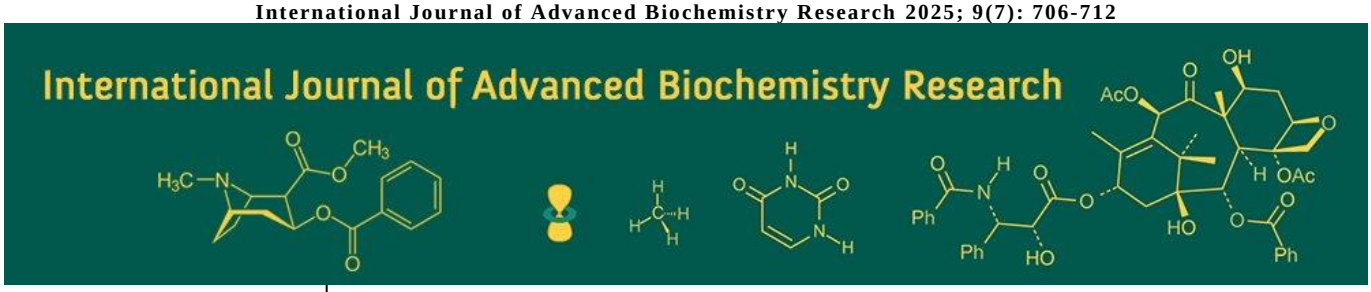

ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating: 5.29
IJABR 2025; 9(7): 706-712
www.biochemjournal.com
Received: 18-05-2025
Accepted: 21-06-2025

Vishwa Gohil
Department of Agricultural Statistics, N.M. College of Agriculture, NAU, Navsari, Gujarat, India

Alok Shrivastva
Department of Agricultural Statistics, N.M. College of Agriculture, NAU, Navsari, Gujarat, India

Naresh Chaudhary
Department of Genetics and Plant Breeding, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

KJ Bhuva
Department of Entomology, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Hemali Pandya
Department of Genetics and Plant Breeding, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Krunal Baria
Ph.D. Department of Genetics and Plant Breeding, Anand Agricultural University, Anand, Gujarat, India

Mori Krinal
Ph.D. Scholar, Department of Plant Biotechnology, Gujarat Biotechnology University, Gandhinagar, Gujarat, India

Shivangi J Vanpariya
Department of Entomology, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Corresponding Author:
Naresh Chaudhary
Department of Genetics and Plant Breeding, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Evolutionary insights into the Flowering Locus T (FT) gene in soybean (*Glycine max*)

Vishwa Gohil, Alok Shrivastva, Naresh Chaudhary, KJ Bhuva, Hemali Pandya, Krunal Baria, Mori Krinal and Shivangi J Vanpariya

DOI: https://www.doi.org/10.33545/26174693.2025.v9.i7i.4825

Abstract
Background: The *Flowering Locus T* (*FT*) gene plays a pivotal role in the regulation of flowering time and seasonal adaptation in plants. In this study, the *FT* gene from *Glycine max* was analyzed to explore its evolutionary relationships across leguminous species. Phylogenetic reconstruction revealed a high degree of sequence conservation, particularly among closely related genera, suggesting functional stability of *FT* in flowering regulation. The findings provide insight into the evolutionary dynamics of the *FT* gene and its potential role in legume breeding programs targeting climate-resilient flowering traits.
Methods: In present study, the nucleotide sequence of *Glycine max FT* (*GmFT*) was used to identify homologous sequences within the *Fabaceae* family through BLASTn analysis. As well as for identify protein homologous sequences. A total of 18 *Fabaceae FT* homologs, along with *Arabidopsis thaliana FT* as an out group, were aligned using the CLUSTAL W algorithm in MEGA 11. Using MEME and SMART-EMBL motif and domain predicted.
Result: The Tamura 3-parameter model with Gamma distribution (T92 + G) was identified as the best-fit nucleotide substitution model, based on the lowest BIC and AIC values. For protein JTT (Jones-Taylor-Thornton) substitution model found to best model. Phylogenetic reconstruction was performed using the Maximum Likelihood (ML) method reliability assessment using 1000 bootstrap replicates. Based on lower E-value selected motif submit to TOMTOM for finding transcription factor. With the help of domain identified functionality.
Conclusion: The evolutionary analysis of the FT gene in legumes reveals a strong conservation across species, alongside distinct lineage-specific variations. These differences likely contribute to the unique flowering patterns observed among leguminous crops. Understanding such genetic diversity provides crucial insights for breeding programs. Ultimately, this knowledge can guide targeted strategies to optimize flowering time and enhance stress resilience in legumes.

Keywords: Flowering Locus T, BLASTn, BLASTp, ClustalW, substitution model, phylogenetic analysis, domain and motif

Introduction
The transition from vegetative to reproductive growth is a critical phase in the plant life cycle, intricately regulated by both genetic and environmental cues. Among the central regulators of this transition is the *Flowering Locus T* (*FT*) gene, which encodes a small protein functioning as florigen, a mobile flowering signal. Initially characterized in *Arabidopsis thaliana*, *FT* integrates signals from multiple flowering pathways including photoperiod, ambient temperature and hormonal signaling to induce floral development at the appropriate time (Kobayashi & Weigel, 2007; Corbesier *et al.*, 2007) [13, 5].
The *FT* gene is a member of the phosphatidylethanolamine-binding protein (PEBP) family and has been found to be highly conserved across a wide range of plant species, including cereals, legumes and other angiosperms. In legumes, *FT* homologs play essential roles not only in the regulation of flowering time but also in regional adaptability, maturity group determination and ultimately, crop yield potential (Kong *et al.*, 2010; Jung *et al.*, 2012) [14, 12]. As legume species are grown in diverse agro-climatic zones, variations in *FT* gene sequences may represent molecular adaptations to environmental pressures such as photoperiod and temperature.

Phylogenetic analysis serves as a powerful tool for exploring the evolutionary trajectories of genes across species. Comparative phylogenetics of *FT* gene homologs among legumes and model plants provides valuable insights into sequence conservation, gene duplication events and divergence patterns. These evolutionary insights are critical for understanding the molecular basis of flowering regulation and for designing breeding strategies aimed at optimizing phenological traits in legume crops under changing climatic conditions.

Phylogenetic analysis in legumes plays a crucial role in enhancing crop breeding programs by elucidating genetic relationships and function of the gene. Wickland and Hanzawa (2015) [18] studied that phylogenetic analysis of *FT/TFL* gene family indicated that *FT* and *TFL* homologs were clustered in two major distinct groups, in line with their antagonistic function as flowering inducer and repressor, respectively. Flores *et al.* (2018) [7] studied that phylogenetic analysis of GTPases suggest that the number of family members and the primary sequence of small GTPases are well conserved between legume and non-legume plants. Krishna *et al.* (2022) [15] studied the phylogenetic analysis using *LprPhyA3* DNA and protein sequences indicated that *phytochrome* gene evolved from a common ancestry root but diverged into different clades during evolution. Abbas *et al.* (2023) [1] assessed the phylogenetic analysis of *R. solani* isolates across different legumes indicated that the distinct clades or subclades formed by the isolates correspond to their specific anastomosis groups (AGs) and subgroups, rather than being determined by their host legume crop. Phylogenetic anlaysis at genomic scale may unravel the hidden aspects of evolutiornay journey of legumes. Goyal *et al.* (2024) [9] studied that the phylogenetic analysis of PLD and POD genes showed strong evolutionary conservation aligned with their expression profiles, suggesting conserved functions. In contrast, the DMP genes exhibited greater divergence, indicating possible gene duplication and functional diversification. Gohil *et al.* (2025) [8] studied that unravelling the Evolutionary Blueprint of *Phytochrome A4* in Legumes: A Molecular Phylogenetic Approach. The phylogenetic analysis of the *PhyA4* gene at both protein and nucleotide levels showed clear and consistent clustering among leguminous species, indicating its evolutionary conservation and divergence across taxa.

Materials and Methods

Data Acquisition

In this study, sequence data for legume species were retrieved in FASTA format from the National Center for Biotechnology Information (NCBI) database (Anon., 2025) [2]. The focus was placed on flowering-related genes, specifically the *Flowering Locus T* (*FT*) gene.

Analysis Process of Build Phylogenetic Tree

The present study carried out phylogenetic analysis was conducted using Molecular Evolutionary Genetics Analysis version 11 (MEGA11) software (Hall, 2013 and Kumar *et al.*, 2018) [10, 16]. The following step-wise procedure was followed:

a. Data collection

The first step in constructing the phylogenetic tree involved developing a high-quality dataset comprising DNA and protein sequences retrieved from GenBank (NCBI) as a primary source in FASTA format.

b. Homology search

The retrieved sequences were used as queries in BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify homologous sequences within the NCBI GenBank database. Sequences from the *Fabaceae* family showing significant similarity were selected for further analysis.

c. Sequence curation

Selected sequences were screened to ensure completeness and quality. Only full-length or near full-length sequences were retained.

d. Multiple Sequence Alignment (MSA)

The curated sequences, including *Arabidopsis thaliana* as an outgroup, were aligned using the CLUSTAL W algorithm in MEGA11 Sievers and Higgins, 2018.

e. Model selection

The best-fit nucleotide or protein substitution model was identified using MEGA11 go to Models and select Find Best DNA/Protein Models. Select the alignment file and run the model test. The model with the lowest Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC) value was selected.

f. Phylogenetic tree construction

The phylogenetic tree was constructed using the selected substitution model. Choose Phylogeny in which select Construct/Test Neighbor-Joining (NJ), Maximum Likelihood (ML) or UPGMA Tree, depending on the method selected. Load the aligned sequence file. Set bootstrap replications to 1000 for statistical support. Generate the tree and export it in image format for documentation (Felsenstein, 1985) [6].

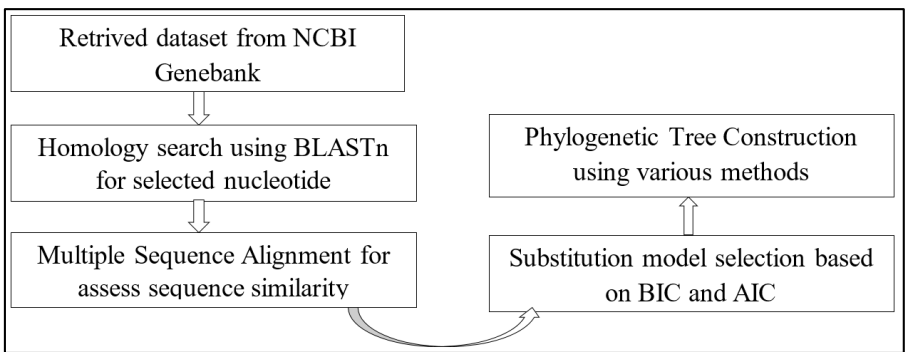

Fig 1: Flow chart of Phylogeny tree construction

Results and Discussion

Sequence retrieval, alignment and model selection for phylogenetic analysis of FT nucleotides

The processed *Glycine max Flowering Locus T* (*GmFT*) nucleotide sequence was used as a query in a BLASTn search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify homologous sequences within the NCBI nucleotide database. Homologous *GmFT* sequences from members of the *Fabaceae* family were retrieved for comparative analysis. Multiple sequence alignment of *FT* nucleotide sequences of 18 *Fabaceae* species, along with *Arabidopsis thaliana*, was performed using the CLUSTAL W algorithm implemented in MEGA 11 software. Nucleotide substitution analyses were conducted using sequences with complete deletion of gaps or missing data. Determine the most suitable substitution model, the "Find Best DNA/Protein Models" function in MEGA 11 was used. Based on the lowest Akaike Information Criterion (AIC = 22248.94) and Bayesian Information Criterion (BIC = 22,537.22) values, the Tamura 3-Parameter (T92) model was determined to be the best-fitting model for further evolutionary analyses (Table 1).

Table 1: Nucleotide substitution model of *FT*

Model	BIC	AIC	Gamma	Freq A	Freq T	Freq C	Freq G
T92 + G	22537.23	22248.94	0.75	0.31	0.31	0.19	0.19
HKY + G	22556.65	22264.66	0.75	0.30	0.32	0.18	0.21
TN93 + G	22565.89	22266.22	0.75	0.30	0.32	0.18	0.21
T92 + G + I	22569.19	22284.89	0.83	0.31	0.31	0.19	0.19
GTR + G	22572.31	22249.61	0.74	0.30	0.32	0.18	0.21
GTR + G + I	22579.32	22260.60	0.95	0.30	0.32	0.18	0.21
HKY + G + I	22587.77	22288.10	0.84	0.30	0.32	0.18	0.21
TN93 + G + I	22596.87	22289.53	0.85	0.30	0.32	0.18	0.21
T92 + I	22773.12	22496.50	n/a	0.31	0.31	0.19	0.19

Table 2: Substitution matrix using Tamura 3-parameter Model of *FT*

From\To	A	T	C	G
A	-	0.0714	0.0452	0.1034
T	0.0714	-	0.1034	0.0452
C	0.0714	0.1635	-	0.0452
G	0.1635	0.0714	0.0452	-

Table 2 represents a substitution matrix. It shows the rates at which one nucleotide changes to another nucleotide. Tamura 3-parameter model is a specific evolutionary model used in genetics to describe the rate of nucleotide substitutions. It takes into account differences in rates between transitions and transversions and G + C-content bias**.** The numbers represent the probability of one nucleotide changing to another within a given time period. Higher numbers indicate a higher rate of substitution**.** The rate of G to A and C to T (0.1635) was higher than the rate of T to A and C to A (0.0714), suggesting that transitions (purine to purine) were more frequent than transversions (purine to pyrimidine) in this specific model parameterization. The rate of G to A and C to T (0.1635) was the highest in the Table 2, indicating this particular transition was the most likely among the possibilities considered.

Construction of phylogeny tree of FT nucleotide

A phylogenetic tree was constructed using the Maximum Likelihood (ML) method based on the Tamura 3-Parameter (T92) model. Account for rate heterogeneity among sites, a discrete Gamma distribution (+ G) with five rate categories was applied. The initial tree for the ML analysis was generated using the Neighbor-Joining (NJ) method and Nearest-Neighbor Interchange (NNI) was employed as the ML heuristic search strategy. The best-scoring ML tree was selected to represent the evolutionary relationships among the 18 *Fabaceae* species analyzed.

A phylogenetic tree is a diagrammatic representation of the evolutionary relationships between different organisms. The phylogenetic tree illustrates the formation of distinct clades, reflecting evolutionary changes among the sequences. Species sharing a more recent common ancestor were placed closer together on the tree. The *GmFT* nucleotide sequence was compared with *FT* gene sequences from other plant species, primarily from the *Fabaceae* family. Phylogenetic analysis revealed that the *FT* gene evolved from a common ancestral root but diverged into different clades over the course of evolution.

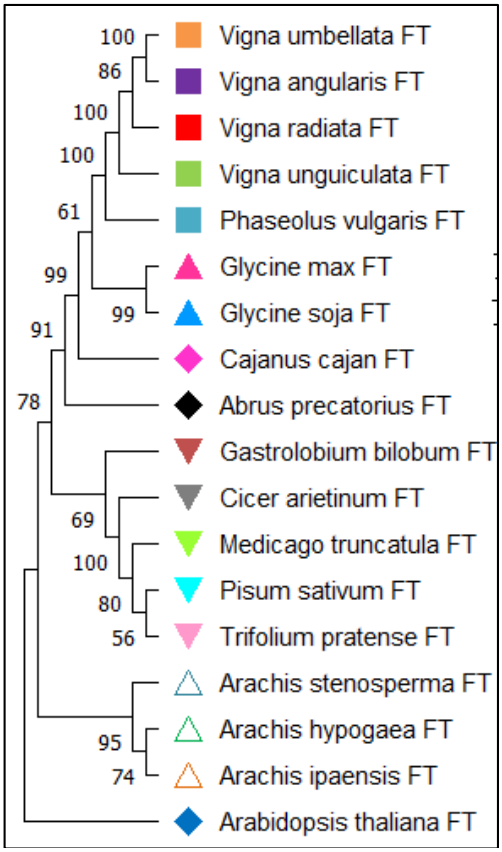

Fig 2: Phylogenetic tree of the FT nucleotide sequences of 18 different taxa (The tree was constructed using the maximum likelihood method with 1000 bootstrap replications) function

Figure 1 represented that *FT* sequences of *Vigna* species including *Vigna umbellata*, *V. angularis*, *V. radiata* and *V. unguiculata* cluster tightly together with high bootstrap support (100), indicating a strong evolutionary relationship among these species. *Phaseolus vulgaris* was also grouped closely with the *Vigna* clade, although with a lower bootstrap value (61), suggesting a slightly more distant relationship. The *Glycine* species (*G. max* and *G. soja*) form a strongly supported clade (bootstrap = 99), indicating high similarity between their FT gene sequences. These were closely related to *Cajanus cajan* and *Abrus precatorius*, supported by moderate to high bootstrap values (91 and 78, respectively). Further down the tree, *Gastrolobium bilobum*, *Cicer arietinum*, *Medicago truncatula*, *Pisum sativum* and *Trifolium pratense* form a group that represents more distantly related legumes, with varying bootstrap support. The *Arachis* species (*A. stenosperma*, *A. hypogaea*, and *A. ipaensis*) form a distinct and well-supported clade (bootstrap = 95), indicating strong evolutionary conservation of *FT* genes within this genus. *Arabidopsis thaliana*, used as an outgroup was placed on a separate branch, confirming its evolutionary distance from the *Fabaceae* family.

Phylogeny analysis of Flowering Locus T protein in Legumes

In present study, the tree of phylogeny analysis focuses on the Flowering Locus T (FT) protein. The processed *Glycine max* Flowering Locus T (GmFT) protein sequence was used as a query in BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify homologous sequences in the NCBI protein database. Homologous sequences from species within the *Fabaceae* family were retrieved. Multiple sequence alignment of 18 *Fabaceae FT* species, was performed using the CLUSTAL W algorithm in MEGA 11 (Molecular Evolutionary Genetics Analysis). The optimal substitution model was determined using the "Find Best DNA/Protein Models" function in MEGA 11. Nucleotide substitutions were analyzed using sequences with complete deletion of gaps or missing data.

Table 3: Substitution model of *FT* protein

Model	BIC	AIC	Gamma
JTT + G	4202.08	3996.17	0.70
JTT + G + I	4210.16	3998.22	0.70
JTT + I	4219.78	4013.87	n/a
cpREV + G	4226.24	4020.34	0.75
LG + G	4226.86	4020.95	0.71
LG + G + I	4234.94	4022.99	0.71
cpREV + G + I	4237.01	4025.06	0.75
WAG + G	4240.84	4034.93	0.72
LG + I	4246.61	4040.71	n/a

The results of substitution model selection for the FT gene, evaluated based on Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) values presented in Table 3. These criteria help identify the most appropriate evolutionary model by balancing model complexity with goodness-of-fit. Lower BIC and AIC scores indicate a better-fitting model.

Among the tested models, JTT + G (Jones-Taylor-Thornton model with Gamma distribution) exhibited the lowest BIC (4202.08) and AIC (3996.17) values, indicating it as the best-fitting model for the FT protein sequence data (Jones *et al.*, 1992) [11]. The model accounts for amino acid substitution probabilities based on empirical data, providing an accurate representation of evolutionary relationships. Bootstrap analysis with 1000 replicates was conducted to assess the reliability of the inferred phylogeny.

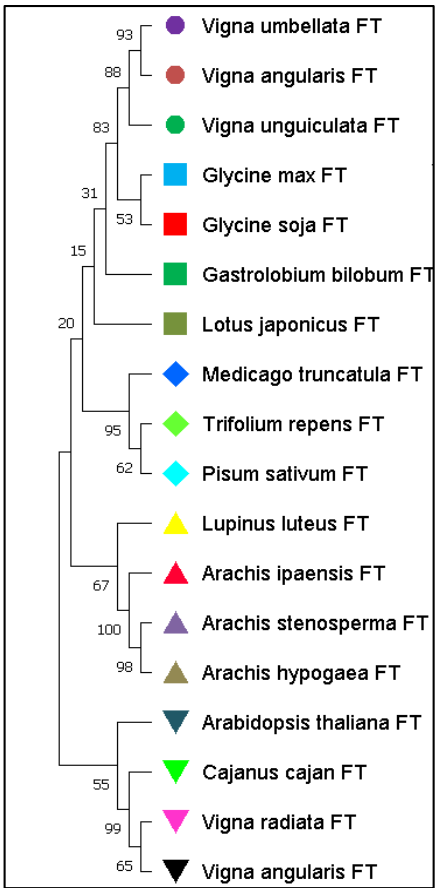

Fig 3: Phylogenetic tree of the *FT* protein sequences of 18 different taxa (The tree was constructed using the maximum likelihood method with 1000 bootstrap replications) function

The phylogenetic tree constructed based on FT (Flowering Locus T) gene sequences from various leguminous species, along with *Arabidopsis thaliana* as an outgroup, revealed distinct clustering patterns reflecting evolutionary relationships. Species belonging to the same genus grouped closely, indicating a high level of *FT* gene conservation within genera. For instance, *Vigna umbellata*, *Vigna angularis* and *Vigna unguiculata* formed a strongly supported clade (bootstrap values 93, 88 and 83, respectively), highlighting conserved evolutionary lineage within the *Vigna* genus. Similarly, *Glycine max* and its wild relative *Glycine soja* clustered together, although with moderate bootstrap support (53), suggesting some sequence divergence possibly due to domestication events.

Another notable clade included *Medicago truncatula*, *Trifolium repens*, *Pisum sativum*, and *Lupinus luteus*, which are all members of the *Papilionoideae* subfamily. This grouping, particularly between *Medicago* and *Trifolium*, was supported by high bootstrap values (up to 95), indicating close evolutionary proximity. *Arachis ipaensis*, *Arachis stenosperma* and *Arachis hypogaea* (cultivated peanut and its wild progenitors) formed a well-supported cluster (bootstrap values 100 and 98), reaffirming the genetic similarity of FT genes within the *Arachis* lineage.

Interestingly, *Arabidopsis thaliana* formed an independent branch, confirming its position as a non-legume outgroup and effectively rooting the tree. The positions of *Cajanus cajan*, *Vigna radiata* and an additional *Vigna angularis* entry were placed in a moderately supported clade (bootstrap values ~65), suggesting evolutionary divergence within the *Vigna* lineage or potential gene duplication events. Overall, the tree topology supports the hypothesis that the FT gene is evolutionarily conserved across legumes, with clade-specific variations reflecting genetic divergence, domestication and adaptation across different ecological niches.

Prediction of Common Motifs in *FT* Sequence

Conserved motif detection within the gene's DNA sequences was carried out using the online tool MEME (Multiple Em for Motif Elicitation), available at https://meme-suite.org/meme/ (Bailey *et al.*, 2009) [3]. Identification of conserved motifs in the *FT* was performed by comparing DNA sequences from these 17 *FT* sequences from different species. The analysis resulted in three conserved sites present in pool of studied sequence (Fig. 3). The best result was a 50-nt-long motif AATGGGTGTGAGTTCAAACCCTCACAAGTTGTCAA CCAACCAAGAGTAGC which was found in only 17 analysed sequences (Table 4). The E-value for best match discussed here was 3.7×10^{-377}. Subsequently, the motif was compared using TOMTOM (Beshir and Kebede, 2021) [4] with a database of Arabidopsis thaliana DAP motifs which resulted in identification of ND_tnt.AT1G63040_col_a_m1 (AT1G63040) motif containing transcription factor (p value 2.10×10^{-3} and E value $1.83\times10^{+0}$) (Fig. 4) and Table 4 represented that position of observed motif.

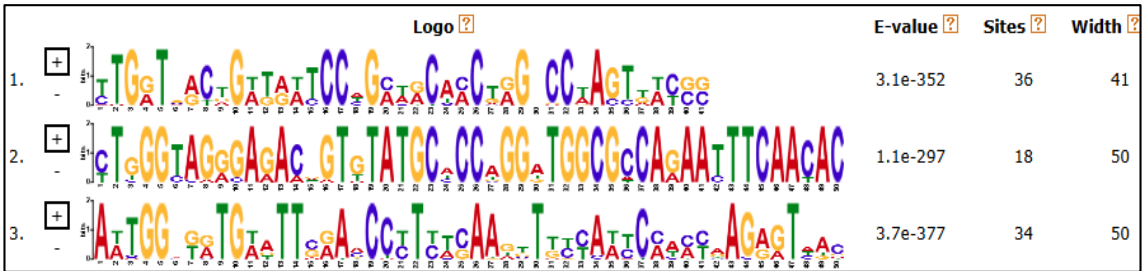

Fig 4: Motif that observed in *FT* sequence of legumes

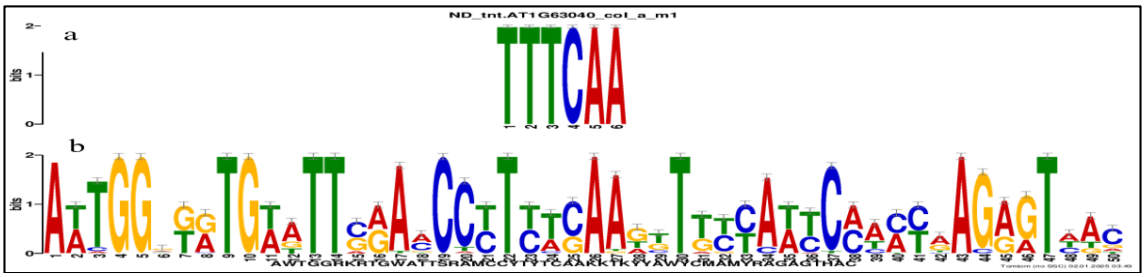

Fig 5: The sequence-logo comparison of conserved motifs between "a" *Arabidopsis thaliana* ND_tnt.AT1G63040_col_a_m1 (AT1G63040) and "b" the motif found in putative *FT* gene

Table 4: Predicted conserved motifs in 17 sequences of *Flwering Locus T* gene

Sequence	P-value	Motif 1	Motif 2	Motif 3
GmFT	7.88e-55	279-319	462-511	123-172
GsFT	1.11e-52	296-336	479-528	140-189
CcFT	7.63e-52	258-298	441-490	102-151
VrFT	3.66e-56	369-409	552-601	213-262
VumFT	2.31e-56	275-315	458-507	119-168
VaFT	8.59e-56	251-291	434-483	95-144
VunFT	4.72e-54	276-316	459-508	120-169
PvFT	1.19e-55	259-299	442-491	103-152
ApFT	1.02e-57	199-239	382-431	43-92
GbFT	3.48e-50	254-294	437-486	98-147
PsFT	4.81e-49	418-458	601-650	262-311
MtFT	1.74e-51	199-239	382-431	43-92
TpFT	4.35e-48	300-340	483-532	144-193
AhFT	7.37e-47	234-274	417-466	78-127
AsFT	5.82e-48	243-283	426-475	87-136
AiFT	1.47e-49	199-239	382-431	43-92
CaFT	3.36e-50	319-359	502-551	163-212
AtFT	2.05e-32	1151-1191	2171-2220	51-100 (scanned)

GmFT: Glycine max,GsFT: Glycine soja, CcFT: Cajanus cajan, VrFT: Vigna radiata, VumFT: Vigna umbellate, VaFT: Vigna angularis, VunFT: Vigna unguiculate, PvFT: Phaseolus vulgaris, ApFT: Abrus precatorius, GbFT: Gastrolobium bilobum, PsFT: Pisum sativum, MtFT: Medicago truncatula, TpFT: Trifolium pratense, AhFT: Arachis hypogaea, AsFT: Arachis stenosperma, AiFLD: Arachis ipaensis, CaFT: Cicer arietinum, AtFT: Arabidopsis thaliana

Domain Analysis of GmFT and GsFT

Domain analysis of GsFT (*Glycine Soja*) sequence retrieved from NCBI GenBank database after performing BLASTp search with GmFT (*Glycine max*) that encodes a protein with 176 amino acids was carried out using EMBL-SMART platform. This protein displayed normal features of FT with one domains *viz*., Pfam PBP (protein binding). GmFT Pfam PBP at 28 to 162 with E Value ($1.8e^{-14}$). Also for GsFT Pfam PBP at 28 to 162 with E Value ($5.30e^{-14}$). Pfam PBP key role in ligand binding, nutrient sensing, hormone signalling and transport. Position of this predicted domain mention in predicted protein structure of GmFLC in Fig 6.

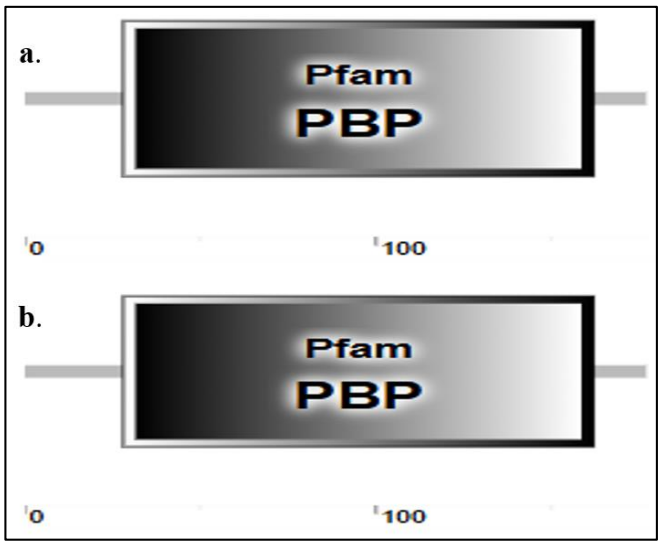

Fig 6: "a" Domain structure analysis of amino-acid sequence for GmFT and "b" Domain structure analysis of GsFT. The numbers represent position of amino acid

Fig 7: Predicted domain labelled in predicted protein structure of GmFT (PBP stands for Periplasmic Binding Protein)

Conclusion

The phylogenetic, motif and domain analyses of the *Flowering Locus T* (*FT*) gene across 18 leguminous species, with *Arabidopsis thaliana* as an outgroup, reveal a strong evolutionary conservation alongside lineage-specific divergence. The Tamura 3-Parameter (T92) and Jones-Taylor-Thornton model (JTT + G) model, selected based on the lowest BIC and AIC values, ensured accurate nucleotide and protein-based phylogenetic inference. Maximum Likelihood-based phylogenetic trees, constructed using both nucleotide and protein sequences, showed well-supported clades aligning with taxonomic relationships, especially within *Vigna*, *Arachi* and *Glycine* genera. Motif analysis identified three conserved regions, including a 50-nt motif with a highly significant E-value, while domain analysis confirmed the presence of the conserved Pfam PBP domain in *Glycine max* and *Glycine soja*. These findings highlight the functional conservation of FT across legumes, while also reflecting adaptations shaped by ecological and domestication pressures providing a valuable foundation for breeding strategies focused on flowering time regulation and stress resilience.

References

1. Abbas A, Ali A, Hussain A, Ali A, Alrefaei AF, Naqvi SAH, Baloch FS. Assessment of genetic variability and evolutionary relationships of *Rhizoctonia solani* inherent in legume crops. Plants. 2023;12(13):2515.
2. Anonymous. NCBI (National Center for Biotechnology Information) [Internet]. 2025 [cited 2025 Jan 18]. Available from: https://www.ncbi.nlm.nih.gov/
3. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research. 2009;37(Web Server):W202-W208.
4. Beshir JA, Kebede M. *In silico* analysis of promoter regions and regulatory elements (motifs and CpG islands) of the genes encoding for alcohol production in *Saccharomyces cerevisiae* S288C and *Schizosaccharomyces pombe* 972h. Journal of Genetic Engineering and Biotechnology. 2021;19(1):8.
5. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, Coupland G. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science. 2007;316(5827):1030-1033.
6. Felsenstein J. Phylogenies and the comparative method. The American Naturalist. 1985;125(1):1-15.
7. Flores AC, Via VD, Savy V, Villagra UM, Zanetti ME, Blanco F. Comparative phylogenetic and expression analysis of small GTPases families in legume and non-legume plants. Plant Signaling and Behavior. 2018;13(2):e1432956.
8. Gohil V, Shrivastava A, Modha K, Pandya H, Chaudhary N, Varshney N, Patel P, Garde Y. Unravelling the evolutionary blueprint of phytochrome A4 in legumes: A molecular phylogenetic approach. Journal of Biology and Biotechnology. 2025;28(7):537-544.
9. Goyal L, Kaur M, Mandal M, Panda D, Karmakar S, Molla KA, Bhatia D. Potential gene editing targets for developing haploid inducer stocks in rice and wheat with high haploid induction frequency. 3 Biotech. 2024;14(1):14.
10. Hall BG. Building phylogenetic trees from molecular data with MEGA. Molecular Biology and Evolution. 2013;30(5):1229-1235.
11. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Bioinformatics. 1992;8(3):275-282.
12. Jung CH, Wong CE, Singh MB, Bhalla PL. Comparative genomic analysis of flowering genes between *Arabidopsis* and soybean. Plant Molecular Biology Reporter. 2012;30(5):1069-1080.

13. Kobayashi Y, Weigel D. Move on up, it's time for change: mobile signals controlling photoperiod-dependent flowering. Genes & Development. 2007;21(19):2371-2384.
14. Kong F, Liu B, Xia Z, Sato S, Kim BM, Watanabe S, Yamada T, Tabata S, Kanazawa A, Harada K, Abe J. Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. Plant Physiology. 2010;154(3):1220-1231.
15. Krishna S, Modha K, Parekh V, Patel R, Chauhan D. Phylogenetic analysis of phytochrome A gene from *Lablab purpureus* (L.) Sweet. Journal of Genetic Engineering and Biotechnology. 2022;20(1):9.
16. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution. 2018;35(6):1547-1549.
17. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Science. 2018;27(1):135-145.
18. Wickland DP, Hanzawa Y. The Flowering Locus T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms. Molecular Plant. 2015;8(7):983-997.