Structural And Functional Analysis of CCT Family Genes In Pigeonpea

Kishor U. Tribhuvan

ICAR-National Institute for Plant Biotechnology (NIPB), New Delhi 110012, Delhi; ICAR-Indian Institute of Agricultural Biotechnology (IIAB), Ranchi 834010, Jharkhand

Tanvi Kaila

ICAR-National Institute for Plant Biotechnology (NIPB), New Delhi 110012, Delhi

Harsha Srivastava

ICAR-National Institute for Plant Biotechnology, New Delhi 110012, Delhi

Antara Das

ICAR-Central Institute for Subtropical Horticulture, Lucknow 226101, Uttar Pradesh

Kuldeep Kumar

ICAR-National Institute for Plant Biotechnology, New Delhi 110012, Delhi

Kumar Durgesh

ICAR-Indian Agricultural Research Institute (IARI), New Delhi 110012, Delhi

Rekha Joshi

ICAR-Indian Agricultural Research Institute (IARI), New Delhi 110012, Delhi; ICAR Central Agricultural Research Institute

Binay K. Singh

ICAR-Indian Institute of Agricultural Biotechnology, Ranchi 834010, Jharkhand

Nagendra K. Singh

ICAR-National Institute for Plant Biotechnology (NIPB), New Delhi 110012, Delhi

Kishor Gaikwad (✉ kish2012@gmail.com)

ICAR-National Institute for Plant Biotechnology (NIPB), New Delhi 110012, Delhi

Research Article

Keywords: Flowering-related genes, Photoperiod sensitivity, CCT family genes, Pigeonpea

Posted Date: July 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-724094/v1>

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

Version of Record: A version of this preprint was published at Molecular Biology Reports on November 20th, 2021. See the published version at <https://doi.org/10.1007/s11033-021-06860-6>.

Abstract

Pigeonpea (*Cajanus cajan* (L) Millsp) is a short-day plant in which the flowering is highly sensitive to photoperiod. A better understanding of the genes modulating photoresponse and flowering time is critical to developing photoperiod insensitive pigeonpea cultivars for cultivation across the seasons. We identified 33 CCT family genes (*CcCCT1- CcCCT33*) in *C. cajan* and localized them on 10 chromosomes and nine genomic scaffolds. The structural analysis of CCT family genes revealed a considerable variation in length and distribution of exons and introns. Based on the type of domain(s), we classified the CCT family genes into CCT motif family (CMF) type, CONSTANS like (COL) type, Pseudo-response Regulator (PRR) type, and GATA and tifi containing CCT (GTCC) type. The CCT family genes of *C. cajan* exhibited an extensive orthologous relationship with the CCT family genes of other legume species. We also observed significant sharing of CCT family genes among the legume species. *Glycine max* exhibited the maximum sharing of CCT family genes with *C. cajan*. The analysis of CCT family proteins-based phylogenetic relationships revealed a general congruence with the legumes' taxonomic relationships. The expression analysis of CCT family genes of pigeonpea demonstrated that *CcCCT4* and *CcCCT23* are the active CONSTANS (CO) in ICP20338. In contrast, only *CcCCT23* is active in MAL3, explaining the differential response of ICP20338 and MAL3 to photoperiod. The chromosomes of *C. cajan* contain a variable number of CCT family genes. A majority of these genes are localized in the centromeric regions. The COL type CCT genes are structurally highly diverse and contain a variable number of B-box domains. The CCT family genes of different legume species exhibit all three kinds of relationships: one-to-one, many-to-one, and many-to-many types. The photoperiod insensitive cultivar ICP20338 contains *CcCCT4* and *CcCCT23*, while the photoperiod sensitive cultivar MAL3 contains only *CcCCT23* as active CONSTANS (CO), which may be the plausible reason for their differential photoperiod response.

Introduction

Flowering is a critical event in a plant's life cycle [1]. It is regulated by many intrinsic and extrinsic factors. The intrinsic factors comprise mainly plant's age, hormonal balance, and nutritional status, while the extrinsic factors include, majorly, temperature and photoperiod [2]. Photoperiod-mediated flowering regulation occurs through modulation of the expression of a cascade of flowering-related genes [3]. The major flowering-related genes are *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), *Pseudo-response Regulator* (*PRR*), *FLAVIN-BINDING KELCH REPEAT F-BOX 1* (*FKF1*), *LATE ELONGATED HYPOCOTYL* (*LHY*), *CYCLING DOF FACTORS* (*CDFs*), *TOC1*, *GIGANTEA* (*GI*), *FLOWERING LOCUS T* (*FT*), and *CONSTANS* (*CO*) [4]. Among these, *FT* and *CO* are the primary flowering-related genes. These genes are the most studied flowering-related genes and well-characterized in Arabidopsis [5, 6]. The *FT* gene product 'florigen' triggers the induction of flowering [7]. It is synthesized in the leaves' companion cells and translocated to the shoot apex through the phloem [6, 8, 9]. At the shoot apex, it interacts with 14-3-3, a class of tetratricopeptide repeat (TPR) proteins and FD proteins, and forms a 'florigen activation complex' (FAC), which induces transcription of floral meristem identity genes such as *APETALA1* (*AP1*) and *LFY*, leading to flower development [10]. Activation of *FT* transcription in response to favorable photoperiod conditions is mediated primarily through the CO transcription factors [11]. The CO transcription factors sense the favorable photoperiod condition through their interactions with the products of several photo-inducible genes like *GI*, *FKF1*, *CCA1*, *LHY*, and *PRR* [12–17]. The CO transcription factor gene is a member of the *CCT* (CONSTANS, CO LIKE; COL, and TIMING OF CAB; TOC1) gene family, possessing characteristic CCT and B-box domains in their protein products [18]. The CCT family genes include CONSTANS like (COL), PRR, CCT motif family (CMF), and GATA/tifi containing CCT (GTCC) genes [19]. These genes regulate photoperiod-dependent flowering in various plant species [20–21]. Moreover, they play vital roles in various physiological processes, including stomatal opening, photomorphogenesis, abiotic stress response, hormone metabolism, and help determine an array of morphological traits like shoot branching, sink size, plant height, and grain number [19].

Cajanus cajan (L) Millsp., commonly known as pigeonpea or red gram, is a highly drought-tolerant pulse crop mostly cultivated in the world's tropical and subtropical parts [22]. Globally, it is grown in about 5.61 million hectares producing nearly 4.42 million tons of seed grain. India is the largest producer, contributing around 74.8% (3.31 million tons) of the global pigeonpea production [23]. Still, it imports more than 95% of the total global import volume of pigeonpea to meet its domestic demand. We can improve the situation only by enhancing pigeonpea's productivity in India, which hovers around 780kg/ha for the last five decades (<http://faostat.fao.org/>). Sensitivity to biotic and abiotic stresses, and most importantly, strong photoperiod sensitivity are the critical constraints limiting the country's productivity of pigeonpea. Pigeonpea requires short-day (< 12hr) photoperiod conditions for induction of flowering [24]. Nevertheless, variation in terms of photoperiod response exists among the pigeonpea germplasm and its wild relatives. Improvement in understanding of the molecular mechanisms underlying photoperiod response in pigeonpea and identifying the trait determinants is critical for breeding photoperiod insensitive cultivars, thereby enabling its cultivation across the seasons with varying photoperiod regimes. Towards this, we report the identification, chromosomal localization, and structural and functional analysis of CCT family genes in pigeonpea. We also report an account of the orthologroups, intergeneric/interspecific sharing and phylogenetic relationships of CCT family genes of pigeonpea and other legumes.

Material And Methods

2.1. Plant materials

Three biological replicates of photoperiod insensitive (ICP20338) and photoperiod sensitive (MAL3) genotypes [25] of *C. cajan* were grown under long-day (16hr light /8hr dark) and short day (12hr light/12hr dark) photoperiod conditions with 30°C temperature and 60% relative humidity at National Phytotron Facility, ICAR-IARI, New Delhi. We isolated the total RNA from the vegetative leaf (VL) and reproductive leaf (RL) tissues from both the short day (SD) and the long day (LD) grown plants. Since ICP20338 enters the reproductive phase 45–50 days after sowing (DAS), we harvested the VL tissues 30 DAS and RL tissues 50 DAS from both the SD and LD-grown plants. The photoperiod sensitive genotype MAL3 enters the reproductive phase 115–120 DAS; hence, we harvested the VL tissues from MAL 3 50 DAS and RL tissues 120 DAS from both the SD and LD- grown plants.

2.2. RNA isolation and cDNA synthesis

Total RNA was isolated from VL and RL tissues using Spectrum Plant Total RNA Kit (Sigma Aldrich, USA). We removed the residual DNA in the total RNA through DNase (Thermo Fisher Scientific, Waltham, USA) treatment and quantified the total RNA using Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA). We assessed the quality and the integrity of the isolated RNA through electrophoresis on 1.2% denaturing agarose gel. The cDNA synthesis was performed using the RevertAid first-strand cDNA synthesis kit (Thermo Fisher Scientific, Waltham, USA) as per the manufacturer's instructions.

2.3. Genome-wide identification of CCT family genes

We identified the CCT family genes in the *C. cajan* genome available at the Legume Information System (LIS) database (<https://legumeinfo.org/>) using 'CCT DOMAIN PROTEIN' as a keyword. The genes were located on the chromosomes using the BLAST search option available at the LIS database. We named all the CCT family genes with a prefix Cc representing *Cajanus cajan* and a numeral suffix in ascending order denoting their chromosomal positions [26]. The chromosome-wise distribution of CcCCT family genes was depicted using Mapchart software 2.3 [27]. The protein sequences of CCT family genes of *Vigna angularis*, *Vigna radiata*, *Vigna unguiculata*, *Glycine max*, *Medicago truncatula* and *Cicer arietinum* were downloaded from the LIS database. The protein sequences of CCT family genes of *Arabidopsis thaliana* were downloaded from the Uniprot protein database (<https://www.uniprot.org/uniprot/>).

2.4 Identification of centromere positions of *C. cajan* chromosomes

We identified the centromere positions of the *C. cajan* chromosomes using the procedure given by Melters et al. [28]. The chromosome-wise FASTA files of *C. cajan*, downloaded from the NCBI database (accession: GCA_000340665.1), were used as a local database, while the centromeric repeat sequence of *C. cajan* (ACTTGCTACACCTGGGGAGACTAATAACCAACACAGATGCACAACATAGCATGTAATTGGTTTACTGTTTCATTGGTTCTCTCTAATTCTTCAACTGACTT) identified by Melters et al. [28] was used as the query sequence. The blast hits showing more than 85% similarity were designated as the centromeric positions of the chromosomes.

2.5 Domain prediction, classification, and structural analysis of CcCCT family genes

We predicted conserved domains in the amino acid sequences of the CcCCT family genes using the 'Web CD-search tool' available at <https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>, and the exon-intron structures in the CcCCT family genes using the online software Gene Structure Display Server 2.0 (GSDS 2.0) (<http://gsds.cbi.pku.edu.cn/>). We classified the CcCCT family genes into COL, CMF, PRR, and GATA and tifi containing CCT (GTCC) types based on the presence of additional domains and/or CCT domain.

2.6 Prediction of orthologous relationships, intergeneric/interspecies sharing, and phylogenetic analysis of CCT family genes in legume species

We compared the amino-acid sequences of the CCT family proteins of the seven common legumes, including *C. cajan*, *V. angularis*, *V. radiata*, *V. unguiculata*, *G. max*, *M. truncatula* and *C. arietinum*, and the non-legume model species *A. thaliana* through the pair-wise BLASTP analysis keeping the sequence similarity and the E-value cutoff minimum at 90% and 1e-10, respectively. We presented the connectome using Circos software (version 0.69). The orthologous relationship between the CCT family genes of different species was inferred using the OrthoFinder software (version 2.2.6). The multiple sequence alignments of amino acid sequences of the CCT family genes of seven common legumes and *A. thaliana* were performed using the Clustalw software. The phylogenetic analysis was performed using the MEGA 10.1 software [29], and the tree was constructed using the maximum likelihood method with the bootstrap value of 1000.

2.7 Expression analysis of CcCCT family genes using transcriptome data

The RNA-seq data generated from the RL tissues of the genotypes ICP20338 and MAL3 grown under SD and LD conditions were analyzed to determine the level of expression of the CcCCT family genes. Firstly, we identified the CcCCT family genes' transcripts corresponding to the CDS through BLAST-based sequence identification analysis. The transcripts' ID was then used to extract the expression values (FPKM) of the CcCCT family genes from all the four samples derived from ICP20338-SD, ICP20338-LD, MAL3-SD, and MAL3-LD used in the study.

2.8 Expression analysis of CcCCT family genes through qRT-PCR

We designed the qRT-PCR primers for the CcCCT family genes using PrimerQuest™ Tool (<https://eu.idtdna.com/Primerquest/Home/Index>) of Integrated DNA Technologies (IDT). The primers were custom synthesized at the Bioengineering Services, India Private Ltd. The details of qRT-PCR primers are given in Supplementary Table 1. The qRT-PCR assay was performed in a Light Cycler qRT-PCR (Roche) instrument using a Brilliant III Ultra-Fast SYBR Green qPCR master mix kit (Agilent Technologies, USA). Each qRT-PCR reaction comprised of 0.5µl of each primer, 2µl of cDNA, 10µl SYBR green master mix, and 7.5µl DEPC treated water. We used the *tubulin* gene-specific primers of *C. cajan* as an internal control. The qRT-PCR reaction conditions were set at 94°C for 3 min, followed by 40 cycles of 94°C for 30s, 60°C for 15s and 72°C for 20s. The melting curve analysis followed with 10s at 95°C and then 10s each at 0.5°C increments between 56°C and 95°C. The fold change expression of CCT genes in RL tissues was calculated by $\Delta\Delta CT$ method, considering the expression values of CCT genes in VL tissues as reference.

Results And Discussion

3.1. Genome-wide identification and chromosomal localization of CCT family genes of pigeonpea

We identified a total of 33 CCT family genes in the genome of *C. cajan* (Table 1). These genes were designated as *CcCCT1* to *CcCCT33* in ascending order based on their chromosomal positions. All the 33 CcCCT family genes were dispersed on 10 *C. cajan* chromosomes and 9 genomic scaffolds (Fig. 1). Chromosomes 3, 11, and 7 contained a relatively higher number of CcCCT family genes, with 6, 4, and 3 genes. Each of the chromosomes 2, 4, 6, and 9 contained 2 CcCCT family genes, while chromosomes 1, 8, and 10 contained a single CcCCT family gene. Nine CcCCT (*CcCCT25-CcCCT33*) genes could not

be placed on any of the chromosomes or anchored scaffolds. The genes *CcCCT7* and *CcCCT8*, and *CcCCT15* and *CcCCT16* were located in close vicinity, as pairs of repeat genes, on chromosomes 3 and 7. The BLAST analysis using the conserved centromeric tandem repeats as query against the *C. cajan* chromosome sequences revealed the tentative physical positions of centromeres in all the 11 chromosomes of *C. cajan* (Table 2). The chromosomal localization study of the CcCCT family genes indicated that many of these genes were localized in the chromosomes' centromeric regions, explaining the substantial structural and functional conservation of CCT family genes [30].

3.2. Domain prediction and structural analysis of CcCCT family genes of pigeonpea

Domain prediction of CcCCT family genes using 'web CD-search tool' revealed that among the 33 CcCCT family genes, protein products of 12 genes contained only the CCT domain, while that of the remaining 21 genes contained either of the B-box, PRR, or ZnF-GATA/tifi domains along with the CCT domain. Based on the type of domain, we classified all the 33 CcCCT family genes into four classes comprising CMF type, COL type, PRR type, and GTCC type. The twelve CCT family genes containing only CCT domains were classified as CMF type, while another 12 CCT family genes containing 1–3 B-box and a single CCT domain were considered COL types. The remaining two classes comprised five PRR types containing REC (receiver) and CCT domains, and four GTCC type containing GATA and/or tifi domain along with CCT (Fig. 2). The study indicated that *C. cajan* CCT family is composed of a broad group of genes. The genes' protein products also displayed a vast diversity within the gene family members (Fig. 2). Among the 12 COL type CcCCT proteins, two proteins contained three B-box domains, four proteins containing two B-box domains, and six proteins containing a single B-box domain along with a CCT domain. An integrated comparative and phylogenetic analysis of the four sub-families of CCT genes in *Poaceae* has revealed that the non-B-box domain-containing CMF types have evolved from the B-box domain-containing COL types through the successive loss of B-box domains. Evidence suggests that COL types continue to lose their B-box domains from three to two and two to one and then to none [31].

The structural analysis revealed a considerable variation in the length and structure of the CcCCT family genes (Fig. 2). Among the 33 CcCCT family genes, *CcCCT33* was the smallest (44kb), while *CcCCT11* was the largest (116kb). The COL type CcCCT family genes varied in length from 3–5 kb and contained either one or three introns. Similarly, the CMF type CcCCT family genes varied from 3–6 kb with 1–4 introns, while GTCC type CcCCT family genes ranged from 5–7 kb with 5–9 introns. The PRR type CcCCT family genes were the longest, ranging from 6–16 kb, and contained 7–8 introns (Fig. 2). The structure of the CCT family genes observed in the study showed significant similarity with the structure of the CCT family genes of *A. tauschii* [30]. The molecular weight of protein products of CcCCT family genes ranged from 19.4 kD (*CcCCT28*) to 84.8 kD (*CcCCT11*) with the isoelectric points (pI) ranging from 4.37 (*CcCCT18*) to 9.77 (*CcCCT16*) (Table 1). The range of the molecular weight of the protein products of CcCCT family genes and their pIs were comparable to those observed in *A. tauschii* [30].

3.3. Orthology, intergeneric/interspecific sharing, and phylogenetic analysis of CCT family genes

Besides 33 CCT family genes identified in *C. cajan*, the genome-wide survey of CCT family genes in other legumes led to the identification of an additional 258 CCT family genes, including 36 in *V. radiata*, 37 each in *V. angularis* and *C. arietinum*, 38 in *M. truncatula*, 40 in *V. unguiculata*, and the maximum 70 in *G. max*. We may attribute the presence of a significantly higher number of CCT family genes in *G. max* to genome duplication events during its evolution from common ancestors [32]. The analysis of CCT family genes for the prediction of orthogroups revealed all three kinds of relationships, namely one-to-one, many-to-one, and many-to-many types, between the orthologous genes of the different legume species [33]. All of the 33 CCT family genes identified in *C. cajan* grouped into nine paralogous groups and nine singletons (Table 3). Among the nine paralogous groups, five groups had only two paralogues, while the remaining four groups comprised 3–4 paralogues. All the nine paralogous groups and the singletons had the corresponding orthologs in all the other six common legumes. However, only five paralogous groups and six singletons had their corresponding orthologs in *A. thaliana*. The results indicated that orthologous genes across the different legume species and *A. thaliana* had a variable number of paralogs in other species. These results support the earlier reports indicating that the proportion of duplicated genes retained in a genome is highly variable, and as a result, the number of paralogous genes in a paralogous group varies significantly among different species [34]. Non-adapted species tend to possess a higher number of duplicated genes which provides an opportunity for neofunctionalization and thus adaptive evolution of the species [35–37]. A significant variation in the number of members in different paralogous groups may also be attributed to the fact that gene duplication variation may allow subfunctionalization if the duplication occurs in a gene performing more than one functions. In such a case, the duplicated gene pair might adapt to serve the individual function separately than the parental gene [38].

The Circos map revealed an extensive sharing of CCT family genes among the legumes. It indicated that *C. cajan* shared the highest proportion of CCT family genes with *G. max*, followed by *V. unguiculata*, *V. radiata*, *V. angularis*, *C. arietinum* and *M. truncatula* (Fig. 3). These results are, in general, congruent with the taxonomic relationships of these legume species [39].

The phylogenetic analysis based on the alignment of amino acid sequences of the 291 legume-based and 26 well-characterized *A. thaliana* CCT family genes showed that the CCT family genes could be grouped in seven different clusters designated as clusters A–G (Fig. 4). Cluster-wise distribution of the protein products of the CCT family genes exhibited all GTCC types' affiliation to cluster A. The CMF type CCT family genes were distributed in clusters B and C, while COL type CCT family genes were distributed in clusters D, E, and F. All the PRR type CCT family genes showed their affiliation to cluster G. The *C. cajan* CCT family genes were dispersed in all the clusters of the phylogenetic tree. Cluster B contained the highest number of CcCCT family genes (9). Clusters C and D each contained 3 CcCCT family genes; clusters A and E contained 4 CcCCT family genes, and clusters F and G contained 5 CcCCT family genes. *C. cajan* and *G. max* CCT family genes showed maximum similarity and co-clustered in the phylogenetic tree.

As evident in other species, the long-term evolution and anthropogenic-driven selection pressures during domestication of the common legumes may have resulted in a significant inter and intra-species/generic variation in the functional and regulatory aspects of CCT family genes. Nevertheless, the co-clustering of well-characterized CCT family genes from *Arabidopsis* can reference the functional research of the CCT family genes of *C. cajan* and other legumes used in the study.

3.4 Expression analysis of *CcCCT* family genes in photoperiod sensitive and insensitive pigeonpea genotypes

In the photoperiod-sensitive plant species, COL, PRRs, and CMF type of CCT family genes play a crucial role in photoperiod dependent flowering by either modulating the expression of *FT* or activation of photoperiod independent pathways of flowering [19, 21, 26]. The CO protein controls flowering induction events in plants through interaction with environmental stimuli and plants' circadian rhythm proteins. Under favourable conditions, it activates the expression of *FT* genes to induce flowering. Transcription and translation of *FT* occur in the plant's reproductive leaves, and translated protein products are translocated to the plant's shoot apex through the phloem [40]. Hence, reproductive leaves (RL) are ideal tissues for studying expression profiles and identifying probable candidate gene(s) for photoperiod sensitivity. We analyzed the expression profile of all the 33 *CcCCT* family genes using transcriptome data obtained from RL tissues of ICP20338 and MAL3 grown under SD and LD conditions. The results indicated a significantly higher expression of only three *CcCCT* family genes, namely *CcCCT4*, *CcCCT23*, and *CcCCT32* in RL tissues (Fig. 5). We further analyzed the expression profile of these three genes using real-time PCR. The expression analysis of *CcCCT4* indicated its higher expression in RL tissues compared to the VL tissues of ICP20338 under both SD and LD conditions. However it did not show its expression in MAL 3.

Compared to the VL tissues, *CcCCT23* also exhibited higher expression in the RL tissues of both SD and LD-grown ICP20338 and SD-grown MAL 3. Its expression in LD-grown MAL3 was, however, non-detectable. The expression of *CcCCT32* in VL and RL tissues of ICP20338 and MAL 3 under SD and LD conditions did not show significant variation (Fig. 6). ICP20338 is a photoperiod insensitive genotype and flowers under both SD and LD conditions. On the other hand, MAL3 is photoperiod sensitive and flowers only under SD condition. The expression of *CcCCT23* in the SD-grown and not in the LD-grown MAL 3 genotype strongly indicates that *CcCCT23* is the candidate gene for the synthesis of CONSTANS protein in pigeonpea.

Moreover, the expression of *CcCCT4* only in ICP20338 and not in MAL3 indicates the possibility of two active CONSTANS genes, namely *CcCCT4* and *CcCCT23* in ICP20338, making it photoperiod insensitive. On the other hand, only *CcCCT23* as a functional CONSTANS gene makes MAL3 photoperiod sensitive.

Declarations

Funding: Not applicable

Conflicts of interest/Competing interests: The author(s) declare that there is no competing interest.

Availability of data and material: Not applicable

Code availability: Not applicable

Authors' contributions:

KG, NKS, and KUT, conceived and designed the study. KG designed the experimental plan. KD, RJ and KK participated in sample maintenance and collection. TK, AD and KUT conducted all wet-lab experiments. KUT and HS performed the data analysis; KUT, BKS, KK, KD, NKS and KG compiled and interpreted the data and wrote the manuscript. All authors have read and approved the compiled manuscript.

Ethics approval: Not applicable

Consent to participate: Not applicable

Consent for publication: Not applicable

Acknowledgement

We sincerely acknowledge the Director, ICAR-NIPB, New Delhi; Director, ICAR-IARI, New Delhi; Director, ICAR-IIAB, Ranchi; and the Incharge, National Phytotron Facility, ICAR- IARI, New Delhi for providing the necessary facilities to carry out the research work.

References

- [1] Lang A.(1965) Physiology of flower initiation. In: Lang A. (eds) Differenzierung und Entwicklung / Differentiation and Development. Handbuch der Pflanzenphysiologie / Encyclopedia of Plant Physiology; 15. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-50088-6_39
- [2] Trevaskis B, Hemming MN, Dennis ES, et al. (2007) The molecular basis of vernalization induced flowering in cereals. Trends Plant Sci, 12: 352–357. <https://doi.org/10.1016/j.tplants.2007.06.010>
- [3] Agliassa C, Narayana R, Berteau CM, et al. (2018) Reduction of the geomagnetic field delays Arabidopsis thaliana flowering time through downregulation of flowering-related genes. Bioelectromagnetics, 39: 361-374. <https://doi.org/10.1002/bem.22123>
- [4] Song YH, Shim JS, Kinmonth-Schultz HA, et al. (2015) Photoperiodic flowering: time measurement mechanisms in leaves. Korea Annu. Rev. Plant Biol, 66: 441–464. <https://doi.org/10.1146/annurev-arplant-043014-115555>
- [5] Putterill JJ, Robson F, Lee K, et al. (1995) The *CONSTANS* gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell, 80:847-857. [https://doi.org/10.1016/0092-8674\(95\)90288-0](https://doi.org/10.1016/0092-8674(95)90288-0)

- [6] Corbesier L, Vincent C, Jang S, et al. (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science*, 316: 1030–1033. <https://doi.org/10.1126/science.1141752>
- [7] Eliezer L, Brian A, Yuval E. (2014) Florigen and anti-florigen – a systemic mechanism for coordinating growth and termination in flowering plants. *Frontiers in Plant Science*, 5: 465. <https://doi.org/10.3389/fpls.2014.00465>
- [8] Tamaki S, Matsuo S, Wong HL, et al. (2007) *Hd3a* Protein is a mobile flowering signal in rice. *Science*, 316: 1033–1036. <https://doi.org/10.1126/science.1141753>
- [9] Cheng X, Li G, Tang Y, et al. (2018) Dissection of genetic regulation of compound inflorescence development in *Medicago truncatula*. *Development*, 145(3):dev158766. <https://doi.org/10.1242/dev.158766>
- [10] Taoka K, Ohki I, Tsuji H, et al. (2011) 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature*, 476:332–335. <https://doi.org/10.1038/nature10272>
- [11] Golembeski GS, Imaizumi T. (2015) Photoperiodic regulation of florigen function in *Arabidopsis thaliana*. 13: e0178. <https://doi.org/10.1199/tab.0178>
- [12] David KM, Armbruster U, Tama N, et al. (2006) *Arabidopsis* GIGANTEA protein is post-transcriptionally regulated by light and dark. *FEBS Lett*, 580: 1193–1197. <https://doi.org/10.1016/j.febslet.2006.01.016>
- [13] Niwa Y, Ito S, Nakamichi N, et al. (2007) Genetic linkages of the circadian clock associated genes *TOC1 CCA1* and *LHY* in the photoperiodic control of flowering time in *Arabidopsis thaliana*. *Plant Cell Physiol*, 48: 925–937. <https://doi.org/10.1093/pcp/pcm067>
- [14] Sawa M, Nusinow DA, Kay SA, et al. (2007) *FKF1* and *GIGANTEA* complex formation is required for day-length measurement in *Arabidopsis*. *Science*, 318: 261–265. <https://doi.org/10.1126/science.1146994>
- [15] Fornara F, Panigrahi KCS, Gissot L, et al. (2009) *Arabidopsis* DOF transcription factors act redundantly to reduce *CONSTANS* expression and are essential for a photoperiodic flowering response. *Dev Cell*, 17: 75–86. <https://doi.org/10.1016/j.devcel.2009.06.015>
- [16] Nakamichi N, Kiba T, Henriques R, et al. (2010) *PSEUDO-RESPONSE REGULATORS 97* and *5* are transcriptional repressors in the *Arabidopsis* circadian clock. *Plant Cell*, 22: 594–605. <https://doi.org/10.1105/tpc.109.072892>
- [17] An H, Roussot C, Suarez LP, et al. (2004) *CONSTANS* acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. *Development*, 131:3615–3626. <https://doi.org/10.1242/dev.01231>
- [18] Khanna R, Kronmiller B, Maszle DR, et al. (2009) The *Arabidopsis* B-Box Zinc finger family. *Plant Cell*, 21: 3416–3420. <https://doi.org/10.1105/tpc.109.069088>
- [19] Li Y, Xu M. (2017) CCT family genes in cereal crops: A current overview. *Crop J*, 5: 449–458. <https://doi.org/10.1016/j.cj.2017.07.001>
- [20] Gao H, Jin M, Zheng XM, et al. (2014) *Days to heading-7* a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proc Natl Acad Sci*, 111: 16337–16342. <https://doi.org/10.1073/pnas.1418204111>
- [21] Beales J, Turner A, Griffiths S, et al. (2007) A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor Appl Genet* 115: 721–733. <https://doi.org/10.1007/s00122-007-0603-4>
- [22] Fuller DQ, Murphy C, Kingwell-Banham E, et al. (2019) *Cajanus cajan* (L.) Millsp. origins and domestication: The South and Southeast Asian archaeobotanical evidence. *Genet Resour Crop Evol*, 66, 1175–1188. <https://doi.org/10.1007/s10722-019-00774-w>
- [23] Srivarsha J, Jahagirdar JE, Kumar CVS, et al. (2018) Performance of Parents and Hybrids of Pigeonpea (*Cajanus cajan* (L.) Millsp.) in terms of Yield and Yield Contributing Characters. *Int. J. Pure App. Biosci*, 6(1): 1271-1275. DOI: <http://dx.doi.org/10.18782/2320-7051.5803>
- [24] Velez-Colon R, Garrison SA. (1998) Growth maturity and flowering of pigeon peas (*Cajanus cajan* L Millsp.) at high latitudes. *J Agric Univ*, 73: 223–229. <https://doi.org/10.46429/jaupr.v73i3.6460>
- [25] Tribhuvan KU, Das A, Srivastava H, et al.(2020) Identification and characterization of PEBP family genes reveal CcFT8 a probable candidate for photoperiod insensitivity in *C. cajan*. *3 Biotech*, 10, 194. <https://doi.org/10.1007/s13205-020-02180-x>
- [26] Zhang L, Li Q, Dong H, et al. (2015) Three CCT domain-containing genes were identified to regulate heading date by candidate gene-based association mapping and transformation in rice. *Sci. Rep*, 5: 7663. <https://doi.org/10.1038/srep07663>
- [27] Voorrips RE. (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered*, 93: 77–78. <https://doi.org/10.1093/jhered/93.1.77>
- [28] Melters DP, Bradnam KR, Young HA et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*, 14, R10. <https://doi.org/10.1186/gb-2013-14-1-r10>

- [29] Kumar S, Stecher G, Li M, et al. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, 35:1547-154. <https://doi.org/10.1093/molbev/msy096>
- [30] Zheng X, Li X, Ge C, et al. (2017) Characterization of the CCT family and analysis of gene expression in *Aegilops tauschii*. *PLoS ONE*, 12(12): e0189333. <https://doi.org/10.1371/journal.pone.0189333>
- [31] Cockram J, Thiel T, Steuernagel B, et al. (2012) Genome Dynamics Explain the Evolution of Flowering Time CCT Domain Gene Families in the Poaceae. *PLoS ONE*, 7(9): e45307. <https://doi.org/10.1371/journal.pone.0045307>
- [32] Shoemaker R, Olson T, Kanazin V. (1996) Soybean genome organization: evolution of a legume genome. In: Gustafson JP, Flavell RB (eds) *Genomes of plants and animals*. Stadler Genetics Symposia Series. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-0280-1_11
- [33] Zahn-Zabal M, Dessimoz C, Glover NM. (2020) Identifying orthologs with OMA: A primer. *F1000Research*; 9: 27. <https://doi.org/10.12688/f1000research.21508.1>
- [34] Wolfe KH. (2001) Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet*, 2:333–341. <https://doi.org/10.1038/35072009>
- [35] Nowak MA, Boerlijst MC, Cooke J, et al. (1997) Evolution of genetic redundancy. *Nature*, 338:167–171. <https://doi.org/10.1038/40618>
- [36] Taverna, D.M. and Goldstein, R.M. (2000) The evolution of duplicated genes considering protein stability constraints. *Pac. Symp. Biocomput*, 69–80.
- [37] Wagner A.(2000) The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics*, 154:1389–1401.
- [38] Force A, Lynch M, Pickett FB, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151:1531–1545.
- [39] Lee C, Yu D, Choi HK, et al. (2017) Reconstruction of a composite comparative map composed of ten legume genomes. *Genes Genomics*, 39(1):111-119. <https://doi.org/10.1007/s13258-016-0481-8>
- [40] Tsuji H, Taoka K. (2014) Chapter five-Florigen Signaling. In: Machida Y, Lin C, Tamanoi F (eds). *The Enzymes*. Academic Press, 35: 113-144. <https://doi.org/10.1016/B978-0-12-801922-1.00005-1>.

Tables

Table 1
Characteristics of *C. cajan* CCT family genes.

Sr. No	Gene ID	Name of the gene	Location	Protein length (aa)	Molecular weight (D)	Theoretical pI
1.	C. cajan_20131	CcCCT1	Chr1	365	39953.05	5.55
2.	C. cajan_04552	CcCCT2	Chr2	282	31855.17	5.02
3.	C. cajan_05381	CcCCT3	Chr2	241	28600.85	4.57
4.	C. cajan_08639	CcCCT4	Chr3	393	44226.40	5.28
5.	C. cajan_08715	CcCCT5	Chr3	528	59236.29	5.86
6.	C. cajan_08842	CcCCT6	Chr3	238	26983.79	5.14
7.	C. cajan_10458	CcCCT7	Chr3	345	38588.05	5.75
8.	C. cajan_10578	CcCCT8	Chr3	346	38774.43	5.97
9.	C. cajan_10962	CcCCT9	Chr3	586	64938.20	6.54
10.	C. cajan_21434	CcCCT10	Chr4	255	29460.77	9.32
11.	C. cajan_21442	CcCCT11	Chr4	777	84877.17	8.06
12.	C. cajan_11568	CcCCT12	Chr6	314	35353.80	5.41
13.	C. cajan_13216	CcCCT13	Chr6	181	21052.56	9.11
14.	C. cajan_17635	CcCCT14	Chr7	327	36734.21	8.01
15.	C. cajan_17916	CcCCT15	Chr7	255	28049.55	7.64
16.	C. cajan_17917	CcCCT16	Chr7	180	20058.86	9.77
17.	C. cajan_15626	CcCCT17	Chr8	368	41626.56	5.46
18.	C. cajan_22520	CcCCT18	Chr9	412	45428.06	4.37
19.	C. cajan_22947	CcCCT19	Chr9	336	37395.46	5.31
20.	C. cajan_13596	CcCCT20	Chr10	175	20159.33	8.46
21.	C. cajan_00921	CcCCT21	Chr11	366	40639.58	8.30
22.	C. cajan_02757	CcCCT22	Chr11	388	42180.59	4.62
23.	C. cajan_03622	CcCCT23	Chr11	412	44689.53	7.42
24.	C. cajan_03709	CcCCT24	Chr11	368	41250.69	8.97
25.	C. cajan_32117	CcCCT25	Scaffold000063	304	33458.92	6.19
26.	C. cajan_36557	CcCCT26	Scaffold000230	427	47126.94	5.83
27.	C. cajan_26721	CcCCT27	Scaffold000400	370	41109.70	5.01
28.	C. cajan_24786	CcCCT28	Scaffold000401	165	19464.68	8.63
29.	C. cajan_39146	CcCCT29	Scaffolds132091	625	69387.24	6.21
30.	C. cajan_39868	CcCCT30	Scaffold132067	624	69124.16	6.11
31.	C. cajan_31872	CcCCT31	Scaffold132202	416	46139.56	5.28
32.	C. cajan_32168	CcCCT32	Scaffold133489	308	33597.63	7.01
33.	C. cajan_29262	CcCCT33	Scaffold137859	411	44814.16	5.15

Table 2
Probable centromeric positions of *C. cajan* chromosomes.

Sr. No	Database ID	Chromosome No.	Length of Chromosome	Centromere position	
				Start	Stop
1.	CM003603.1	Chr 1	17676265	7907841	7907859
2.	CM003604.1	Chr 2	36921656	10643125	10643224
3.	CM003605.1	Chr 3	29476785	10196466	10196486
4.	CM003606.1	Chr 4	12502730	4909048	4908978
5.	CM003607.1	Chr 5	5133333	2115182	2115201
6.	CM003608.1	Chr 6	23791330	9297812	9297832
7.	CM003609.1	Chr 7	19579009	6095224	6095201
8.	CM003610.1	Chr 8	20383548	12135398	12135354
9.	CM003611.1	Chr 9	10489139	7762458	7762477
10.	CM003612.1	Chr 10	22571102	932227	932196
11.	CM003613.1	Chr 11	48970052	16968062	16968088

Table 3
Orthologs of CcCCT family genes in different legume species.

Species OrthoID	A. Thaliana	C. arietinum	C. cajan	G. max	M. truncatula	V. angulata	V. unguiculata	V. radiata
OG0000000	COL1, COL2, COL3, COL4, COL5, CONS	Ca05961.1, Ca08365.1, Ca09606.1, Ca21484.1	CCT1, CCT23, CCT27, CCT32	04G058900.1, 06G059600.1, 08G255200.1, 13G050300.1, 13G093800.1, 17G066600.1, 18G278100.2, 19G039000.1	1g013450.1, 3g105710.1, 4g128930.1, 7g018170.1	Vang02g13360.1, Vang04g02640.1, Vang05g09750.2, Vang11g14640.1	Vigun03g379600.1, Vigun04g068400.1, Vigun09g217500.1	Vradi01g07820.1, Vradi04g09860.1, Vradi05g11170.1, Vradi07g07500.1, Vradi07g21950.1
OG0000001	APRR3, APRR5, APRR7, APRR9	Ca07096.1, Ca11231.1, Ca14266.1, Ca18487.1	CCT9, CCT11, CCT29, CCT30	03G261300.3, 04G228300.2, 06G136600.3, 07G049400.10, 10G048100.1, 13G135900.5, 16G018000.1, 19G260400.2	1g067110.7, 3g092780.1, 4g061360.4, 7g118260.1, 8g024260.1	Vang0010ss00800.1, Vang0084s00090.1, Vang0238s00030.1	Vigun07g078600.2, Vigun09g004700.2, Vigun09g140300.2, Vigun10g152100.2, Vigun11g149100.1	Vradi0179s00390.1, Vradi02g05430.1, Vradi05g00340.1
OG0000002	GAT24, GAT25, GAT28	Ca04105.1, Ca14939.1, Ca14940.1, Ca18523.1	CCT15, CCT16, CCT25	02G215600.1, 04G096200.1, 04G096300.1, 06G097900.1, 06G098000.1, 08G221800.10, 14G182800.1	3g100040.1, 3g100050.1, 5g066510.1	Vang0017s00340.2, Vang0017s00360.1, Vang01g08240.1, Vang01g09310.1	Vigun08g109700.3, Vigun08g127200.1, Vigun09g180800.2, Vigun09g180900.2, Vigun09g276400.1	Vradi0153s00520.1, Vradi0180s00180.1, Vradi05g07550.1, Vradi05g07560.1
OG0000003	COL16, COL6, COL7	Ca11846.1, Ca15551.1	CCT4, CCT17, CCT31	05G233700.1, 07G091400.3, 08G041100.1, 09G184600.3, 10G274300.1, 20G115600.1	1g110870.1, 8g104190.1	Vang0070s00170.1, Vang02g04910.1, Vang06g00840.1, Vang10g01650.1	Vigun02g042100.1, Vigun03g017700.1, Vigun04g021700.2, Vigun07g270500.1	Vradi0169s00070.1, Vradi01g04550.1, Vradi07g27680.1, Vradi08g21370.1
OG0000004	APRR1	Ca12864.1, Ca12988.1, Ca21411.1	CCT5	04G166300.2, 05G025000.2, 06G196200.2, 17G102200.3	3g037390.1, 4g108880.1	Vang09g02620.1, Vang11g11310.2	Vigun03g337800.2, Vigun09g076300.1	Vradi03g08890.1
OG0000005	COL10, COL9	Ca03946.1, Ca13653.2	CCT33	02G223700.2, 13G009300.2, 14G190400.1, 20G060400.2	3g082630.2, 5g069480.5	Vang01g09960.2, Vang03g16900.3	Vigun06g006800.2, Vigun08g116900.2	Vradi06g07140.1
OG0000006		Ca00342.1	CCT10, CCT20	12G074400.1, 13G331000.1, 15G219800.1, U033700.1	2g096080.1, 4g061823.1, 4g061910.1	Vang0010ss00750.1, Vang0021ss01680.1	Vigun05g253600.1, Vigun11g150000.1	Vradi02g05350.1, Vradi04g02440.1
OG0000007		Ca08569.1, Ca09217.1, Ca11178.1	CCT18, CCT22	10G134400.1, 10G190300.1, 20G085100.2, 20G200400.2	1g044785.1, 1g073350.1	Vang06g09620.1, Vang06g25290.1	Vigun07g011700.1, Vigun07g181300.1	Vradi08g00870.1
OG0000008	COL13	Ca07767.1	CCT7, CCT8	02G152900.1, 03G209800.1, 10G021400.1, 19G207100.1	7g108150.1	Vang04g15980.3, Vang06g17920.4	Vigun01g187100.3, Vigun07g116500.2	Vradi03g05670.1, Vradi08g09990.1
OG0000009		Ca22540.1	CCT12, CCT13	01G221100.1, 05G209800.1, 08G016500.3, 11G022600.2	5g010120.1, 8g098725.2	Vang07g06400.1, Vang10g06540.1	Vigun02g177100.1, Vigun03g042600.1	Vradi07g30750.1, Vradi11g02090.1
OG0000010		Ca09779.1, Ca14178.1	CCT2, CCT6	04G236400.1, 06G127900.1	1g008220.1, 3g091340.1	Vang0090s00480.2, Vang01g10650.1	Vigun08g080300.2, Vigun09g150200.2	Vradi06g10270.1, Vradi10g01250.1
OG0000011	COL14, COL15	Ca06714.1	CCT14	16G050900.2, 19G099700.2	7g083540.1	Vang04g13110.1	Vigun01g075800.1	Vradi04g04920.1
OG0000012	COL11, COL12	Ca00755.1	CCT26	12G196100.2, 13G306400.2	2g088900.1	Vang09g03140.1	Vigun05g225000.5	Vradi05g22760.1
OG0000013	CIA2	Ca27287.1	CCT24	13G100700.4, 17G059100.2	4g127420.1	Vang11g15310.1	Vigun03g387700.1	Vradi07g22580.1
OG0000014	COL8	Ca03756.1, Ca03757.1	CCT3	02G230800.1, 14G197900.1	5g072780.1	Vang01g08550.1	Vigun08g124000.1	
OG0000015		Ca01591.1	CCT28	08G168900.1, 15G258200.1	2g068730.1	Vang0022ss02670.1	Vigun01g054600.1	Vradi05g18830.1

Species OrthoID	A. Thaliana	C. arietinum	C. cajan	G. max	M. truncatula	V. angulata	V. unguiculata	V. radiata
OG0000016		Ca16555.1	CCT19	07G182700.1, 08G067200.1	4g093730.1	Vang0064ss00450.2	Vigun03g124600.1	Vradi07g13890.1
OG0000017		Ca24426.1	CCT21	16G067000.1, 19G079600.2	7g032240.1	Vang04g12460.1	Vigun01g065500.1	Vradi0259s00090.1
OG0000018		Ca10375.2			1g102900.2		Vigun07g235400.4	

Figures

Figure 1

Chromosome-wise distribution of *C. cajan* CCT family genes. The orange colour lines indicate the paralogues relationship between the genes. CEN represents the centromeric positions of the chromosomes.

Figure 2

Phylogenetic relationship, gene structure and protein domain prediction of *C. cajan* CCT family genes.

Figure 3
Sharing of CCT family genes among the legume species.

Figure 4

Maximum likelihood phylogenetic tree of CCT family genes of *C. cajan* and other legume species. The yellow colour highlights the distribution of *C. cajan* CCT family genes among the various clades of the phylogenetic tree.

Figure 5

Expression analysis of CCT family genes using transcriptome data obtained from the RL tissues of SD and LD-grown ICP20338 and MAL3 genotypes of *C. cajan*.

Figure 6

Real-time PCR-based expression analysis of CcCCT4, CcCCT23 and CcCCT32 using RNA isolated from VL and RL tissues of SD and LD grown ICP20338 and MAL3 genotypes of *C. cajan*. X axis accounts for the log2fold expression of CCT gene in RL tissue. Y axis depicts the pigeonpea genotypes grown under different photoperiod conditions.

Supplementary Files

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

- [SupplementaryS1..docx](#)