exactly the same when compared with that in Arabidopsis (Robson et al. 2001; Griffiths et al. 2003).

Sequence Structure Analysis of CsaCOL Members

MEME program was used to predict the sequence structure information. Ten kinds of conserved motifs were presented by numbers from 1 to 10 (Fig. 2a; Table S2). All the COL members in cucumber include two conserved domains, one or two B-boxes (motif 1) at the N-terminus and one CCT domain (motif 2) near the C-terminus. In addition to the representative B-box and CCT domains, each group CsaCOLs has their specific motif composition. For example,
three CsaCOL proteins in group III, CsaCOL1, CsaCOL9, and CsaCOL10, share the same composition of motifs, and the specific motif is motif 4. All the CsaCOL members in group I contain a valine-proline motif (VP motif, motif 3) near the CCT domain in the C termini, which is important for the interaction with COP1 gene (Gangappa and Botto 2014). The conserved motifs of CsaCOL proteins show a certain similarity within groups.

The alignment of amino acid sequences of CsaCOLs identified three key domains of COL family genes, namely B-box1, B-box2, and CCT domains (Fig. S1). Then, the representative domains of COL proteins were shown separately (Fig. 2b). The CCT domain among the CsaCOLs is highly conserved, and the conservation is 45.24%. The amino acid sequences of B-box1 domain in the three groups are not identical, but the five cysteine residues (C in green) are fully conserved among twelve CsaCOL members, with the consensus sequence unified as C-X2-C-X8-C-X2-D-X-A-X-L-C-X2-C-D-X3-H-X8-H. The CCT and B-box1 are relatively conserved domains of COLs. Group III COL members have no B-box2 domain. In the group I and II, the B-box2 domains also contain five cysteine residues, even though the interval sequences are different, indicating that the cysteine residues are conserved in the B-box domains of the COL proteins. The sequence conservation in normal B-box2 is up to 78.95% (group I CsaCOLs) and only 34.48% in the divergent B-box2 (group II CsaCOLs). The representative differences of CsaCOL proteins are in the sequence of the B-box2.

Collinearity Analysis of COL Genes Between Cucumber and other Plant Species

The collinearity analysis was carried out between cucumber and five plant species (Fig. 3), including three dicots (melon, watermelon, and Arabidopsis) and two monocots (maize and rice). The number of COL collinearity gene pairs between cucumber and melon, watermelon, Arabidopsis, maize, and rice are 11, 10, 7, 1, and 0 in order (Table S4). Cucumber, melon, and watermelon belong to the cucurbit crops, and the comparative genomic analysis yields more conserved collinear blocks, along with more collinearity gene pairs of the COL genes (one on one). However, relatively few collinearity gene pairs were detected between cucumber and maize.
or rice. Multiple-to-one and one-to-many phenomena were detected in the collinearity analysis with the model plant Arabidopsis. For example, both CsaCOL3 and CsaCOL4 have a collinearity relationship with AthCOL5. In addition, CsaCOL9 has a collinearity relationship with both AthCOL6 and AthCOL16. The collinearity analysis between cucumber and cucurbit crops showed that COL genes in group I, II, and III participate in the formation of collinearity gene pairs; however, only group I and III COL members in cucumber produced collinearity with those in Arabidopsis and maize, implying that the COL sequences in group I and III are relatively conserved in the evolutionary history. In addition to the comparison of cucumber and rice, CsaCOL5 has a collinearity relationship in the other four comparisons, which indicates that CsaCOL5 is highly conserved among different plant species (Table S4).

**Cis-Acting Elements in the Promoter Regions of CsaCOL Genes**

The composition of cis-acting elements was detected in the CsaCOL genes’ promoter regions (Fig. 4; Table S5). Five elements evolved in light-responsive were found, such as G-Box, GT1-motif, GATA-motif, ACE, and 3-AF1-binding site, indicating that COL genes can be used as a light sensor in flowering plants (Simon et al., 2015). Among them, GT1-motif and G-box cis-acting elements were detected in seven and six promoter regions, respectively. Plant hormones-responsive elements evolved under the action of plant hormones.
elements were also detected in the promoter regions, mainly correlated to GA (GARE-motif), MeJA (CGTCA-motif), and Auxin (AuxRR-core). In addition, development-related element CAT-box was found in five promoter regions, and its specific function is related to meristem expression. The circadian element was detected in CsaCOL3, which is corresponding to the circadian expression pattern of COL genes (Campoli et al. 2012; Kikuchi et al. 2012). The compositions of cis-acting elements are related to the possible functions and expression patterns of the COL genes.

**Diurnal Rhythm Expression Patterns of CsaCOL Genes Under SD and LD Photoperiod Regimes**

Previous study suggested that the expression of COL genes has been shown to be affected by the circadian clock and...
light-dependent diurnal oscillations (Campoli et al. 2012; Kikuchi et al. 2012). In order to check the daily oscillation rhythm, we detected the expression levels of cucumber COL genes under SD (8 h/16 h) and LD (16 h/8 h) regimes every 4 h in 2 days.

Under SD treatment (Fig. 5; Table S6), eight CsaCOL genes showed a diurnal expression pattern. Six of them, CsaCOL1, CsaCOL2, CsaCOL3, CsaCOL5, CsaCOL6, and CsaCOL8, showed a significant diurnal rhythm, peaking at dawn, and descending to different troughs. In addition, the expression peak of CsaCOL9 and CsaCOL11 appeared after dusk, then reaching their troughs during the day. And both of CsaCOL9 and CsaCOL11 hold a relative high expression level when compared with above six rhythm expression genes.

Different diurnal expression profiles were presented under LD regime (Fig. 6; Table S7), and the number of genes with daily oscillation rhythm is also eight. CsaCOL1, CsaCOL2, CsaCOL3, CsaCOL5, CsaCOL6, and CsaCOL8 reached their peaks at the same time that was 4 h after dawn. CsaCOL9, CsaCOL11, and CsaCOL12 presented their expression peak at dusk or 4 h after dusk.

The diurnal rhythm expression patterns of CsaCOL genes showed different peaks and troughs under SD and LD photoperiod regimes. Summarizing the rhythm expression results, seven COLs in cucumber, CsaCOL1, CsaCOL2, CsaCOL3, CsaCOL5, CsaCOL8, CsaCOL9, and CsaCOL11, presented the diurnal expression patterns under both SD and LD photoperiod regimes.

**Expression Profiles of CsaCOL Genes at Different Developmental Stages**

In order to better understanding the functions of CsaCOL genes in the flowering pathway, their expression patterns at different developmental stages were analyzed. The short-day treatment 8 h/16 h (day/night) was considered to be a relatively suitable photoperiod regime for the flowering of XIS cucumber ‘SWCC8’ (Bo et al. 2010). The expression analysis was carried out under 8 h/16 h (day/night) regime.

Previous study showed that COL family genes play critical roles in the flowering induction process (Suárez-López et al. 2001; Li et al. 2018). In the pre-experiment, the average flowering time of ‘SWCC8’ was 80 DAS. Then the expression profiles were detected every 10 days until 80 DAS. Comparative analysis of the slices at 10 DAS (Fig. 7a) and 30 DAS (Fig. 7b) showed that the floral primordia appeared at 30 DAS, indicating that the cucumber plants changed from vegetative to reproductive growth at 30 DAS. The expression profiles of CsaCOL genes were detected, and their expression patterns could be classified into two types (Fig. 8; Table S8). In the first type, which showed their highest expression level before or after the floral induction stage, comprised CsaCOL2, CsaCOL3, CsaCOL5, and CsaCOL11. In the second type, the highest expression profile was presented close to the flowering induction phase (30 DAS), including CsaCOL1, CsaCOL8, and CsaCOL9.

Previous study showed that COL genes have been correlated with the flowering induction by affecting the
expression of florigen gene FT (Kobayashi et al. 1999). The expression level of CsaFT was also detected in this study (Fig. 8). The florigen gene CsaFT had an up-regulated expression trend around 30 DAS. The consistent higher transcripts’ accumulation of CsaFT and CsaCOL1, CsaCOL8, and CsaCOL9 indicated that the three CsaCOL genes may play positive roles in the flowering induction process. In addition, the expression level of CsaCOL8 is significantly higher when compared with other CsaCOL genes. And the opposite expression pattern of other four CsaCOL members (CsaCOL2, CsaCOL3, CsaCOL5, and CsaCOL11) and CsaFT suggested that they may inhibit the expression of CsaFT, thereby delaying the flowering process.

**Prediction the Interaction Network Between CsaCOL and Photoperiodic Flowering-Related Genes**

Previous study showed that CO was the hub gene of the photoperiod-mediated flowering pathways (Shim et al. 2017). To further elucidate the function of CsaCOL genes,
the interaction relationships between CsaCOL and photoperiodic flowering-related genes were predicted according to the network of Arabidopsis. Ultimately, three COL genes in cucumber, CsaCOL3, CsaCOL5, and CsaCOL8, participated in the photoperiodic flowering network. Genes with more interaction gene pairs are generally thought to be more important. Based on this principle, CsaCOL8 had more interaction partners (9) compared with CsaCOL3 (6) and CsaCOL5 (3), indicating that CsaCOL8 may be more active in the photoperiod-associated flowering network (Fig. S2; Table S9).

The predicted network was complex for containing multiple interaction gene pairs, so we selected the critical elements to draw the schematic diagram (Fig. 9). Light is perceived by the specialized photoreceptors, such as CsaPHYA, CsaPHYB, and CsaCRY1. The cucumber plants measure the light signal by the internal oscillators, i.e., the genes (CsaLHY, CsaTOC1, CsaGI, and CsaFKF1)
regulated by the circadian clock, all of which have direct or indirect effects on *CsaCOLs'* expression. *COL* genes affect the flowering process by regulating the expression of florigen gene *FT* (Kobayashi et al. 1999). Combined with the expression results (Fig. 8), *CsaCOL8* directly promotes the expression of *CsaFT*, while *CsaCOL3* and *CsaCOL5* suppress it.
Discussion

CO has the function of measuring day length, and meanwhile CO plays a central role during the photoperiod-regulated flowering process (Putterill et al. 1995; Suárez-López et al. 2001). Benefiting from the genome sequencing technology, CO and COL family genes have been found and characterized in different plants, for example, seventeen COL genes in Arabidopsis (Robson et al. 2001), sixteen and nine in rice and barley (Griffiths et al. 2003), thirteen in sugar beet (Chia et al. 2008), twelve in soybean (Wu et al. 2014), eleven in Chrysanthemum lavandulifolium (Fu et al. 2015), twenty five in Chinese cabbage (Song et al. 2015), and twenty in radish (Hu et al. 2018). The numbers of COL genes are not proportional to the size of corresponding plant genome (Table S1). In our study, ignoring the exceptions, most of the CsaCOL genes are structurally and evolutionarily conserved (Figs. 2, 3), and they have diverse expression patterns (Figs. 5, 6, 8).

Structural and Evolutionary Characteristics of the COL Genes

According to the definition of COL genes in Arabidopsis, all the COL genes have one CCT domain (C-terminus), and the differences come from the numbers of B-box domain (N-terminus), with group I COL members containing two B-box, group II one normal B-box, one diverged B-box and group III only including one B-box (Robson et al. 2001; Griffiths et al. 2003). Twelve COL genes in cucumber are also classified into three groups (Fig. 1), but not every CsaCOL gene fits the grouping principle in Arabidopsis. For example, CsaCOL8 just contains one B-box domain (Table S1), while CsaCOL8 is classified into group I in the phylogenetic analysis (Fig. 1). According to the previous analysis, the similar phenomenon occurs in multiple plants, such as Brachypodium distachyon, Brassica rapa, and Citrus clementina (Song et al. 2015).

Gene’s structure affects its function. Therefore a comprehensive understanding of its structural features is necessary. The study of COL genes in Lilium formosolongi showed that LfCOL13, LfCOL14 and LfCOL15, which only contain one B-box domain, were classified into group II (Li et al. 2018). In this study, the exception CsaCOL8 just includes one B-box domain and belongs to a member of group I (Fig. 1; Table S1). Since it might be insufficient to comprehend the features of a gene just based on the number of B-box or phylogenetic analysis. The VP motif was detected in all of the group I COL members in cucumber (Fig. 2a, motif 3), which also has been verified in the group I COL genes in Arabidopsis (Gangappa and Botto 2014) and Banana (Chaurasia et al. 2016). The VP motif may be one of the criteria for judging the structural characteristics of group I COL genes. In addition to VP motif, the nuclear localization signals (NLSs) and other novel motifs also play important roles in the function of COL genes (Crocco and Botto 2013). It has been suggested that the B-box and CCT domains are necessary structures of COL members, and the radiated variation into other motifs and phylogenetic analyses also should be taken into consideration.

The evolution and origin of COL genes in plants are well worth exploring. According to the sequence logo results (Fig. 2), the B-box1 and CCT domains, namely the structural features of group III COL genes, are highly conserved when compared with others. Previous study shows that early COL proteins only contain B-box1 and CCT domains in green plants (Crocco and Botto 2013), implying that group III members may be the origin of COL proteins. The high similarity between B-box2 and B-box1 in group I COLs suggested that B-box2 may originate from the replication event of B-box1. Because of the differences of amino acid sequence and protein length (Fig. 2b), the B-box2 in group II COL members may derived from the mutation of B-box2 in group I. In summary, we proposed that group III members are the early COL proteins; thereafter, through the replication event, the group I COLs are generated; finally, group II COL members appeared to be initiated by gene mutation. The origin and differentiation of COLs need further exploration.

Multiple Expression Patterns of COL Genes

Previous studies have shown that some of the COL genes exhibit a distinct diurnal rhythm expression pattern under
different photoperiod regimes (Suárez-López et al. 2001; Li et al. 2018). For example, in Arabidopsis, CO showed distinct diurnal expression patterns under LD and SD light regimes, with the expression peak occurring at dusk and night under LD condition; however, CO reached its peak only at night under SD treatment (Turck et al. 2008). In the LD plant Lilium × formolongi, the rhythm expression patterns of LfCOL genes were also different (Li et al. 2018). In this study, under 8 h/16 h and 16 h/8 h regimes, the diurnal expression patterns of CsaCOL genes were varied in the time points and the expression level of the peak (Figs. 5, 6). These results suggested that CsaCOL genes are sensitive to photoperiod, thus, intriguing different responses according to the external photoperiod conditions. The discovery of light-responsive cis-acting elements also demonstrates that the CsaCOL genes typically respond to photoperiod changes (Fig. 4).

The COL genes play critical roles in the photoperiodic flowering induction process (Suárez-López et al. 2001); however, the functions of COL genes are diverse among plant species (Chaurasia et al. 2016; Li et al. 2018). Some of the studies highlighted the importance of group I COL members. For example, in wild and domesticated cotton, eight COL genes in group I were demonstrated to be involved in the photoperiod-regulated flowering process (Zhang et al. 2015). Previous study of banana showed that the group I type COL genes were also regulated by the photoperiod regime (Chaurasia et al. 2016). However, not all the COL genes, participating in the flowering network, belong to group I members. In Lilium × formolongi, COL genes in group I (LfCOL5), II (LfCOL9), and III (LfCOL6) play positive roles in the flowering induction process. In this study, group I (CsaCOL2, CsaCOL3, CsaCOL5, CsaCOL8), II (CsaCOL11), and III (CsaCOL1, CsaCOL9) COL genes were involved in the flowering induction network, and the functions of CsaCOL members were varied (Fig. 8). These results demonstrated that CsaCOL genes have different roles, and even the same group members may have diverse functions.

Conclusions

In conclusion, this study systematically analyzed the COL family genes in cucumber. The twelve CsaCOL genes are distributed on six chromosomes of the cucumber genome and can be divided into three groups through phylogenetic analysis. Conserved domain analysis reveals high similarity of CsaCOLs within groups, and the COL genes are evolutionarily conserved in different plants. Multiple CsaCOL members showed light-dependent diurnal rhythm changes, which echo the detection of light-responsive cis-acting elements in their promoter regions. Three CsaCOL genes are involved in the flowering network; notably, CsaCOL8 plays a positive role at the flowering induction phase. Our study provides a comprehensive understanding of cucumber COL genes on structure and expression regulation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00344-021-10420-4.

Acknowledgements This work was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions. The authors would like to thank Heshon Ochieng Obel for the revision of language.

Author Contributions JFC, JL, and ZT planned and designed the work. ZT, XDQ, and HW participated in the preparation of plant materials. ZT and HW conducted all experiments. ZT, XDQ, and HW analyzed the data. JFC, JL, and ZT wrote the manuscript. All authors read and approved the manuscript.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. You may not use the material in any way that suggests the licensor endorses you or your use of the material. If you wish to use material from this article outside the scope of the Creative Commons licence, permission from the copyright holder may be required. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309(5737):1052–1056. https://doi.org/10.1126/science.1115983

Almada R, Cabrera N, Casaretto JA, Ruiz-Lara S, Gonzalez Villanueva E (2009) VvCO and VvCOL1, two CONSTANS homologous genes, are regulated during flower induction and dormancy in grapevine buds. Plant Cell Rep 28(8):1193–1203. https://doi.org/10.1007/s00299-009-0720-4

Bo KL, Chen LZ, Qian CT, Zhang SX, Chen JF (2010) Short-day treatments induce flowering of Xishuangbanna cucumber. China Cucurbits Veg 23(4):1–3

Bo KL, Ma Z, Chen JF, Weng YQ (2015) Molecular mapping reveals structural rearrangements and quantitative trait loci underlying traits with local adaptation in semi-wild Xishuangbanna cucumber (Cucumis sativus L. var. xishuangbannensis Qi et Yuan). Theor Appl Genet 128(1):25–39. https://doi.org/10.1007/s00122-014-2410-z

Campoli C, Drosse B, Searle I, Coupland G, von Korff M (2012) Functional characterisation of HvCO1, the barley (Hordeum vulgare)
flowering time ortholog of CONSTANS. Plant J 69(5):868–880. https://doi.org/10.1111/j.1365-313X.2011.04839.x

Chaurasia AK, Patil HB, Azeez A, Subramaniam VR, Krishna B, Sane AP, Sane PV (2016) Molecular characterization of CONSTANS-Like (COL) genes in banana (Musa acuminata L., AAA Group, cv. Grand Nain). Physiol Mol Biol Plants 22(1):1–15. https://doi.org/10.1007/s12298-016-0345-3

Cheng XF, Wang ZY (2005) Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in Arabidopsis thaliana. Plant J 43(5):758–768. https://doi.org/10.1111/j.1365-313X.2005.02491.x

Chia TY, Müller A, Jung C, Mutasa-Gottgens ES (2008) Sugar beet contains a large CONSTANS-LIKE gene family including a CO homologue that is independent of the early-bulbing (B) gene locus. J Exp Bot 59(10):2735–2748. https://doi.org/10.1093/jxb/ern129

Corbesier L, Vincent C, Jang SH, Fornara F, Fan QZ, Searle I, Giai-kounitis A, Farron a S, Gisso t L, Turnbull C, Coupland G (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science 316(5827):1030–1033. https://doi.org/10.1126/science.1141752

Crocce CD, Botto JF (2013) BBX proteins in green plants: insights into their evolution, structure feature and functional diversification. Gene 531(1):44–52. https://doi.org/10.1016/j.gene.2013.08.037

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14(6):1188–1190. https://doi.org/10.1101/gr.139004

Fu JX, Wei LW, Dai SL (2015) Identification and characterization of the CONSTANS-like gene family in the short-day plant Chrysanthemum lavandulifolium. Mol Genet Genomics 290(3):1039–1054. https://doi.org/10.1007/s00438-014-0977-3

Gangappa SN, Botto JF (2014) The BBX family of plant transcription factors. Trends Plant Sci 19(7):460–470. https://doi.org/10.1016/j.tplants.2014.01.010

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant Physiol 131(4):1855–1867. https://doi.org/10.1104/pp.102.016188

Hassidim M, Harir Y, Yakir E, Kron I, Green RM (2009) Over-expression of CONSTANS-LIKE 5 can induce flowering in short-day grown Arabidopsis. Planta 230(3):481–491. https://doi.org/10.1007/s00425-009-0958-7

Hu TH, Wei QZ, Wang WH, Hu JH, Mao WH, Zhu QM, Bao CL (2018) Genome-wide identification and characterization of CONSTANS-like gene family in radish (Raphanus sativus). PLoS ONE 13(9):e0204137. https://doi.org/10.1371/journal.pone.0204137

Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA (2005) FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. Science 309(5732):293–297. https://doi.org/10.1126/science.1110586

Ji D, Liu XG, Jia W, Liu HJ, Li QW, Peng Y, Du YF, Wang YB, Yin YJ, Zhang XH, Liu Q, Deng M, Li N, Cui XY, Hao DY, Yan JB (2018) ZmCOL3, a CCA gene represses flowering in maize by interfering circadian clock and activating expression of ZmCCT. J Integr Plant Biol 60(6):465–480. https://doi.org/10.1111/jipb.12632

Kikuchi R, Kawahigashi H, Oshima M, Ando T, Handa H (2012) The differential expression of HvCO9, a member of the CONSTANS-like gene family, contributes to the control of flowering under short-day conditions in barley. J Exp Bot 63(2):773–784. https://doi.org/10.1093/jxb/err299

Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286(5446):1960–1962. https://doi.org/10.1126/science.286.5446.1960

Ledger S, Strayer C, Ashton F, Kay SA, Putterill J (2001) Analysis of the function of two circadian-regulated CONSTANS-LIKE genes. Plant J 26(1):15–22. https://doi.org/10.1046/j.1365-313x.2001.01003.x

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombouts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30(1):325–327. https://doi.org/10.1093/nar/30.1.325

Li YF, Zhao YQ, Zhang M, Jia GX, Zaccai M (2018) Functional and evolutionary characterization of the CONSTANS-LIKE family in Lilium×formollongi. Plant Cell Physiol 59(9):1874–1888. https://doi.org/10.1093/pcp/pcy105

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25(4):402–408. https://doi.org/10.1016/S1046-5923(01)00038-9

Pan YP, Qu SP, Bo KL, Gao ML, Haider KR, Weng YQ (2017) QTL mapping of domestication and diversifying selection related traits in short-fruit sne-wild Xishuangbanna cucumber (Cucumis sativus L. var. xishuangbanensis). Theor Appl Genet 130(7):1531–1548. https://doi.org/10.1007/s00122-017-2908-2

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The CONSTANS gene of arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80(6):847–857. https://doi.org/10.1016/0092-8674(95)90288-0

Qi CZ, Yuan ZZ, Li YX (1983) A new type of cucumber-Xishuangbanna cucumber. Acta Hortic Sin 10(4):259–264

Robwon F, Costa MMR, Hepworth SR, Vizir I, Pineiro M, Reeves PH, Putterill J, Coupland G (2001) Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by mutant alleles and transgenic plants. Plant J 28(6):619–631. https://doi.org/10.1046/j.1365-313x.2001.01163.x

Sawa M, Nusinow DA, Kay SA, Imaizumi T (2007) FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science 318(5848):261–265. https://doi.org/10.1126/science.1146994

Shim JS, Kubota A, Imaizumi T (2017) Circadian clock and photoperiodic flowering in Arabidopsis: CONSTANS is a hub for signal integration. Plant Physiol 173(1):5–15. https://doi.org/10.1104/pp.16.01327

Simon S, Rühl M, de Montaigu A, Wötzel S, Coupland G (2015) Evolution of CONSTANS regulation and function after gene duplication produced a photoperiodic flowering switch in the Brassicaceae. Mol Biol Evol 32(9):2284–2301. https://doi.org/10.1093/molbev/msv110

Song YH, Ito S, Imaizumi T (2010) Similarities in the circadian clock and photoperiodism in plants. Curr Opin Plant Biol 13(5):594–603. https://doi.org/10.1016/j.rpb.2010.05.004

Song XM, Duan WK, Huang ZN, Liu GF, Wu P, Liu TK, Li Y, Hou XL (2015) Comprehensive analysis of the flowering genes in Chinese cabbage and examination of evolutionary pattern of CO-like genes in plant kingdom. Sci Rep 5(1):1–16. https://doi.org/10.1038/srep14631

Srikant R, Schmid M (2011) Regulation of flowering time: all roads lead to Rome. Cell Mol Life Sci 68(12):2013–2037. https://doi.org/10.1007/s00018-011-0673-y

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature 410(6832):1116–1120. https://doi.org/10.1038/35074138

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.
conditions. Plant Biotechnol 28:439–446. https://doi.org/10.5511/plantbiotechnology.11.0823b
Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K (2007) 
Hd3a protein is a mobile flowering signal in rice. Science 
316(5827):1033–1036. https://doi.org/10.1126/science.1141753
Turck F, Fornara F, Coupland G (2008) Regulation and identity of 
florigen: FLOWERING LOCUS T moves center stage. Annu Rev 
Plant Biol 59:573–594. https://doi.org/10.1146/annurev.arplant. 
59.032607.092755
Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupl 
land G (2004) Photoreceptor regulation of CONSTANS protein in 
photoperiodic flowering. Science 303(5660):1003–1006. https:// 
doi.org/10.1126/science.1091761
Wang SH, Li HB, Li YY, Li Z, Qi J, Lin T, Yang XY, Zhang ZH, 
Huang SW (2020) FLOWERING LOCUS T improves cucumber 
adaptation to higher latitudes. Plant Physiol 182(2):908–918. 
https://doi.org/10.1104/pp.19.01215
Wu F, Price BW, Haider W, Seufferheld G, Nelson R, Hanzawa Y 
(2014) Functional and evolutionary characterization of the 
CONSTANS gene family in short-day photoperiodic flowering in 
soybean. PLoS ONE 9(1):e85754. https://doi.org/10.1371/journ 
al.pone.0085754
Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, 
Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) 
Hd1, a major photoperiod sensitivity quantitative trait locus in 
rice, is closely related to the Arabidopsis flowering time gene 
CONSTANS. Plant Cell 12(12):2473–2483. https://doi.org/10. 
1105/tpc.12.12.2473
Zhang R, Ding J, Liu CX, Cai CP, Zhou BL, Zhang TZ, Guo WZ 
(2015) Molecular evolution and phylogenetic analysis of eight 
COL superfamily genes in group I related to photoperiodic regu 
lation of flowering time in wild and domesticated cotton (Gos 
sypium) species. PLoS ONE 10(2):e0118669. https://doi.org/10. 
1371/journal.pone.0118669

Publisher’s Note Springer Nature remains neutral with regard to 
jurisdictional claims in published maps and institutional affiliations.