



Genome-wide identification and characterization of CONSTANS-like gene family in peas (*Pisum sativum*)

Amna BiBi

PMAS Arid Agriculture University Rawalpindi. Email: amna181bib@gmail.com

Sharjeel Ahmad

Bahauddin Zakariya University Multan. Email: Sharjeelkhankazmi@gmail.com

Ezzah Rao

Bahauddin Zakariya University Multan. Email: Raoezzah231@gmail.com

Shakir shehzad

Zhengzhou University, School of Agriculture Science, Henan, China

Email: ssqureshi.14@gmail.com

Muhammad Irfan (Corresponding Author)

Quaid I Azam University Islamabad. Email: irfanchandia50@gmail.com

Abstract

CONSTANS-like (CO-like, COL) genes play an important role in the circadian clock rhythms, which ensures regular development through complicated biological clocks. COL genes are also involved in plant responses to stress. COL genes are zinc-finger transcription factor. COL genes are also involved in regulating growth and development, and stress responses. COL genes, primarily active in the leaves, play a key role in the photoperiod pathway that regulates flowering. These genes exhibit varying expression levels across different plant varieties, and their amino acid sequences differ, potentially explaining their diverse functions. In this study, 40 PsCOLs genes with complete conserved domain was identified from the genome of *Pisum sativum* and analysed by bioinformatics. 40 PsCOLs were distributed on 7 chromosomes, encoding 40 PsCOL proteins with different physical and chemical properties. We analyzed their gene structure, phylogenetic relationships, synteny, and expression levels in different tissues. The bioinformatical analysis identified 40 COL transcription factors in the pea genome. The phylogenetic tree constructed with *Arabidopsis thaliana* and other pulses indicates that PsCOLs of different clusters have different biological functions. The conserved motif prediction showed that the number and distribution of motifs on each PsCOLs is varied. The results provide a foundation for the study of COL transcription factors of *Pisum sativum* and provide more reference information of the function of COL genes in flowering.

Key words: *Pisum sativum*, COL (CONSTANS LIKE), circadian clock, flower induction



Introduction

The shift from vegetative growth to bolting and flowering is vital for a plant's reproductive success. The transition to flowering in plants is controlled by complex genetic networks influenced by various plant hormones and environmental factors like light, temperature, and day length. In *Arabidopsis thaliana*, about 180 genes are involved in regulating flowering time through six main pathways: vernalization, autonomous, photoperiod, gibberellin (GA), ambient temperature, and age pathways. These genes work together to ensure flowering happens at the right time. For example, the gene FLOWERING LOCUS C (FLC) acts as a key repressor, integrating signals from both the autonomous and vernalization pathways. Other important genes, such as FLOWERING LOCUS T (FT), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), and LEAFY (LFY), act as convergence points, coordinating signals from multiple pathways to initiate flowering.

CONSTANS-like (COL) genes are essential in controlling the flowering of plants through the photoperiod pathway. They work by integrating signals from the circadian clock, light, and specific flowering time genes. As transcription activators specific to the phloem, COL genes promote flowering by boosting the transcription of the FT and TSF genes. However, COL's activity is finely tuned; it is degraded in the dark by the ubiquitin ligase COP1 and in the morning by a pathway activated by the photoreceptor PHYB. This regulation ensures that CONSTANS-like (COL) genes activate the transcription of FT and TSF genes only during long days. This precise timing allows plants to flower at the most appropriate time, aligning their reproductive cycle with favorable environmental conditions.

COL genes are part of an *Arabidopsis* gene family containing 16 other genes encoding transcription factors. These genes generally have two key conserved domains: a C-terminal CCT domain specific to plants, and an N-terminal zinc finger B-box domain, which is also found in animals. CONSTANS-like (COL) genes are part of the larger B-BOX (BBX) transcription factor family. This family is categorized into five groups based on the presence of one or two B-BOX motifs and the presence or absence of the CCT domain. Previous studies have shown that most COL genes with a CCT domain play a role in controlling flowering time in some plant species. By understanding these mechanisms, we can gain valuable insights into how plants adapt to their environments and improve crop breeding strategies.

The availability of the *Arabidopsis* genome sequence and its annotation has opened new possibilities for comparing COL transcriptional regulators. Additionally, various tools and databases, like the PlantTFDB database, have been developed to identify, cluster, align, and analyze plant transcription factors.

In *Arabidopsis*, CO-like genes are grouped into three main categories:

1. **Group I:** Includes AtCO and AtCOL1 to AtCOL5, which have two B-boxes.
2. **Group II:** Contains AtCOL6 to AtCOL8 and AtCOL16, each with one B-box.
3. **Group III:** Consists of COL9 to COL15, which have one B-box and an additional diverged zinc finger domain.



Peas (*Pisum sativum* L.) are an annual diploid ($2x=2n=14$) species belonging to the Fabaceae family, and they are an economically significant vegetable crop grown worldwide. The main edible parts of peas are their seeds, which are rich in proteins, vitamins, and various medicinal compounds. For successful cultivation, it's crucial to manage the timing of flowering and pod development, as this affects the quality and yield of the pea crop, preventing premature flowering and ensuring optimal production of the desired edible seeds.

The genome sequence of peas (*Pisum sativum* L.) has been a significant focus of research due to its importance as a vegetable crop. The Pea Genome International Consortium launched a program to develop a high-quality reference draft sequence for peas. The reference genome sequence of the pea cultivar 'Caméor' was produced, with a genome assembly length of approximately 3.92 Gb, composed of 24,623 scaffolds. This comprehensive genome sequence provides valuable insights into the genetic basis of various traits and facilitates the improvement of pea crops through genome-informed breeding strategies.

In this study we analyzed their gene structure, phylogenetic relationships, synteny, and expression levels in different tissues. The results provide insights into the genetic networks regulating flowering in peas.

Materials and Methods

Discovering and Cataloging COL Genes in Peas

The genome, genes and corresponding protein sequences of the peas were downloaded from the Pulse Crop Database (<https://www.pulsedb.org/>).

Analyzing the Evolutionary Relationships of COL Genes

Protein sequences from various species were collected for a phylogenetic study in the *Planta* journal. The sequences were aligned using the Clustal X2 program with the Gonnet protein weight matrix. A maximum likelihood phylogenetic tree was then built using the MEGA program (v6.06) with the Jones-Taylor-Thornton (JTT) model. The analysis used 1000 bootstrap replicates and the full CDS sequence for a partial 70% length. Uniform rates and homogeneous lineages were used, with gaps/missing data treated using partial deletion with a site coverage cutoff of 70%. Branch frequencies higher than 50% were shown in the results. The figure was beautified with information from the group using iTOL software (<https://itol.embl.de/>).

Breakdown of Gene Structure and Key Motifs

The gene structure was analyzed using the Gene Structure Display Server tool (<http://gsds.cbi.pku.edu.cn/>, v2.0). MEME software (<http://meme.nbcr.net/meme/>, v4.12.0) was used to search for motifs among the proteins. The motif search window ranged from 10 to 100 base pairs (bp). Only motifs that appeared in at least 3 different protein sequences were kept. These motifs were illustrated in two separate figures, matching the phylogenetic trees. The top 10 motifs with the lowest E-values were highlighted and shown in a specific pattern.



Gene expression analysis

Gene expression analysis was carried out using TBtool in the form of heatmap.

Results and Discussion

Genome-wide identification of COL genes in peas

The bioinformatics analysis discovered 40 CONSTANS-like (COL) transcription factors in the pea genome. All the pea COL genes were found to be well conserved. They were designated as `PsCOL' followed by a serial number and sorted based on the E-value of their CCT domain. (S1 Table).

#	Name	Unique name	Type	Organism	Source	Location
1	Psat1g004200.1	Psat1g004200.1_Ps_Cameor_v1a	mRNA	Pisum sativum	Pisum sativum Cameor genome v1a	chr1LG6: 6179629 .. 6182102
2	Psat2g054040.1	Psat2g054040.1_Ps_Cameor_v1a	mRNA	Pisum sativum	Pisum sativum Cameor genome v1a	chr2LG1: 96225535 .. 96227308
3	Psat2g184400.1	Psat2g184400.1_Ps_Cameor_v1a	mRNA	Pisum sativum	Pisum sativum Cameor genome v1a	chr2LG1: 422791882 .. 422793178
4	Psat3g094880.1	Psat3g094880.1_Ps_Cameor_v1a	mRNA	Pisum sativum	Pisum sativum Cameor genome v1a	chr3LG5: 192414403 .. 192418130
5	Psat3g146040.1	Psat3g146040.1_Ps_Cameor_v1a	mRNA	Pisum sativum	Pisum sativum	chr3LG5: 283230002 ..



					Cameo r genom e v1a	283232920
6	Psat4g049800.1	Psat4g049800.1_Ps_Cameor_v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr4LG4: 83108351 .. 83114427
7	Psat5g013840.1	Psat5g013840.1_Ps_Cameor_v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 23305529 .. 23306747
8	Psat5g047080.1	Psat5g047080.1_Ps_Cameor_v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 85890798 .. 85900143
9	Psat5g071920.1	Psat5g071920.1_Ps_Cameor_v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 131131454 .. 131135493
10	Psat5g077440.1	Psat5g077440.1_Ps_Cameor_v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 139279057 .. 139282202
11	Psat5g105240.1	Psat5g105240.1_Ps_Cameor_v1a	mRN A	Pisum sativum	Pisum sativu m	chr5LG3: 187478164 ..



					Cameo r genom e v1a	187481433
12	Psat5g199360.1	Psat5g199360.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 404947942 .. 404956872
13	Psat5g200320.1	Psat5g200320.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 407506973 .. 407511905
14	Psat5g265480.1	Psat5g265480.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 522397014 .. 522401790
15	Psat5g267200.1	Psat5g267200.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 526039434 .. 526041217
16	Psat5g282000.1	Psat5g282000.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 547667071 .. 547668574
17	Psat6g004440.1	Psat6g004440.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m	chr6LG2: 2980326 .. 2984035



					Cameo r genom e v1a	
18	Psat6g107360.1	Psat6g107360.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr6LG2: 179347411 .. 179352245
19	Psat6g124840.1	Psat6g124840.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr6LG2: 233659887 .. 233675414
20	Psat6g131680.1	Psat6g131680.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr6LG2: 252713969 .. 252717993
21	Psat6g223840.1	Psat6g223840.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr6LG2: 441355586 .. 441358728
22	Psat6g228360.1	Psat6g228360.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr6LG2: 459822830 .. 459824927
23	Psat7g011960.1	Psat7g011960.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Pisum sativu m	chr7LG7: 18571500 .. 18574594



					Cameo r genom e v1a	
24	Psat7g031600.1	Psat7g031600.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr7LG7: 51210245 .. 51212076
25	Psat7g037680.1	Psat7g037680.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr7LG7: 63533061 .. 63536130
26	Psat7g101680.1	Psat7g101680.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr7LG7: 166775673 .. 166780076
27	Psat7g221720.1	Psat7g221720.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr7LG7: 445763618 .. 445769891
28	Psat7g260760.1	Psat7g260760.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr7LG7: 488732676 .. 488734148
29	Psat7g260880.1	Psat7g260880.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr7LG7: 488750731 ..



					Cameo r genom e v1a	488754266
30	Psatos3732g024 0.1	Psatos3732g0240.1_Ps_Came or_v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	scaffold037 32: 88975 .. 90064
31	Psatos1849g008 0.1	Psatos1849g0080.1_Ps_Came or_v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	scaffold018 49: 103114 .. 105287
32	Psat3g193520.1	Psat3g193520.1_Ps_Cameor_v 1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr3LG5: 414219815 .. 414222685
33	Psat5g032240.1	Psat5g032240.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 61613251 .. 61615757
34	Psat6g025680.1	Psat6g025680.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr6LG2: 20093622 .. 20095774
35	Psat3g024360.1	Psat3g024360.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr3LG5: 52122308 .. 52125247



					Cameo r genom e v1a	
36	Psat7g210000.1	Psat7g210000.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Cameo r genom e v1a	chr7LG7: 416598234 .. 416601139
37	Psat5g289520.1	Psat5g289520.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr5LG3: 555813839 .. 555815702
38	Psat2g155840.1	Psat2g155840.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr2LG1: 386656323 .. 386658929
39	Psat2g060680.1	Psat2g060680.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr2LG1: 105544927 .. 105549131
40	Psat6g005440.1	Psat6g005440.1_Ps_Cameor_ v1a	mRN A	Pisum sativum	Pisum sativu m Cameo r genom e v1a	chr6LG2: 3748932 .. 3751167

Table S1

Phylogenetic Tree

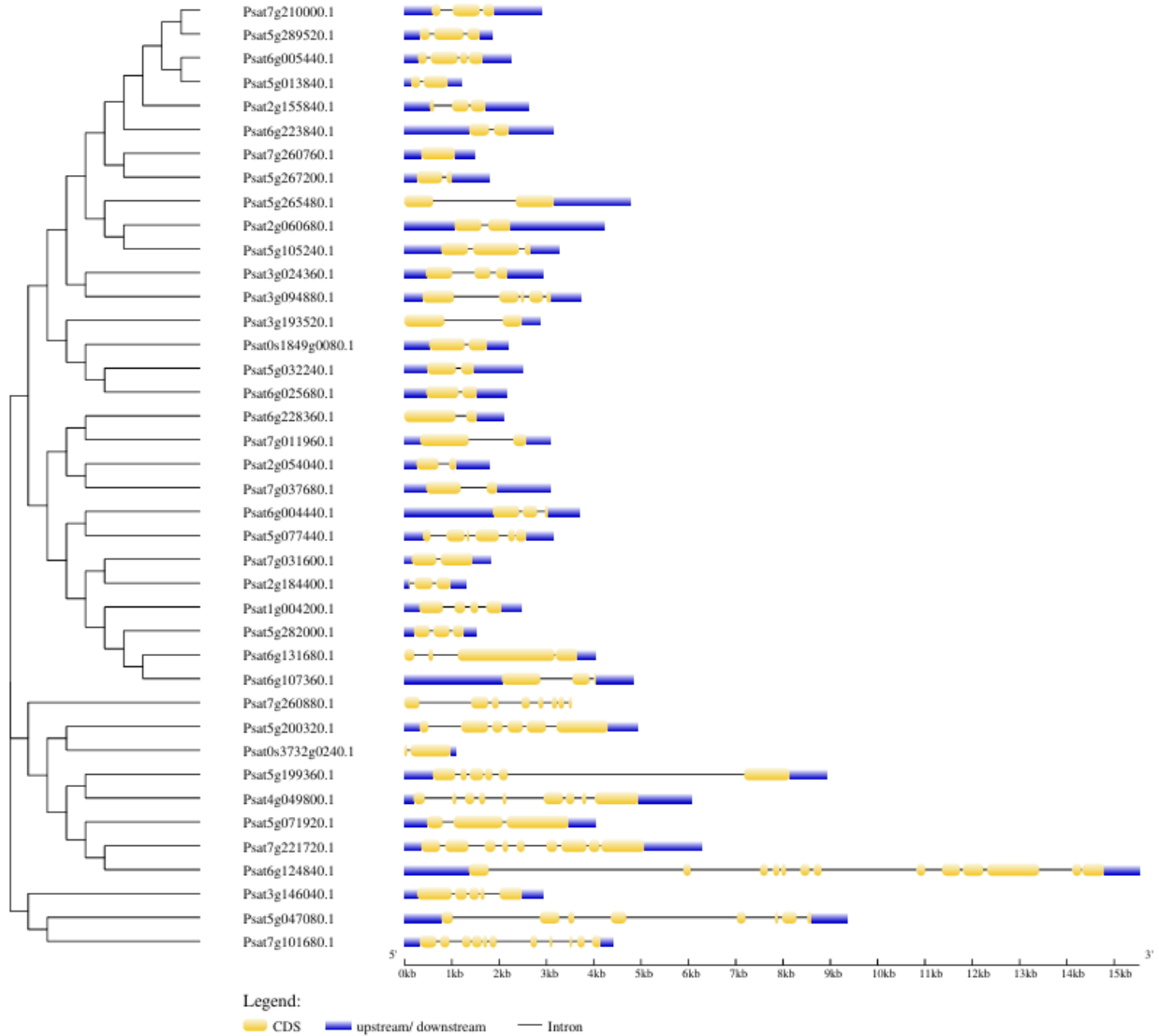


To provide insight into the evolution of COL genes in different sequences of the pea plant, we performed a comparative analysis using the genes from a total of 40 gene sequences of pea genome.



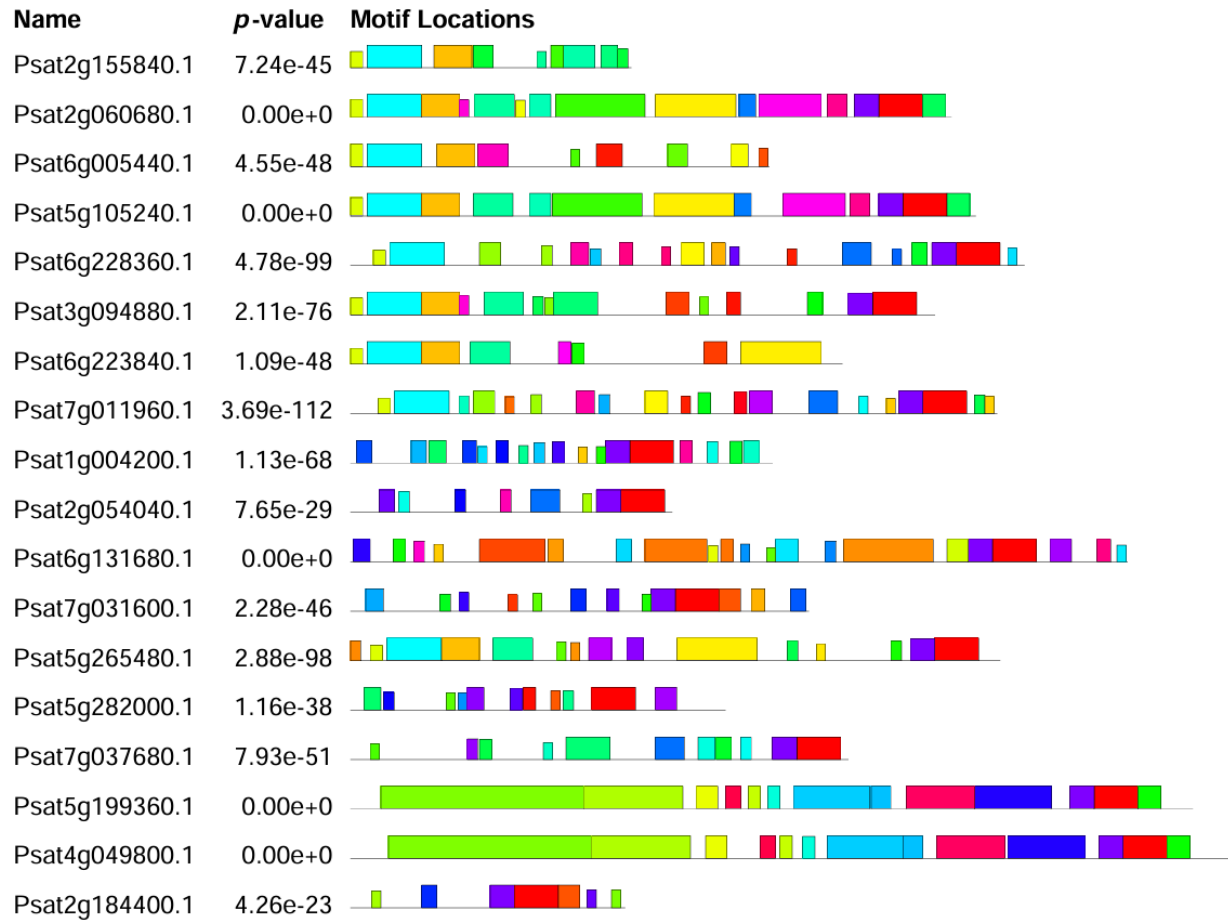
Gene structure

To compare the pea COL genes, their exon-intron structures were predicted, and the results are shown in Fig



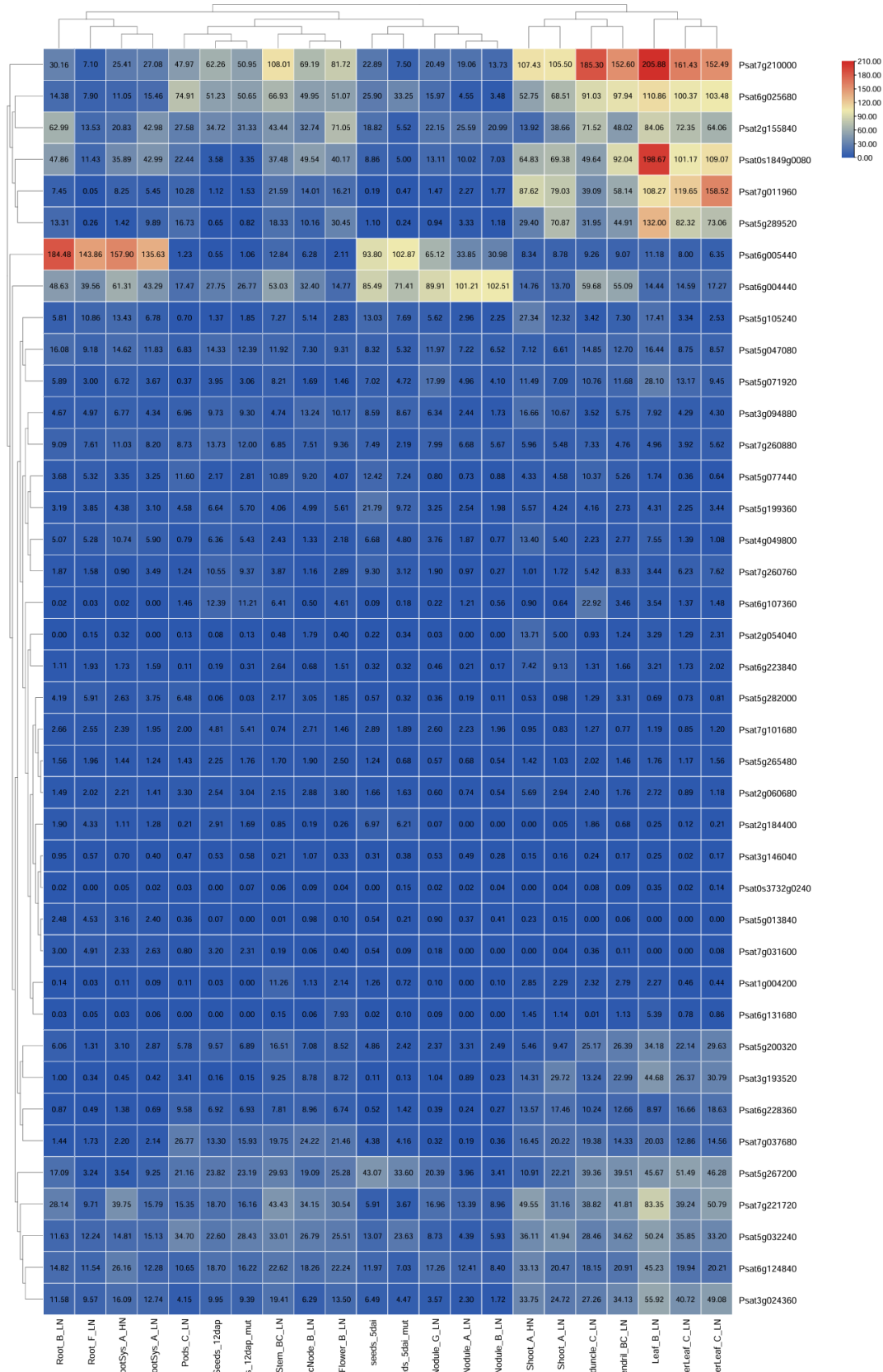
Motif Analysis

To explore the conserved domains and motifs, MEME software was employed to analyze the sequence alignment of the COL proteins in radish. The motifs were listed using serial numbers for Motif 1 to Motif 15 according to the ascending E-value of the alignment.



Expression Analysis

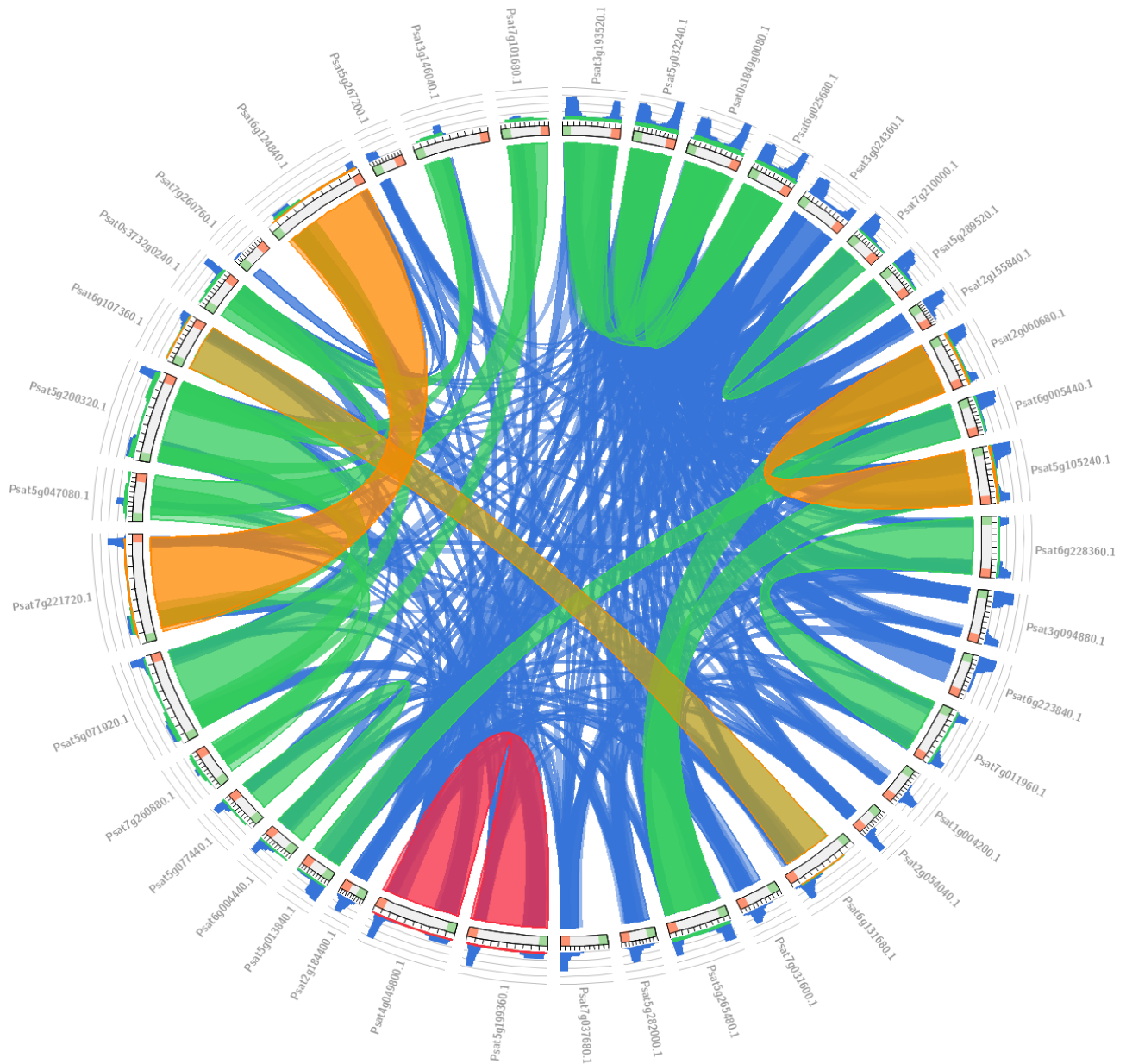
We investigated the expression of each COL gene using published RNA-seq data for different pea tissues during vegetative and reproductive development.





Syntany Analysis

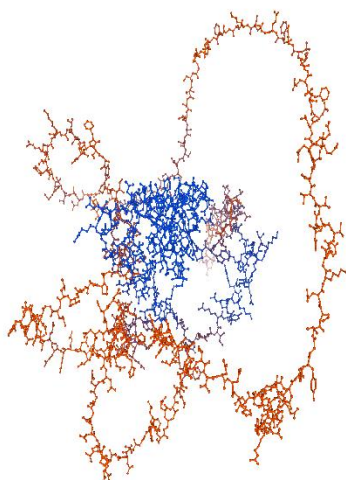
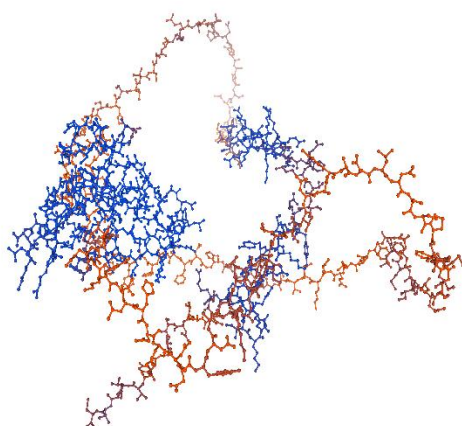
Syntany analysis was performed using Circos Software





3D Protein Structure

3D protein structure was made using Swiss Model Tool



Dialogue Social Science Review (DSSR)

www.thedssr.com

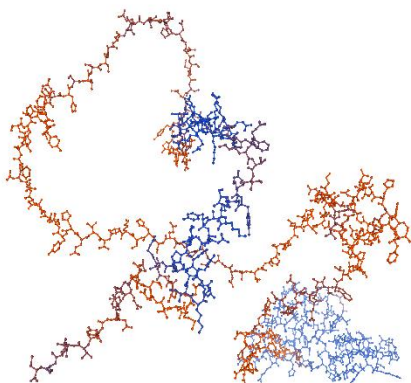
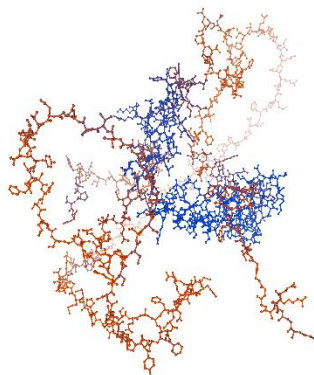
ISSN Online: 3007-3154

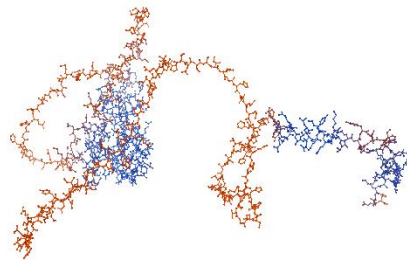
ISSN Print: 3007-3146

Vol. 3 No. 2 (February) (2025)



DIALOGUE SOCIAL SCIENCE REVIEW





References

- AmasinoRM,MichaelsSD(2010)The timing of flowering. *Plant Physiol* 154: 516520. <https://doi.org/10.1104/pp.110.161653> PMID: 20921176
- Nie S, XuL,WangY,HuangD,MulekeEM,SunX,etal.(2015) Identification of bolting-related micro RNAsandtheir targets reveals complex miRNA-mediated flowering-time regulatory networks in radish (*Raphanus sativus* L.). *Sci Rep* 5: 14034.<https://doi.org/10.1038/srep14034> PMID: 26369897
- Srikanth A, Schmid M(2011) Regulation of flowering time: all roads lead to Rome. *Cell Mol Life Sci* 68: 20132037. <https://doi.org/10.1007/s00018-011-0673-y> PMID: 21611891
- Fornara F, de MontaiguA, CouplandG(2010)SnapShot:Control of flowering in *Arabidopsis*. *Cell* 141: 550, 550 e551552. <https://doi.org/10.1016/j.cell.2010.04.024> PMID: 20434991
- Michaels SD, Amasino RM (1999) FLOWERINGLOCUS Cencodesa novel MADS domain proteinthat acts as arepressor of flowering. *Plant Cell* 11: 949956. PMID: 10330478
- LeeJ, Leel(2010) Regulation and function of SOC1, a flowering pathwayintegrator. *J Exp Bot* 61: 22472254. <https://doi.org/10.1093/jxb/erq098> PMID: 20413527
- Berns MC, Nordström K, Cremer F, Tóth R, Hartke M, Simon S, Klasen JR, Bürstel I, Coupland G (2014) Evening expression of *Arabidopsis* GIGANTEA is controlled by combinatorial interactions among evolutionarily conserved regulatory motifs. *Plant Cell* 26:3999–4018. <https://doi.org/10.1105/tpc.114.129437>
- Blazquez MA, Soowal LN, Lee I, Weigel D (1997) LEAFY expression and flower initiation in *Arabidopsis*. *Development* 124:3835–3844. <https://doi.org/10.1242/dev.124.19.3835> Borner R, Kampmann G, Chandler J, Gleissner R, Wisman E, Apel K, Melzer S (2000) A MADS domain gene involved in the transition to



- flowering in Arabidopsis. *Plant J* 24:591–599. <https://doi.org/10.1046/j.1365-313x.2000.00906.x>
- Brandoli C, Petri C, Egea-Cortines M, Weiss J (2020) Gigantea: Uncovering new functions in flower development. *Genes* 11:1142 <https://doi.org/10.3390/genes11101142>
- Choi H-y, Hong J-h, Kang J-y, Kim SY (2000) ABFs, a family of ABA-responsive element binding factors. *J Biol Chem* 275:1723–1730. <https://doi.org/10.1074/jbc.275.3.1723>
- Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J (1999) GIGANTEA: A circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *EMBO J* 18:4679–4688. <https://doi.org/10.1093/emboj/18.17.4679>
- Gray JA, Shalit-Kaneh A, Chu DN, Hsu PY, Harmer SL (2017) The REVEILLE clock genes inhibit growth of juvenile and adult plants by control of cell size. *Plant Physiol* 173:2308–2322. <https://doi.org/10.1104/pp.17.00109>
- Hall A, Bastow RM, Davis SJ, Hanano S, McWatters HG, Hibberd V et al (2003) The TIME FOR COFFEE gene maintains the amplitude and timing of Arabidopsis circadian clocks. *Plant Cell* 15:2719–2729. <https://doi.org/10.1105/tpc.013730>
- Han Y, Zhang X, Wang Y, Ming F (2013) The suppression of WRKY44 by GIGANTEA-miR172 pathway is involved in drought response of Arabidopsis thaliana. *PLoS ONE* 8:e73541. <https://doi.org/10.1371/journal.pone.0124854>
- Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science* 290:2110–2113. <https://doi.org/10.1126/science.290.5499.2110>
- Helfer A, Nusinow DA, Chow BY, Gehrke AR, Bulyk ML, Kay SA (2011) LUX ARRHYTHMO encodes a nighttime repressor of circadian gene expression in the Arabidopsis core clock. *Curr Biol* 21:126–133. <https://doi.org/10.1016/j.cub.2010.12.021>
- Hernandez-Garcia CM, Finer JJ (2014) Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci* 217–218:109–119. <https://doi.org/10.1016/j.plantsci.2013.12.007>
- Jackson SD (2008) Plant responses to photoperiod. *New Phytol* 181:517–531. <https://doi.org/10.1111/j.1469-8137.2008.02681.x>
- Jose J, Bánfalvi Z (2019) The role of GIGANTEA in flowering and abiotic stress adaptation in plants. *Columella* 6:7–18 Kawamura H, Ito S, Yama