PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol 12(1): 1-11
https://doi.org/10.14719/pst.3175

HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

Identification and characterization of genes that regulate flowering in pigeon pea (*Cajanaus cajan*): An *in-silico* exploration

Animesh Pattnaik[1,2], Madhusmita Barik[1], Rukmini Mishra[1] & Jatindra Nath Mohanty[1]*

[1]Department of Botany, School of Applied Sciences, Centurion University of Technology and Management, Jatni, Bhubaneswar 752 050, India

[2]Centre of Excellence in Natural Products and Therapeutics, Department of Biotechnology and Bioinformatics, Sambalpur University, Jyoti Vihar, Burla, Sambalpur 768 019, Odisha, India

*Email: jatindranath.mohanty@cutm.ac.in

OPEN ACCESS

ARTICLE HISTORY
Received: 07 December 2023
Accepted: 30 June 2024
Available online
Version 1.0 : 31 January 2025
Version 2.0 : 05 February 2025

Check for updates

Additional information
Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE
Pattnaik A, Barik M, Mishra R, Mohanty JN. Identification and characterization of genes that regulate flowering in pigeon pea (*Cajanaus cajan*): An *in-silico* exploration. Plant Science Today. 2025; 12(1): 1-11. https://doi.org/10.14719/pst.3175

Abstract

Pigeon pea is a versatile pulse crop extensively cultivated across Latin America, Asia and Africa. It serves as a rich source of protein and fibre. The life cycle of this annual crop is significantly influenced by the timing of flowering, which affects both seed production and the overall growth period. Variation in flowering time is influenced by both biotic and abiotic factors, making it a crucial adaptive trait in flowering plants. In this study, we aim to understand how the genetics of pigeon pea plants regulate their flowering time.

We employed 2 methods, HMM profile search and standalone BLAST search, to identify genes involved in flowering regulation in pigeon pea. Protein sequences of 6 known flowering regulators from *Arabidopsis* and related plants were retrieved from the NCBI database. The entire set of protein sequences from pigeon pea was used as the database for comparison. The top hits with more than 30% identity and known conserved domains were considered true orthologs, resulting in the identification of 6 pigeon pea genes: CcFrigida, CcFrigida Like1, CcFrigida Like2, CcFrigida Essential1, CcTerminal Flowering1 and CcTerminal Flowering2. Through a thorough review, we identified floral repressive genes, such as FLC and its activators, as significant targets for promoting early flowering in plants.

Although considerable progress has been made in understanding the role of MADS-box genes in flower development, we still lack sufficient information about flowering genes and their specific impact on flowering traits in pigeon pea. This investigation will provide details about the biological basis of adaptive traits in this important pulse crop by examining flowering genes in pigeon pea.

Keywords

crop improvement strategies; MADS box genes; pigeon pea; pulse crop

Introduction

The pigeon pea crop is one of the most important pulse crops worldwide. It was domesticated in South Africa approximately 3500 years ago. Today, it is grown all over the world, primarily in Asia, Africa and Latin America (1) covering about 5 million ha of agricultural land. In India, it is predominantly cultivated in the eastern and southern regions. Pigeon pea is closely related to *Cajanaus cajanifolia*, the oldest known species found in the tropical regions of India (2).

The timing of flowering is a pivotal milestone in the life cycle of a plant, profoundly influencing its development. This phenomenon holds exceptional significance for plants that have undergone meticulous domestication, especially annual crops (3). Different species exhibit varied flowering times based on biotic factors (such as competition, pollinators, herbivores) and abiotic factors (such as

photoperiod, temperature, nutrient availability) (3-5). Conditions that promote late blooming lead to a protracted growth phase, which makes it easier to accumulate and distribute resources more effectively for seed production. In contrast, early flowering is linked with a shorter and less consistent growing phase for seed production (6).

The pattern and timing of flowering are important adaptive traits in flowering plants, governed by physiological signals, genes and their interactions (7). Advances in genomic approaches have enhanced our understanding of the molecular mechanisms behind flowering, contributing to improved crop development and production. The availability of genome sequence of a number of plant species has helped to answer fundamental questions in plant biology, including the identification and analysis of genes involved in adaptive traits in crop species (8). One prominent example of evolutionary developmental science in plant species is the identification and analysis of MADS box genes involved in floral development. As a result, identical genes have been discovered in many species, providing information on the diversity and preservation of these genes as well as their roles in plant growth (9).

Floral repressive genes play a central role in promoting early flowering in plants and the involvement of promoter genes further underscores their significance in this process (4). Examples of floral repressive genes include FLC (FLOWERING LOCUS C) and its activators (10, 11). These genes are involved in gibberellin signalling pathways. Variations in the domains of these genes can lead to early flowering (12). Several loss-of-function mutations in FRIGIDA (FRI) result in early flowering. Genetic studies have found that FRIGIDA (FRI) gene as the main promoter of early flowering (12).

Early flowering plants often carry a mutation that causes a loss of function in the FRI allele. However, regardless of whether accessions have functional or non-functional FRI alleles, the variation in blooming time that has been observed appears to be unrelated to FRI. In this context, several accessions with interesting flowering-time characteristics have been identified and described (12). The role of flowering genes in pigeon peas and their specific influence on flowering traits remain insufficiently understood. Information on flowering genes is crucial for any such research. In this study, we identified and characterised the candidate genes and explored possible mechanisms that may regulate flowering traits in *Cajanaus cajan* with the help of computational (*In-silico*) approach.

Materials and Methods

Identification of possible genes involved in flowering trait in pigeon pea

In this research, the TAIR database (https://www.Arabidopsis.org) was used to acquire genetic and molecular biology information connected to the reference plant *Arabidopsis thaliana*. Gene protein sequences associated with flowering traits were extracted from this database. After that, a sequence database was subjected to a protein BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) examination, which resulted in the selection of the top 6 sequences exhibiting the highest degree of similarity.

Further, a local blast (Standalone blast or Offline Blast, ncbi-blast-2.15.0+-x64-win64.tar.gz) was conducted using the query sequences against the genes from the pigeon pea genome (*Cajanaus cajan*). The pigeon pea genome, formatted as a database, was downloaded from the NCBI genome database (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000340665.2/). Sequences with similarity over 30% were chosen. Specifically, sequences with query coverage ranging from 80% to 100% and a p-value threshold of 0.05 were selected for local BLAST analysis. These stringent selection criteria ensured the robustness and statistical significance of the sequence alignments, providing reliable insights into the genetic makeup of the pigeon pea genome. The top-scoring genes from the selected sequences were subsequently retrieved for further analysis.

Annotation of identified genes

The selected top-scoring genes were then used for a gene ontology study using the Blast2go tool. Blast2Go is a universal gene ontology annotation, visualization and analysis tool for functional genomics research (http://blast2go.com/webstart/blast2go1000.jlp). The KEGG Maps and InterPro motifs are also supported by it. Additionally, the application offers a wide array of graphical and analytical tools for annotation manipulation and data mining.

Searching for ESTs of identified flowering related genes

All available pigeon pea (ID: 3821) ESTs (Expressed Sequence Tags) were downloaded from ESTdb, NCBI. A standalone blast was then performed against the complete set of pigeon pea ESTs (as the database), using the identified flowering-related genes as the query. BLAST hits with significant and top identity, with a similarity of over 30% and with query coverage ranging from 80% to 100% and a p-value threshold of 0.05, were considered as ESTs of the genes.

Multiple sequence alignment

Multiple sequence alignments of flowering- related genes from various plants were performed using ClustalW2. ClustalW2 is a useful tool for aligning multiple protein or DNA sequences. It aims to identify the best fit between selected sequences and align them to highlight their distinctions, similarities and identities (http://www.ebi.ac.uk/Tools/msa/clustalw2/help/).

Identification of flowering related genes from pigeon pea using HMM based approach

The results from the ClustalW multiple sequence alignment were used to generate an HMM (Hidden Markov Model) profile. Using this HMM profile, the HMMER-3.0 (https://www.ebi.ac.uk/Tools/hmmer/) software was used to scan sequence databases for homologs of protein sequences. The HMM profile was then used to search the local protein database using HMMER-3.0. HMMER and BLAST hit comparison and parsing was performed.

Gene expression analysis using SRA data and NGS commander software

SRA data for 2 pigeon pea varieties, Asha (**SRX021565**) and Upas120 (**SRX021566**), was downloaded from the SRA database of NCBI (http://www.ncbi.nlm.nih.gov/sra/). Sequence Read Archive stores data for next-generation sequencing platforms.

This SRA data was analysed using NGS commander software version 2.1.4 to find out the gene expression through RNA sequence analysis. NGS commander used for multipurpose areas of NGS like genomics, transcriptomics, epigenomics and as classical sequence analysis tools.

Results

In silico identification of genes involved in flowering trait regulations

To identify genes involved in the regulation of flowering trait, 6 important regulators were selected for this study: Frigida, Frigida -Like1, Frigida-Like2, Frigida-Essential1, Terminal Flowering1 and Terminal Flowering 2. The above-mentioned genes have already been known to play important role in flowering regulation in different plants (14). Protein sequences of above-mentioned genes from *Arabidopsis* and related species were retrieved from GeneBank, NCBI. Two different approaches were used. The first involved building an HMM (Hidden Markov Model) profile and conducting an HMM profile search and the second used a standalone BLAST search with protein sequences of the respective genes from different plants. In both cases, the complete protein sequences from *C. cajan* were used as the database. Results from the HMM profile search and standalone blast were analyzed manually. A self-blast was also performed to remove the redundancy.

Following this approach, top hits with more than 30% identity and with known conserved domains for respective proteins have been considered to be the genuine orthologs of flowering genes in pigeon pea. The identified pigeon pea (*Cajanaus cajan*) genes were named as follows: *CcFrigida (AHL43030), CcFrigida Like1 (P0DKC9), CcFrigida Like2 (Q9C6S2), CcFrigida Essential1 (AEC08893), CcTerminal Flowering1 (Q7G7J6)* and *CcTerminal Flowering2 (AEE79159.1)*.

Table 1 shows the identity of flowering-related genes from pigeon pea to that of *Arabidopsis*. Among the selected genes in pigeon pea, the TERMINAL FLOWERING 2 protein showed the maximum identity (74%) and similarity (88%), whereas CcFrigida essential 1 showed the minimum identity (31%) and similarity (43%) with its respective orthologs in *Arabidopsis*.

In order to see whether the identified proteins had conserved functional annotation, Blast2Go analysis was performed. The results showed that all the 6 proteins shared similar ontology to that of respective *Arabidopsis* genes, with respect to functional role, biological process and cellular functions. The results of the blast2Go analysis are shown in Table 2.

A few important features of the identified pigeon pea flowering genes have also been studied, such as molecular weight, PI and putative sub-cellular localization. Frigida, Frigida-Like 1, Frigida- Essential 1 and Terminal Flowering Locus 2 showed nuclear localization, whereas Frigida-Like 2 showed both nuclear and cytoplasmic localization. Terminal Flowering Locus 1 showed mitochondrial localization (Table 3). The identified proteins showed molecular weight ranging from 61 to 99 kDa and also possessed conserved domains observed in their orthologs in *Arabidopsis*.

The conservation of important domains in the identified pigeon pea flowering genes was also studied by comparing them

Table 1. Identity/similarity of putative flowering genes from pigeon pea with *A. thaliana*

Genes name	Identity	Similarity	E-value
CcFRIGIDA	39%	61%	3e-96
CcFRIGIDA LIKE 1	31%	55%	9e-34
CcFRIGIDA LIKE 2	31%	52%	3e-39
CcFRIGIDA ESSENTIAL 1	31%	43%	4e-24
CcTERMINAL FLOWERING 1	74%	88%	7e-96
CcTERMINAL FLOWERING 2	37%	51%	4e-57

Table 2. Functional annotations of pigeon pea flowering genes as shown by Blast2Go

Gene	Seq. Description	min. e Value	mean Similarity	GOs	Enzyme Codes	InterProScan
FRIGIDA	protein frigida	0	67.9%	P: regulation of multicellular organismal development	NIL	Coil (COILS); Coil (COILS); Coil (COILS); IPR012474 (PFAM); PTHR31791 (PANTHER); PTHR31791:SF3 (PANTHER)
FRIGIDA LIKE-1	frigida-like protein 3	0	88.05%	F: molecular_function; P: biological_process	NIL	Coil (COILS); Coil (COILS); IPR012474 (PFAM); PTHR31791 (PANTHER); PTHR31791:SF4 (PANTHER)
FRIDIGA LIKE-2	frigida-like protein 3	0	88.05%	F: molecular_function; P: biological_process	NIL	Coil (COILS); Coil (COILS); IPR012474 (PFAM); PTHR31791:SF4 (PANTHER); PTHR31791 (PANTHER)
FRIGIDA ESSENTIAL 1	zinc finger c-x8-c-x5-c -x3-h type family isoform 1	0	63.65%	F: binding	NIL	PTHR15242 (PANTHER)
TERMINAL FLOWERING LOCUS 1	dt1(determinant stem)	4.45E-120	96.05%	F: transcription cofactor activity; P: negative regulation of flower development; P: photoperiodism flowering; C: nucleus; C: plasma membrane	NIL	IPR000953 (SMART); IPR007630 (PFAM); IPR011991 (G3DSA:1.10.10. GENE3D); (SUPERFAMILY); IPR016197 (SUPERFAMILY)
TERMINAL FLOWERING LOCUS 2	rna polymerase sigma factor sigb	0	84.5%	P: response to far red light; P: carotenoid biosynthetic process; F: sequence-specific DNA binding transcription factor activity; P: rRNA processing; P: thylakoid membrane organization	EC:2.7.7.6	IPR000943 (PRINTS); IPR008251 (SMART); IPR000953 (SMART); IPR007630 (PFAM); IPR011991 (G3DSA:1.10.10. GENE3D); (SUPERFAMILY); IPR016197 (SUPERFAMILY)

Table 3. Details of molecular weight, PI, sub-cellular localization and conserved domain of flowering related genes from pigeon pea

Gene	PI/Mw (Dalton)	Sub-cellular location (score)	Conserved domains
Frigida	6.19/67897.36	Nuclear (2.273)	Frigida-like protein
Frigida Like 1	6.05/61116.42	Nuclear (2.461)	Frigida-like protein Bin/Amphiphysin/Rvs (BAR) OmOutermrotepinembranepH YqaJ-like viral recombinase
Frigida Like 2	6.05/61116.42	Nuclear (2.352) Cytoplasmic (1.742)	Figida-like protein Bin/Amphiphysin/Rvs (BAR) Outer membrane protein (OmpH-like) YqaJ-like viral recombinase
Frigida Essential 1	7.09/70620.73	Nuclear (4.752)	NIL
Terminal Flowering Locus 1	9.10/19606.52	Mitochondrial (1.305)	PhosphatidylEthanolamineBinding Protein (PEBP)
Terminal Flowering Locus 2	8.052/99023.55	Nuclear (4.282)	Chromatin organization modifier (chromo Sigma70 region (SR) Sigma-70 region Chromo Shadow Domain

with their counterpart in *Arabidopsis thaliana* through sequence alignment using clustalw (Supplementary Fig. 1A-F). CcFrigida essential 1 showed more variations at the N-terminal end and relatively higher conservation at the C-terminal end of the protein (Supplementary Fig. 1A).

Arabidopsis Frigida essential 1 (AtFrigida essential 1)

Amino acid sequence alignment of pigeon pea FRIGIDA with *Arabidopsis* FRIGIDA showed relatively high sequence conservation within the FRIGIDA-like protein domain (Supplementary Fig. 1B). Similarly, CcFIRIGIDA-like 1 showed sequence conservation compared to AtFIRIGIDA-like 1, primarily within the conserved FIRIGIDA-like domain (Supplementary Fig. 1C), which spans most of the length of the protein. Similar sequence conservation within the conserved FIRIGIDA- like domain was also observed in CcFIRIGIDA-like 2. However, an 81 amino acids-long stretch was found inserted in CcFIRIGIDA-like 2 (Supplementary Fig. 1D).

In pigeon pea Terminal flowering 1, an additional 54 amino acid-long stretch was observed at the N-terminal end of the protein (Supplementary Fig. 1E). For Cc Terminal flowering 2, presence of an additional N-terminal extension compared to AtTerminal flowering 2 was observed. However, the C-terminal end showed relatively higher sequence conservation (Supplementary Fig. 1F).

Prediction of SUMOylation of flowering related proteins

SUMOylation is the one of the most important post-translational modifications, playing a pivotal role in several cellular processes, including nuclear-cytosolic transport, transcriptional regulation, biotic and abiotic stress responses and protein stability. Compact proteins known as small ubiquitin-like modifiers (SUMOs) can bind to or separate from a target protein, thereby altering those protein's functions. The identified pigeon pea flowering genes encoding proteins were studied for the possible presence of SUMOylation sites. As shown in Table 4, signature SUMOylation sites were observed in the 6 identified proteins. Two signature SUMOylation sites were found in CcTerminal flowering 1, while 3 sites were identified in CcFrigida and Frigida Essential 1. Four

Table 4. SUMOylation sites in pigeon pea flowering gene protein sequences

Gene	Position	Peptide	Score	Cutoff
CcFrigida	94	SNPNQQVKAEEEEKE	46.88	16
	217		28.021	16
	374	EDVEASLKREAESAA VNCLESHKIDFVKLL	38.054	36.625
CcFrigida Like 1	234	EVSNQDVKKDANLLG	42.742	36.625
	354-358	KMPGVIEVLVNNGRQIDAV	61.262	59.29
	415-419	IEVNERELVALKAVIKCIE	61.369	59.29
	429	IKCIEEHKLDEKYPL	42.466	36.625
CcFrigida Like 2	234	EVSNQDVKKDANLLG	42.742	36.625
	354-358	KMPGVIEVLVNNGRQIDAV	61.262	59.29
	415-419	IEVNERELVALKAVIKCIE	61.369	59.29
	429	IKCIEEHKLDEKYPL	42.466	36.625
CcFrigida Essential 1	29	SFEMLTLKKEELHLK	26.033	16
	36	KKEELHLKSEISCNL	24.476	16
	189	AHQKRELKVEEGVRE	27.478	16
CcTerminal Flowering Locus 1	53	STVNTIPKVEIDGGD	21.518	16
	86-90	DPYLREHLHWIVTDIPGTT	62.164	59.29
CcTerminal Flowering Locus 2	279	AGIQDLLKLEKLQED	26.5	16
	410	VEATYRVKEARKQLY	37.749	36.625
	780	SGSVKRFKRETDPCK	18.092	16
	821-825	DAKTACNIVKIIKPIGYSA	64.369	59.29

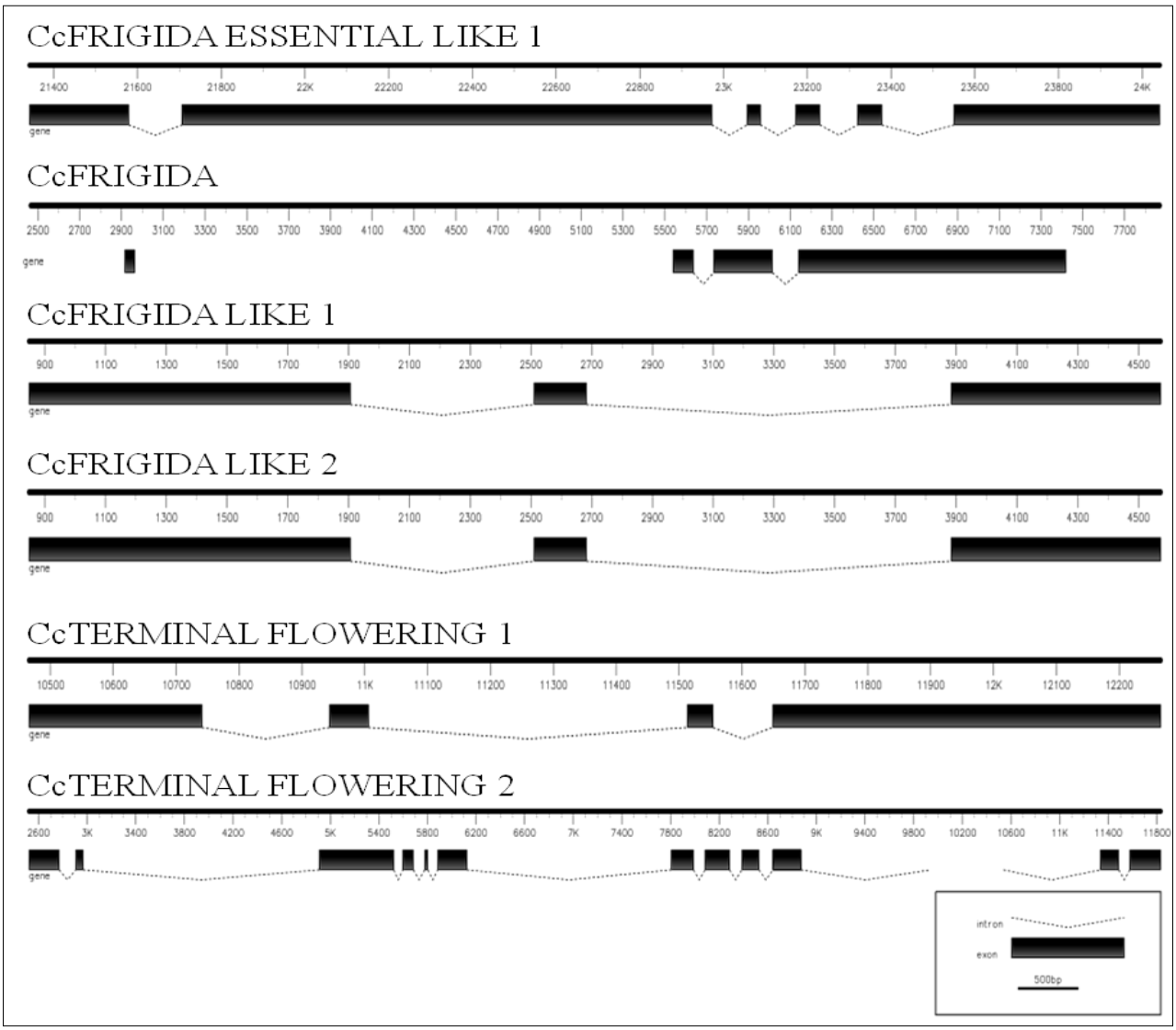

Fig. 1. Genomic structure of pigeon pea flowering genes.

signature SUMOylation sites were observed in CcFrigida-like 1, CcFrigida-like 2 and CcTerminal flowering 2. All the predicted SUMOylation sites have shown significantly high scores, exceeding the cutoff values (Table 4).

Gene structure of pigeon pea flowering related genes

To further understand the flowering-related genes in pigeon pea, their gene structures were studied. For this, results from Augustus gene prediction of whole genome sequence of pigeon pea were considered as input. As shown in Fig. 1, all 6 selected genes showed the presence of introns in their structure. Among the 6 genes studied, CcFrigida Essential 1 showed the maximum number of introns (5 introns), whereas CcFrigida, CcFrigida-like 1, CcFrigida-like 2 had the fewest, with 2 introns in their genomic structures. CcTerminal flowering 1 and CcTerminal flowering 2 showed 3 and 12 introns respectively.

Phylogenetic analysis of pigeon pea flowering related genes

To understand the evolutionary relationship of pigeon pea flowering-related genes, phylogenetic analysis was performed using respective orthologs from other plant species, including *Arabidopsis*, maize (*Zea mays*), brassica (*Brassica rapa*), sesame (*Sesamum indicum*), chickpea (*Cicer arietinum*), soybean (*Glycine max*) and rice (*Oryza sativa*). Amino acid sequences of the respective genes were considered for the phylogenetic analysis (Fig. 2). As shown in Fig. 2A, CcFrigida Essential 1 was placed close to the Frigida Essential 1 of soybean and chickpea. Similarly, CcFrigida was placed along with its respective orthologs from soybean and chickpea (Fig. 2). In the case of Frigida-like 1, CcFrigida-like 1 and CaFrigida-like 1 were grouped together, whereas their soybean orthologs were distantly placed (Fig. 2). However, Frigida-like 2, CcFrigida-like 2 was found to be closer to the maize and rice orthologs than to those of soybean and chickpea (Fig. 2). Similarly, Cc Terminal flowering 1 was found in the same clad as rice and maize (Fig. 2). However, CcTerminal flowering 2 was found to be closer to its orthologs in chickpea and soybean (Fig. 2).

Identification of possible interacting partner of pigeon pea flowering genes

A protein interacts with other proteins to carry out its function within a living thing and these interactions are essential for the organism's growth and development. To get an overview of the possible interacting partners of flowering genes in pigeon pea, their orthologs in *Arabidopsis* were used in the string protein database. As shown in Fig. 3A, Frigida interacts with early flowering, terminal flowering 1, Frigida essential 1 and leafy proteins. Frigida essential 1 interacts with terminal flowering 1, Frigida, TATA binding factor protein and Frigida-like 1 (Fig. 3A). Similarly, Frigida-like 1 shows interacting partners such as Frigida, Frigida essential 1, MAD box affecting flowering 1,

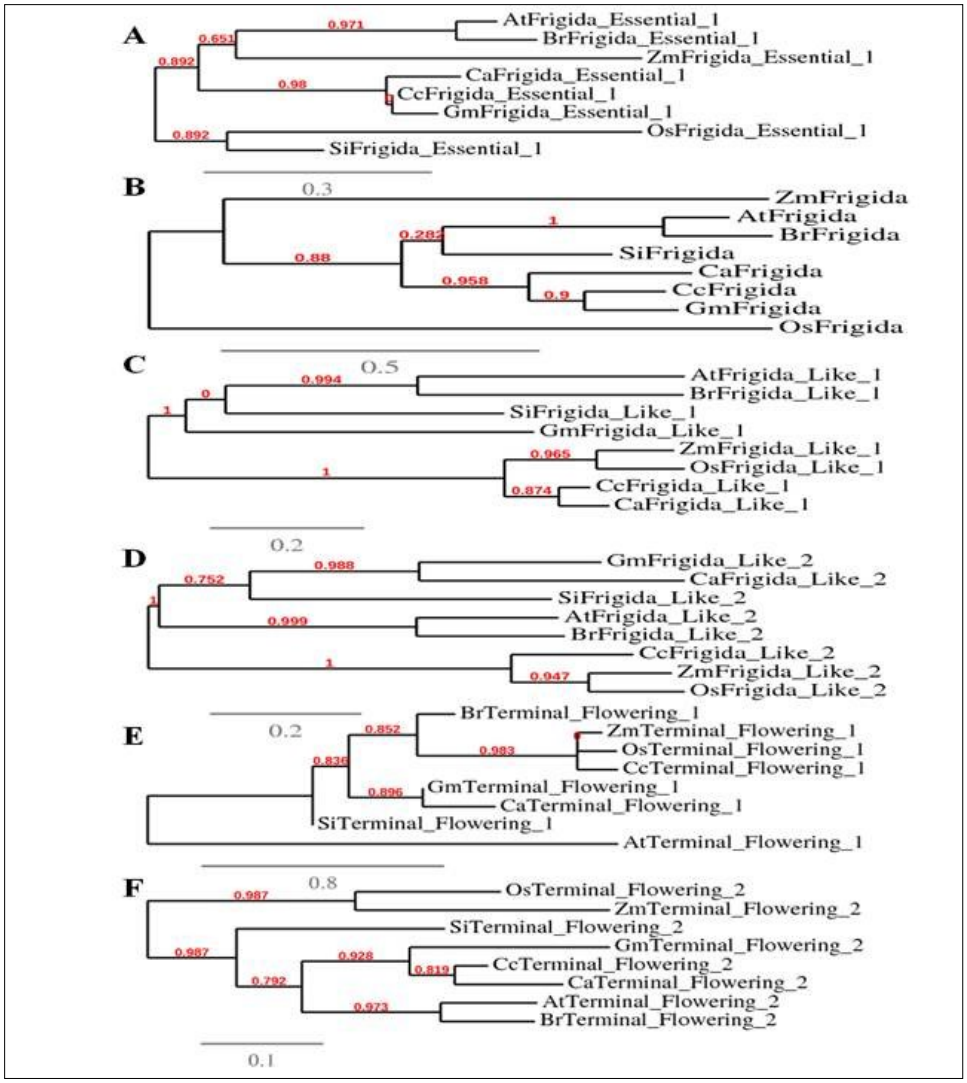

Fig. 2. Phylogenetic analysis of pigeon pea flowering genes with other plant species.

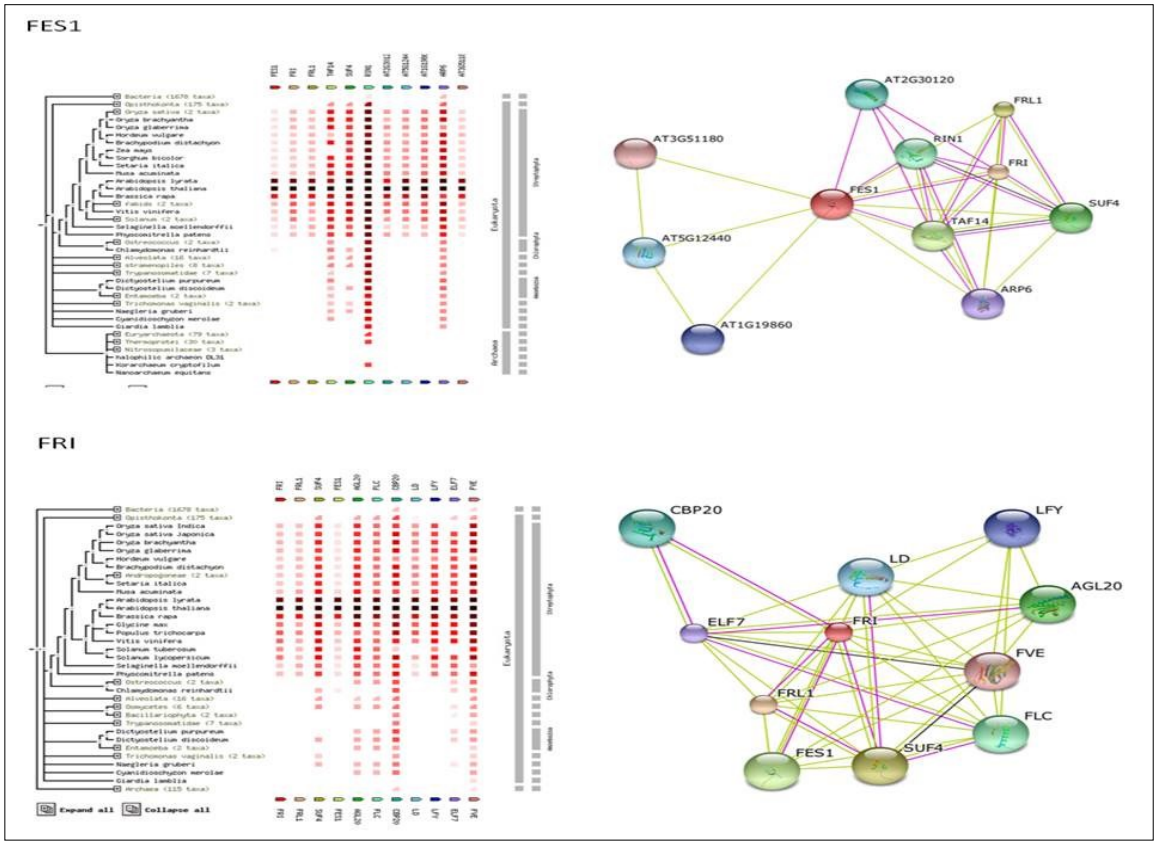

Fig. 3A. Conservation of Frigida Essential 1 and Frigida across flowering plants and their functional partner in different organism as predicted by STRING database.

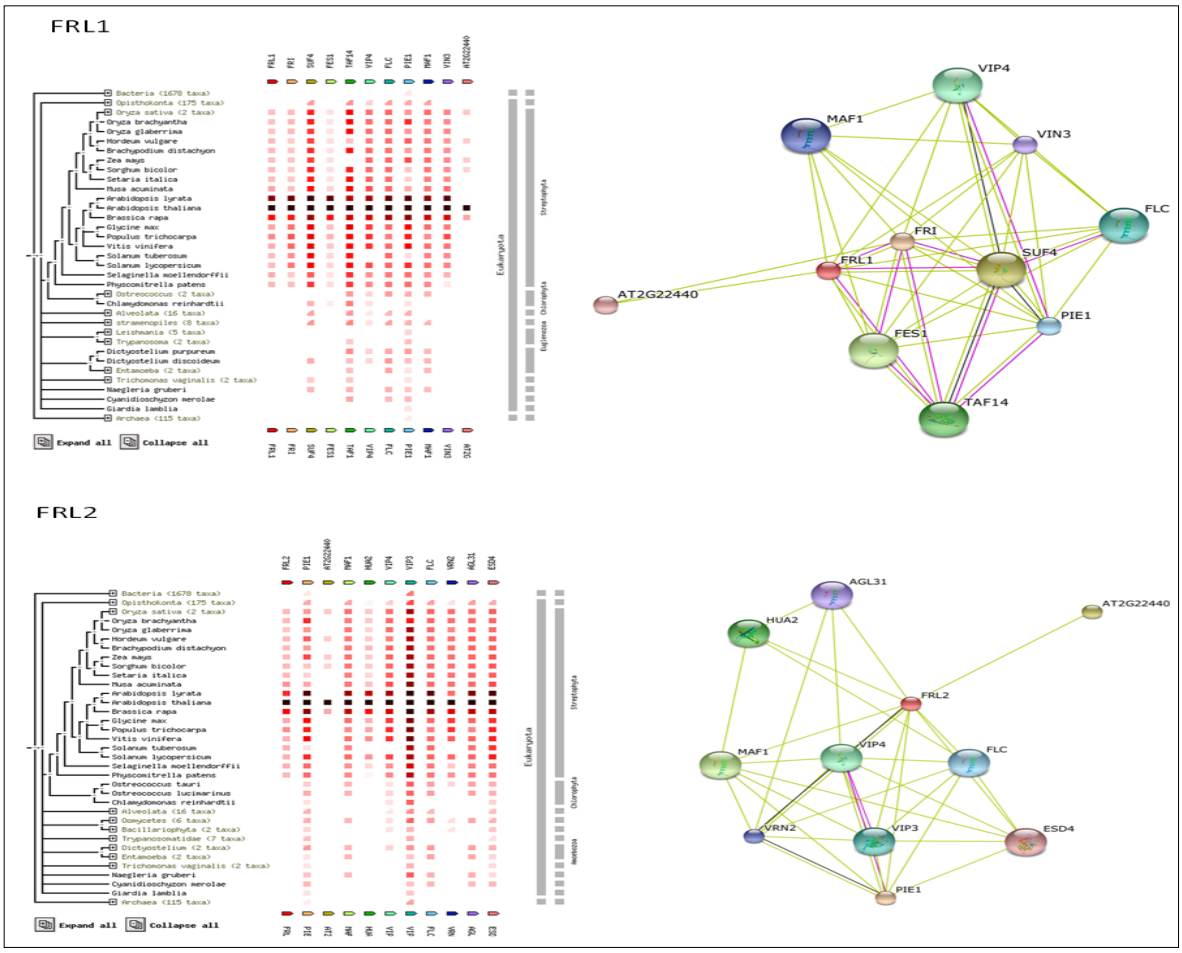

Fig. 3B. Conservation of Frigida Like 1 and Frigida Like 2 across flowering plants and their functional partner in different organism as predicted by STRING database.

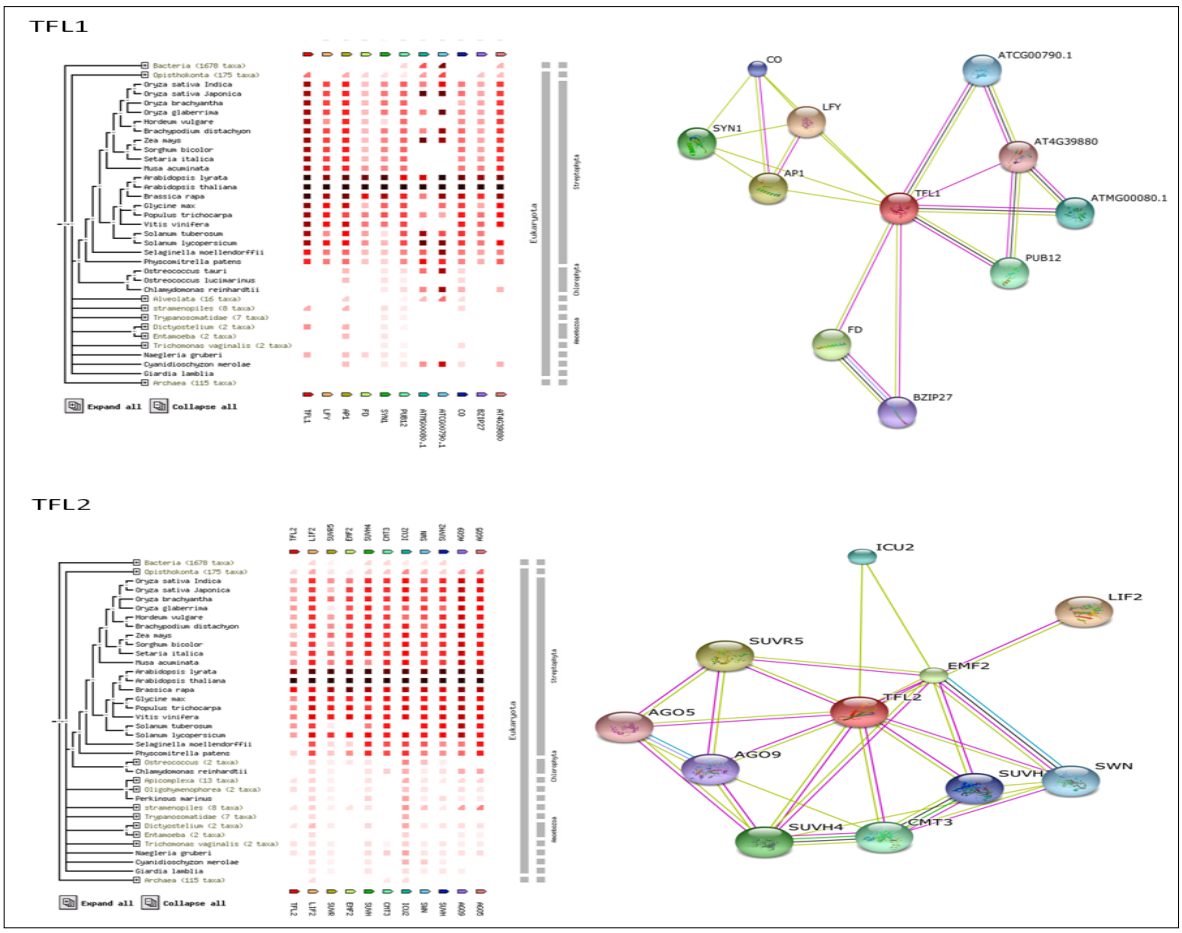

Fig. 3C. Conservation of Terminal flowering 1 and Terminal flowering 2 across flowering plants and their functional partner in different organism as predicted by STRING database.

flowering locus, etc. (Fig. 3B). Frigida-like 2 interacts with mad box affecting flowering 1, flowering locus c and HUA2 (Fig. 3B). Important interacting partners of Terminal flowering1 include Apetala1 and leafy, whereas Terminal flowering 2 interacts with LIF 2, ICU2 and EMF2 (Fig. 3C).

The results reveal that these genes and their interactions are largely conserved among flowering plants, indicating evolutionary stability in flowering mechanisms. Key regulatory pathways essential for flowering processes were found to be preserved, with core genes and their partners showing high levels of sequence similarity and functional conservation across different species. This suggests that the fundamental roles of these genes and their interactions are likely preserved in the development and regulation of flowering in plants (Fig. 3C).

Expression analysis of flowering genes in pigeon pea

Expression of flowering genes was studied using the pigeon pea SRA (Sequence Read Archives) data available in the SRA database at NCBI. RNA-seq data was obtained from the 2 different cultivars i.e. Asha (**SRX021565**) and Upas120 (**SRX021566**). The RNA from these plants were pooled from leaf, stem and root tissues. Expression analysis was performed using NGS commander software version 2.1.4. Expressions data for all the genes were

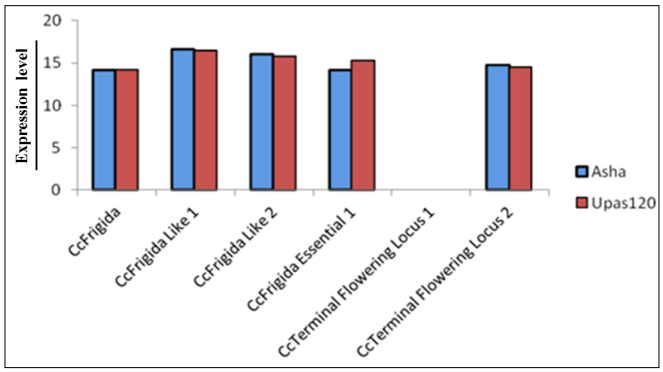

Fig. 4. Expression analysis of flowering related genes in 2 cultivars of pigeon pea. Expression represents log2 of RPKM value for SRA data.

obtained, except for the gene CcTFL1, as shown in Fig. 4.

Discussion

The timing of growing of flowers is an important trait for plant development. When the environment is favourable, delaying blooming prolongs the vegetative growth cycle, providing time for resources to accumulate for seed formation. On the other hand, plants have evolved to blossom early in response to unfavorable or uncertain environmental circumstances, according to reports (15). Early blooming is a favorable characteristic in many agricultural plants. Many genes linked to the control of flowering have been identified and cloned in plants, mainly in rice and *Arabidopsis thaliana*. On the other hand, little is known about the genes that control blooming in pulses. The advent of sequencing technology and the availability of whole genome sequences for agricultural plants have made it possible to use bioinformatics techniques to explore genes of interest (16).

In our research, we aimed to identify the genes that control flowering traits in pigeon pea (*Cajanaus cajan*). As observed, the process of flowering is a crucial factor in plant development, significantly impacting crop yield within an agricultural context. This feature has become even more important in the face of changing climatic conditions. The identified pigeon pea (*Cajanaus cajan*) genes were CcFrigida, CcFrigida Like1, CcFrigida Like2, CcFrigida Essential1, CcTerminal Flowering1 and CcTerminal Flowering2. Based on previously available data, potential genes for this investigation were selected (4, 17, 18). These genes have been shown to play a significant role in the regulations of flowering trait in several plants. The identified genes in our study appear to be transcriptionally active based on expression analysis using SRA data. Some genes show low or no expression due to their specific activation patterns over time and in specific locations (19).

In our study, functional gene ontology (GO) analysis was performed using Blast2Go and the results showed that all 6 proteins (CcFrigida, CcFrigida Like1, CcFrigida Like2, CcFrigida Essential1, CcTerminal Flowering1 and CcTerminal Flowering2) exhibited similar ontologies to their respective *Arabidopsis* genes with respect to molecular function, biological process and cellular components (Fig. 1). A similar observation was also made in Dashehari, which revealed the genetic model of flowering in mango, differing from previous reports (20, 21). Our study also identified important features of these 6 proteins, like molecular weight, PI and putative sub-cellular localization. The proteins Frigida, Frigida Like 1, Frigida Essential 1 and Terminal Flowering Locus 2 exhibited nuclear localization, consistent with the findings of a previous report (21-23). Frigida Like 2 showed both nuclear and cytoplasmic localization, while Terminal Flowering Locus 1 showed mitochondrial localization (24, 25) (Table 3). In our study, the identified proteins showed a normal molecular weight ranging from 61 to 99 kDa. Interestingly, it also possessed conserved domains that are found in orthologs from *Arabidopsis* plant (26).

In this study, we also identified the sequence-conserve domain, using *Arabidopsis* as a reference, through alignment tools along with the identification of flowering-related genes. The results showed more variations at the N-terminal end and relatively higher conservation at the C-terminal end of the protein (Supplementary Fig. 1A). This finding aligns with the previous report on the functional analysis of the flowering gene FRIGIDA, with the help of naturally occurring variation in *A. thaliana* (27).

When discussing the amino acid sequence alignment of these 6 flowering-responsive protein in pigeon pea with *A. thaliana*, we found relatively high sequence conservation across most of the proteins. These data suggests that these proteins play important and conserved roles in the flowering process across both plant species. The observed conservation implies that these proteins likely regulate the timing of flowering, a function that may be conserved throughout evolution. The presence of conserved domains and regions in the aligned sequences suggests functional similarities, implying that the mechanisms controlling flowering may have been preserved across these 2 plant species, as reported earlier (28-30).

Additionally, the presences of signature SUMOylation sites was observed for the 6 identified proteins in our study. All the predicted SUMOylation sites showed significantly higher score than the cutoff values, consistent with previous experiment (Table 4) (31, 32).

During our study, we analyzed the gene structure of the 6 pigeon pea genes related to flowering. Among these genes, Cc Frigida Essential 1 exhibited the highest number of introns, with a total of 5 introns. Interesringly, CcTerminal flowering 2 showed an even higher count of twelve introns, contrary to previous report (33, 34). Similarly, phylogenetic analysis was performed using respective orthologs from other plant species like *Arabidopsis thaliana*, *Zea mays*, etc. and providing relationships as per previous study (Fig. 2) (21, 35).

In this study, we also investigated protein interactions for the 6 genes to identify potential interacting partners of pigeon pea flowering genes. This is a vital step towards determining the expression patterns of these 6 genes, which facilitated to determine crop improvement, high yielding and speeding up the production in cultivars (31, 32, 36).

It is generally accepted that the genes under consideration play a crucial role in determining the blooming characteristics. Loss-of-function mutations in FRIGIDA (FRI) have been found to cause early blooming (13). The FRI gene has been discovered as the main cause of early flowering by genetic techniques. Similarly, a locus known as FRIGIDA-ESSENTIAL 1 (FES1) is critical for upregulating FLC expression (9). It has been shown that the Terminal Flowering Locus (TFL1) controls the timing of flowering in various plants in addition to *Arabidopsis*. Unlike model plants such as *A. thaliana*, which have many TFL1 homologs controlling different aspects of plant development, *Arabidopsis thaliana* has only one gene, TFL1 that regulates both purposes (12, 37).

CcTFL1 has been discovered as a prominent candidate gene for determinacy, by a variety of techniques, including candidate gene sequencing, mapping, QTL analysis, comparative genomics and expression profiling, in studies on flowering and determinacy genes in pigeon peas (37, 38). The presence of sequence variability in pigeon pea, as shown by parameters such as the number of SNPs, SNP frequency, nucleotide diversity and the number of haplotypes among 7 putative genes, clearly reflects the influence of distinct evolutionary constraints (39).

Conclusion

Advancements in bioinformatics and the availability of crop plant genome sequences have created opportunities to understand the genetic components responsible for intricate traits like flowering. This study focuses on understanding the genetic mechanisms governing the timing of flowering in pigeon pea *(Cajanus cajan)*, with the aim of improving crop enhancement strategies. We identified 6 potential flowering regulators in pigeon pea: CcFrigida, CcFrigida Like1, CcFrigida Like2, CcFrigida Essential1, CcTerminal Flowering1 and CcTerminal Flowering2. The identification of these genes provides valuable information for understanding the molecular mechanisms underlying the variation in flowering time in pigeon pea.

By exploring the genetic control of flowering time, we have potentially uncovered significant targets for selection to promote early flowering, which could lead to extended growth periods and enhanced seed production. The results revealed that key genes and their interactions are conserved among flowering plants, indicating evolutionary stability in flowering and suggesting that their crucial roles are preserved across species. The study also highlighted the importance of floral repressive genes like FLC and its activators as potential candidates for manipulation to accelerate flowering in pigeon pea.

In summary, studying the genetic differences and gene expression patterns in pigeon pea plants using advanced data (SRA data) helps us understand plant genetics in a better way. This knowledge will assist in developing better varieties with the desired traits through breeding. Despite certain limitations, our study provides a useful starting point for further research into the flowering gene network in pigeon pea, highlighting potential interacting partners that warrant experimental investigation. Future studies should aim to experimentally validate these interactions to create a more accurate and comprehensive interactome for pigeon pea flowering genes.

Acknowledgements

The authors express gratitude to the Center of Excellence, Natural Products and Therapeutics (COE-NPT), OHEPEE, World Bank, the Department of Biotechnology and Bioinformatics at Sambalpur University for their backing of this research. Additionally, appreciation is extended to Centurion University of Technology and Management, Jatni, Bhubaneswar, for their support and provision of infrastructure.

Authors' contributions

AP and JM done the conceptualization, data curation, investigation, writing - original draft, writing - review and editing. MB done the formal analysis, writing - original draft, writing - review and editing. RM done the writing - review and editing.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

References

1. Mula MG, Saxena KB. Lifting the level of awareness on pigeon pea-a global perspective. Inter Crops Res Institute for the Semi-Arid Trop. 2010;pp. 540.
2. Aja PM, Alum EU, Ezeani N, Nwali BU. Comparative phytochemical composition of *Cajanaus cajan* leaf and seed. Int J Microbiol Res. 2015;6 (1):pp.42–46.
3. Moursi YS, Dawood MFA, Sallam A, Thabet SG, Alqudah AM. Antioxidant enzymes and their genetic mechanism in alleviating drought stress in plants. In: Organic solutes, oxidative stress and antioxidant enzymes under abiotic stressors; 2021. 22:pp.233–62. https://doi.org/10.1201/9781003022879-12
4. Hecht V, Foucher F, Ferrándiz C, Macknight R, Navarro C, Morin J, et al. Conservation of *Arabidopsis* flowering genes in model legumes. Plant Physiol. 2005;137(4):pp. 1420–34. https://doi.org/10.1104/pp.104.057018
5. Du L, Ma Z, Mao H. Duplicate genes contribute to variability in abiotic stress re-sistance in allopolyploid wheat. Plants. 2023;12 (13):pp. 2465. https://doi.org/10.3390/plants12132465

6. Durai AA, Premachandran MN, Govindaraj P, Malathi P, Viswanathan R. Variability in breeding pool of sugarcane (*Saccharum* spp.) for yield, quality and resistance to different biotic and abiot-ic stress factors. Sugar Tech. 2015;17:pp.107–15. https://doi.org/10.1007/s12355-014-0301-x
7. Benedict C, Geisler MJB, Trygg J, Huner NPA. Consensus by democracy. Using meta-analyses of microarray and genomic data to model the cold acclimation signalling pathway in *Arabidopsis*. Plant Physiol. 2006;141(4):pp.1219–32. https://doi.org/10.1104/pp.106.083527
8. Mir RR, Saxena RK, Saxena KB, Upadhyaya HD, Kilian A, Cook DR, et al. Whole-genome scanning for mapping deter-minacy in pigeon pea (*Cajanus* spp.). Plant Breed. 2012a;132:pp.472-78. https://doi.org/10.1111/j.1439-0523.2012.02009.x
9. Liu B, Watanabe S, Uchiyama T, Kong F, Kanazawa A, Xia Z, et al. The soybean stem growth habit gene Dt1 is an ortholog of *Arabidopsis* terminal flower. Plant Physiol. 2010;153:pp.198–210. https://doi.org/10.1104/pp.109.150607
10. He Y, Doyle MR, Amasino RM. PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in *Arabidopsis.* Genes Dev. 2004;18(22):pp. 2774–84. https://doi.org/10.1101/gad.1244504
11. Shen L, Zhang Y, Sawettalake N. A molecular switch for FLOWERING LOCUS C ac-tivation determines flowering time in *Arabidopsis*. Plant Cell. 2022;34(2):pp. 818-33. https://doi.org/10.1093/plcell/koab286
12. Repinski SL, Kwak M, Gepts P. The common bean growth habit gene PvTFL1y is a functional homolog of *Arabidopsis* TFL1. Theor Appl Genet. 2012;124:pp.1539-47. https://doi.org/10.1007/s00122-012-1808-8
13. Werner JD, Borevitz JO, Uhlenhaut NH, Ecker JR, Chory J, Weigel D. FRIGIDA -independent variation in flowering time of natural *Arabidopsis thaliana* accessions. Genet. 2005;170(3):pp.1197–207. https://doi.org/10.1534/genetics.104.036533
14. Roux F, Touzet P, Cuguen J, Corre VL. How to be early flowering: an evolutionary per-spective. Trends Plant Sci. 2006;11(8):pp.375–81. https://doi.org/10.1016/j.tplants.2006.06.006
15. Verslues PE, Bailey-Serres J, Brodersen C, Bukley TN, Conti L, Christmann A, et al. Burning questions for a warming and changing world: 15 unknowns in plant abiotic stress. Plant Cell. 2023;35 (1):pp.67–108. https://doi.org/10.1093/plcell/koac263
16. Tan YC, Kumar AU, Wong YP, Ling APK. Bioinformatics approaches and applications in plant biotechnology. Genet Eng Biotechnol. 2022;20(1):pp. 1–3. https://doi.org/10.1186/s43141-022-00394-5
17. Chien PS, Chen PH, Lee CR, Chiou TJ. TWAS coupled with eQTL analysis reveals the ge-netic connection between gene expression and flowering time in *Arabidopsis*. BioRxiv. 2022;2022–12. https://doi.org/10.1101/2022.12.07.519424
18. Wang X, Miao H, Lv C, Wu G. Genome-wide association study identifies a novel BMI1A QTL allele that confers FLC expression diversity in *Arabidopsis thaliana*. J Exp Bot. 2023;30:pp. 120. https://doi.org/10.1093/jxb/erad120
19. Duan X, Zhang K, Duanmu H, Yu Y. The myosin family genes in soybean: Genome-wide identification and expression analysis. S Afr J Bot. 2023;160:pp. 338–46. https://doi.org/10.1016/j.sajb.2023.06.054
20. Kaur H, Sidhu GS, Mittal A, Yadav IS, Mittal M, Singla D, et al. Comparative transcriptomics in alternate bearing cultivar Dashehari reveals the genetic model of flowering in mango. Front Genet. 2023;10(13):pp. 1061168. https://doi.org/10.3389/fgene.2022.1061168
21. Mohanty JN, Nayak S, Jha S, Joshi RK. Transcriptome profiling of the floral buds and dis-covery of genes related to sex-differentiation in the dioecious cucurbit *Coccinia grandis* (L.) Voigt. Gene. 2017;626:pp. 395–406. https://doi.org/10.1016/j.gene.2017.05.058
22. Kim S, Choi K, Park C, Hwang HJ, Lee I. SUPPRESSOR OF FRIGIDA4, encoding a C2H2-Type zinc finger protein, represses flowering by transcriptional activation of *Arabidopsis* FLOWERING LOCUS C. Plant Cell. 2006;18(11):pp. 2985-98. https://doi.org/10.1105/tpc.106.045179
23. Choi K, Kim J, Hwang HJ, Kim S, Park C, Kim SY, Lee I. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. Plant Cell. 2011;23(1):pp. 289-303. https://doi.org/10.1105/tpc.110.075911
24. Zhang Y, Zeng L. Crosstalk between ubiquitination and other post-translational pro-tein modifications in plant immunity. Plant Commun. 2020;1(4):pp. 1-18. https://doi.org/10.1016/j.xplc.2020.100041
25. Lee J, Oh M, Park H, Lee I. SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY. Plant J. 2008; (5):pp. 832-43. https://doi.org/10.1111/j.1365-313X.2008.03552.x
26. Mohanty JN, Joshi RK. Molecular cloning, characterization and expression analysis of MADS-box genes associated with reproductive development in *Momordica dioica* Roxb. 3 Biotech. 2018;8(3):pp. 150. https://doi.org/10.1007/s13205-018-1176-4
27. Zhang L, Jiménez-Gómez JM. Functional analysis of FRIGIDA using naturally occur-ring variation in *Arabidopsis thaliana*. Plant J. 2020;103(1):pp.154-65. https://doi.org/10.1111/tpj.14716
28. Agarwal G, Garg V, Kudapa H, Doddamani D, Pazhamala LT, Khan AW, et al. Genome-wide dissection of AP2/ERF and HSP90 gene families in five legumes and expression profiles in chickpea and pigeon pea. Plant Biotechnol J. 2016;14(7):pp. 1563-77. https://doi.org/10.1111/pbi.12520
29. Satheesh V, Jagannadham PTK, Chidambaranathan P, Jain PK, Srinivasan R. NAC transcription factor genes: genome-wide identification, phylogenetic, motif and cis-regulatory element analysis in pigeon pea (*Cajanaus cajan* (L.) Millsp.). Mol Biol Rep. 2014;41:pp. 7763-73. https://doi.org/10.1007/s11033-014-3669-5
30. Mendapara I, Modha K, Patel S, Parekh V, Patel R, Chauhan D, et al. Characterization of CcTFL1 governing plant architecture in pigeon pea (*Cajanaus cajan* (L.) Millsp.). Plant. 2023;12(11):pp. 2168. https://doi.org/10.3390/plants12112168
31. Wang H, Feng M, Jiang Y, Du D, Dong C, Zhang Z, et al. Thermosensitive SUMOylation of TaHsfA1 defines a dynamic ON/OFF molecular switch for the heat stress response in wheat. Plant Cell. 2023;pp. 192. https://doi.org/10.1093/plcell/koad192
32. Roy D, Sadanandom A. SUMO mediated regulation of transcription factors as a mechanism for transducing environmental cues into cellular signaling in plants. Cell Mol Life Sci. 2021;78:pp. 2641-64. https://doi.org/10.1007/s00018-020-03723-4
33. Lee HT, Park HY, Lee KC, Lee JH, Kim JK. Two *Arabidopsis* splicing factors, U2AF65a and U2AF65b, differentially control flowering time by modulating the expression or alterna-tive splicing of a subset of FLC upstream regulators. Plant. 2023;12 (8):pp. 1655. https://doi.org/10.3390/plants12081655
34. Chen C, Wanyu X, Gaopu Z, Han Z, Huimin L, Lin W, Tana W. Transcriptome sequencing analysis of flowering related genes n *Pruns sibirica*. Mol Plant Breed. 2022;8:pp. 13. https://doi.org/10.5376/mpb.2022.13.0027
35. Li C, Zhang J, Zhang Q, Dong A, Wu Q, Zhu X, Zhu X. Genome-wide identification and analysis of the NAC transcription factor gene family in garden Asparagus (*Asparagus officinalis*). Genes. 2022;13(6):pp. 976. https://doi.org/10.3390/genes13060976
36. Li C, Lin H, Dubcovsky J. Factorial combinations of protein interactions generate a multiplicity of florigen activation

complexes in wheat and barley. Plant J. 2015;84(1):pp. 70-82. https://doi.org/10.1111/tpj.12960

37. Lembinen S, Cieslak M, Zhang T, Mackenzie K, Elomaa P, Prusinkiewicz P, et al. Diversity of woodland strawberry inflorescences arises from hetero-chrony regulated by TERMINAL FLOWER 1 and FLOWERING LOCUS T. Plant Cell. 2023;35(6):pp. 2079-94. https://doi.org/10.1093/plcell/koad086

38. Mir RR, Kudapa H, Srikanth S, Saxena RK, Sharma A, Azam S, et al. Candidate gene analysis for determinacy in pi-geon pea (*Cajanus* spp.). Theor Appl Genet. 2014;127:pp. 2663-78. https://doi.org/10.1007/s00122-014-2406-8

39. Mir RR, Kumar N, Jaiswal V, Girdharwal N, Prasad M, Balyan HS, Gupta PK. Genetic dissection of grain weight (GW) in bread wheat through QTL interval and association mapping. Mol Breed. 2012;29(4):pp. 963-72. https://doi.org/10.1007/s11032-011-9693-4