

Identification of flavonol synthase gene from *M. pinnata* and its functional elucidation using cowpea genome

Garima Kumari

BIT Mesra: Birla Institute of Technology

Vinod Kumar Nigam

BIT Mesra: Birla Institute of Technology

Dev Mani Pandey (✉ dmpandey@bitmesra.ac.in)

BIT Mesra: Birla Institute of Technology <https://orcid.org/0000-0002-0711-9966>

Research Article

Keywords: Cowpea, Flavonol synthase, Genome wide analysis, *Millettia pinnata*, Polymerase Chain Reaction, Sequencing, Synteny

Posted Date: December 13th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2301362/v1>

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

Abstract

Background

Flavonoids are a class of key polyphenolic secondary metabolites having various roles in plants, like in growth and development, stress defense, and reproduction. Dihydroflavonols are converted into flavonols by flavonol synthase (FLS) and this enzyme belongs to the 2-oxoglutarate-dependent dioxygenases (ODD) superfamily. *Millettia pinnata* is important medicinal plant possessing various phytoconstituents involved in the treatment of various diseases. The FLS is involved in the synthesis of a natural flavonol 3,4',5,7-tetrahydroxyflavone i.e., kaempferol. However, detailed information about FLS gene sequence and its characterization in *M. pinnata* is not available.

Methods

In the current investigation the full-length cDNA sequences of the FLS gene (*MpFLS*) was deduced by isolating RNA from *M. pinnata* leaf, cDNA prearaion, PCR based amplification, sequencing followed by sequence analysis.

Results

The full-length cDNA of *MpFLS* containing 952-bp open reading frame and encoding 310-amino-acid protein has been elucidated, and submitted to NCBI with accession number OM469017. Bioinformatics analysis of *MpFLS* protein showed 85% similarities with *VuFLS* (*Vigna unguilata*). Also, BLAST analysis of *M. pinnata* sequence exhibited maximum similarity with *V. unguilata*. Therefore, to find the possible functional characteristics of *MpFLS*, the genome wide investigation of this gene was performed in cowpea (*V. unguilata*) genome using Phytozome v13 database. Total 17 full length non-redundant *VuFLS* genes were identified in *V. unguilata*. These genes and the corresponding proteins were examined for their physicochemical characteristics, chromosomal location, domain and motif architecture, and phylogeny. Phylogenetic and motif analysis revealed that all the proteins had the 2OG-Fell Oxy domain. The functional annotation analysis revealed their involvement in various biological processes and molecular processes. The majority of the genes at the subcellular level are found in the cytoplasm, followed by the nucleus.

Conclusions

Our research study provides information about *FLS* gene sequence in *M. pinnata* and its detailed characterization. Our result also revealed that *MpFLS* shows similarity in functional characteristics with *VuFLS*.

Highlights

- The isolation, amplification and sequencing of Flavonol synthase from *M. pinnata*, designed as *MpFLS*, has been performed and submitted to NCBI with accession no. OM469017.
- *In-silico* characterization and analysis of FLS reveals that *MpFLS* is closely related to *Vigna. unguilata* flavonol synthase (*VuFLS*).
- The genome-scale identification and characterization of *MpFLS* has been done in cowpea genome (*Vigna. unguilata* v1.2) because *M. pinnata* genome is not available yet. Here, conserved domain, motifs, phylogenetic tree, Gene structure, chromosomal localization and syntenic analysis has been elucidated.

1. Introduction

Modern and traditional medicine continues to benefit from the therapeutic properties of medicinal plants [1]. Traditional remedies are becoming more important due to the negative effects of modern medication, and researchers are now looking into them to determine the scientific basis of their therapeutic effects [2]. A medium-sized glabrous tree with the common names Karanja in Hindi, Pongam in Tamil and Indian Beech in English is called *Pongamia pinnata* (Linn.) Pierre [3]. The majority of Ayurvedic and Siddha practitioners in Tamil Nadu, India utilise *P. pinnata* to cure a variety of illnesses, including diabetes mellitus. At present, different plant species have produced more than 9000 unique flavonoids. As the most prevalent flavonoids, flavonols serve crucial functions in plant physiology, development, and growth, including UV damage protection and flower colour change [4, 5]. They are categorised, for instance, as kaempferols and quercetins, according to the B ring's hydroxylation pattern [6].

Flavonol synthase (FLS), which is typically defined as possessing desaturation and hydroxylation activity (addition of a double bond in dihydroflavonols between C2 and C-3 of (2R, 3R) and is categorised as a bifunctional dioxygenase, converts three common forms of flavonols, including quercetin, kaempferol and myricetin respectively [7]. Belonging to the 2-ODD, the 2nd largest family of higher plants, it catalyses the oxidation reaction with the help of cofactors ferrous iron or 2-oxoglutarate [7]. Takahashi (2007) [8] have studied the FLS cDNA from

Arabidopsis thaliana, *Petroselinum crispum*, *Zea mays*, *Oryza sativa*, strawberry, and *G. max*. Previous reports [9, 10, 11] describes those various copies of FLS was expressed in organ-specific way and also coded by a multicopy gene families in plants. According to Chandran, A.K.N et al., 2006 [12], the elucidation of the morphological agronomic traits in Rice was also done by its analysis and identification using genome-wide study.

The characterization of FLS activity and protein preparation was done first time in cell suspension culture of *Petroselinum crispum* [13]. Additionally, FLS activity has also been shown in the anthers of *Tulipa sp.* and in flower extracts of *Gerbera hybrida*, *Petunia hybrida*, *Dianthus caryophyllus*, and *Matthiola incana* [14, 15]. During flower development, there is a clear change in the activity of FLS in *Petunia*, *Dianthus*, and *Matthiola* which correlates the accretion/accumulation of flavonols. The uncolored flower buds possess high FLS activity, however it starts declining quickly as anthocyanin formation starts. A more thorough description of the *Petunia* FLS clone has been provided. The prior research on FLS enzyme activity correlated with FLS gene expression throughout the development of buds and flowers. An FLS cDNA clone was heterologously expressed in yeast, producing FLS protein. In *Petunia*, FLS antisense expression significantly decreased the synthesis of flavonols in the flower corollas [16]. Also, the flavonol synthase gene was isolated and characterized from *Ginkgo biloba* [17].

Millettia pinnata is an important medicinal plant containing many important phytoconstituents like flavonols, chalcones, flavones etc., which are well-known for their anti-inflammatory, antioxidant, and anti-proliferative, properties that protect against cardiovascular and cancer disease [18]. The FLS also helps in the synthesis of a natural flavonol 3,4',5,7-tetrahydroxyflavone i.e kaempferol [19]. Various parameters of FLSs may also be gathered from the available data of the whole genome sequence. However, information related to FLS family gene or whole genome sequence of *M. pinnata* is not available. Although, FLS of *M. pinnata* shows 85% similarity with *Vigna unguilata* (cowpea). Therefore, we examined the putative FLS at the genome level and its characterization in cowpea.

The current article was studied with the aim to characterise the *M. pinnata* FLS gene. However, FLS gene sequence as well as reports on its functional characterization in *M. pinnata* has not yet been thoroughly studied. Based on PCR amplified *M. pinnata* FLS gene sequence similarity with cowpea FLS gene, and hence detailed genome wide study on cowpea was studied by using the homology search method. Total 17 novel FLS genes in the cowpea genome were retrieved. Further, detailed characterization of exon-intron positions, conserved domains, sub-cellular localisation, gene ontology, evolutionary relationships, synteny, and chromosomal localizations were carried out.

2. Materials And Methods

2.1 Total RNA isolation from leaf of *M. pinnata*

The total RNA isolation was done from *M. pinnata* leaves by using TRIZOL method [20]. All the extracted RNAs were treated with DNase I. Quality of RNA was assessed by performing denaturing electrophoresis on a 1% formaldehyde agarose gel [21]. The quality and concentration of the total RNA was determined by spectrophotometer and the electrophoresed RNA was visualized using gel documentation system.

2.2 cDNA synthesis and PCR gene amplification

The first strand of cDNA was synthesized using manufacturer's protocol (Invitrogen, Life Technologies, USA). The polymerase chain reaction (PCR) amplification of flavonol synthase (FLS) gene was done by following the TAKARA PCR mix protocol. Primers designed for FLS gene amplification are Forward primer- 5'- ATGGAGGTGCTAAGGGTG-3' and Reverse primer- 5'- TCATTGAGGGATCTTATT-3'.

2.3 Sequencing and bioinformatics analysis

PCR amplified product was submitted for sequencing (Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad. Obtained sequence was analysed using blastn, tblastn and blastptools available at NCBI (<http://www.ncbi.nlm.nih.gov>) and MEGAX (<http://www.megasoftware.net/>) [22]. Subsequently, analysed sequence was submitted to NCBI GenBank (<http://www.ncbi.nlm.nih.gov/genbank/submit/>). The open reading frame of gene sequence was identified by ORF finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). The ExpASY translate online tool (<http://web.expasy.org/translate/>) was used to convert the complete ORF nucleotide sequence to protein sequence. The conserved domains and motifs analysis was done by NCBI conserve domain database and MEME suite, respectively. The phylogenetic tree analysis of nucleotide and translated protein sequences was done by MEGA X software. The secondary structure analysis was done by SOPMA and 3-D structure prediction was done by I-TASSER online server. The predicted structure was validated using SaVES server.

2.4 Similarity of Flavonol synthase in *M. pinnata* and *V. unguilata*

The sequence obtained after sequencing was subjected to protein blast with default setting. Based on maximum similarity the FLS sequence of *V. unguilata* was retrieved from NCBI with accession no. QCD92802. Subsequently, physico-chemical characterization of the sequence was

done using ExPASy protParam online tool (<https://web.expasy.org/protParam/>) while its conserved domain analysis was performed using domain finder tool (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

2.5 Identification of the FLS gene family from the *Vigna unguilata* genome

The FLS family genes from the *V. unguilata* genome were identified using the *M. pinnata* FLS protein sequences as a query protein. The transcripts from the cowpea genome that belong to the FLS family were located using the online database "Phytozome v13" (<https://phytozome-next.jgi.doe.gov/>) [23, 24]. All the protein sequences which do not contain the FLS domain were removed. The tool motif finder was used to verify the domains in each transcript. Each protein transcript ID was changed to the gene name in accordance with its location on the *V. unguilata* chromosomes. The CD database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) was used to identify conserved domains in each protein sequence and then it was visualised using TBtools. Using "Phytozome v13" tool (<https://phytozome-next.jgi.doe.gov/>) further details about the genes, including the forward or reverse strand, start-end point, CDS, and protein length (aa) were retrieved. Also, PI, GRAVY, and Mw of each sequence were discovered using the ExPASy protParam tool (<https://web.expasy.org/protParam/>) [25]. The Wolf Psort (<https://wolfpsort.hgc.jp/>) helped to estimate the sub-cellular localization of each FLS protein and is represented using TBtools [26].

2.6 Phylogenetic, motif analysis and gene structure of FLS proteins

MEGA X (Molecular evolutionary genetics analysis) software was used for alignments of all the protein sequences [27]. The tree was constructed using the maximum likelihood approach with 100 bootstrap for the comparative evolutionary analysis of FLS genes in various species. An appropriate evolutionary model that is based on the available data and evaluated can predict the likelihood of character state changes through the calculation of evolutionary history patterns. The FLS protein sequences in cowpea were examined for conserved motifs using the tool MEME available at (<https://meme-suite.org/meme/>) [28]. The annotation of each member was examined by the help of Pfam database (<https://pfam.xfam.org/>), with ten as the default maximum number of motifs [29]. For gene structure analysis, the coding and genomic sequences of each cowpea FLS gene were used from Phytozome v 13. The Gene Structure Display Server 2.0 (<http://gsds.gao-lab.org/>) [30] gives the information of number of exons, introns, and untranslated regions (UTR) in each cowpea FLS gene.

2.7 Chromosome localization, gene duplication and Synteny analysis

Phytozome tool gives the information related to length, position, and number of chromosomes of each FLS gene and then they were mapped to their respective number of the chromosomes. The chromosome localization of FLS genes was depicted by Phenogram Plot (<http://visualization.ritchielab.psu.edu/phenograms/plot>). The genes were assembled according to the following criteria: the identity of the align sequences was greater than 80% and the aligned nucleotide coding sequence spanned 80% of the longer gene [31]. Tandemly duplicated genes have a physical spacing of 50kb between two homologous genes [32]. From phytozome, general feature format (GFF) and genome fasta data of *V. unguilata*, were extracted. The data was run with One Step MCScan for identifying the "syntenic genes" utilised in synteny analysis. Then, TBtools dual system plot was used to visualise orthologous FLS genes on syntenic chromosomal blocks.

2.8 Promoter region analysis and sub-cellular localisation

The upstream sequence upto 1500 bp from ATG, the translation initiation site was obtained from Phytozome v 13. Then PlantCARE (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>), the online tool plant Cis-acting regulatory element was taken for analysing of the promoter or cis-regulatory elements present in each *VuFLS* promoter regions. The online tools Wolf Psort (<https://wolfpsort.hgc.jp/>) was used for predicting the subcellular localisation of each *VuFLS* proteins and then the result was visualized through TBtools [26].

3. Results

3.1 Characterisation of the deduced *MpFLS* protein

Finally, a 952-bp open reading frame full-length cDNA sequence of the *MpFLS* gene was isolated from *M. pinnata*. The deduced *MpFLS* protein had an isoelectric point of 5.10, 340 amino acids, with an estimated molecular weight of 77.856 kDa. The NCBI programme Blastp showed that *MpFLS* had a lot in common with other plant FLSs. In comparison to its counterparts in *V. unguilata*, *G. max*, *A. precatorius*, *C. ternatea*, and *A. hypogrea*, the *MpFLS* displayed 85, 84, 83, 82, and 75% identities. By searching the GenBank conserved domain database and doing a function analysis, it was discovered that *MpFLS* is a member of the 2OG-Fell Oxy super family, which is distinguished by the presence of a conserved 2OG-Fell Oxy domain (<http://www.biochem.ucl.ac.uk/bsm/cath/Gene3D>) as given in supplementary file.

3.2 Similarity of flavonol synthase in *M. pinnata* and *V. unguilata* (cowpea)

The protein blast of FLS sequence of *M. pinnata* shows 85% similarity with *V. unguilata* FLS. Also, physico-chemical similarity of FLS between both the species shows that *M. pinnata* FLS is of 310 amino acids having Mw 36165.70, GRAVY – 0.493, Theoretical pI- 8.37, and has 2OG-Fell

Oxy domain. Whereas, *V. unguilata* FLS is of 331 amino acids having Mw 37944.60 Da, GRAVY – 0.448, Theoretical pI-6.20 and also possess 2OG-Fell-Oxy as conserved domain (Supplementary file). The above results show maximum similarities between *M. pinnata* and *V. unguilata* FLS sequences. Also, whole genome sequence of *M. pinnata* is not available till now, hence, further genome-scale study of FLS was done taking *V. unguilata* genome.

3.3 Genome wide identification of FLS genes

Genome wide analysis of FLS gene has been carried out in many plants as well as in model plant. However, in *M. pinnata* plant the genome analysis of this gene family is still not done. Hence, the goal of the current study was to thoroughly study the FLS, family at the level of the genome. The FLS sequence isolated and sequenced from *M. pinnata* shows 85% similarity with *V. unguilata* FLS sequence. Therefore, the whole genome analysis of FLS was done in cowpea genome. Pfam database (<https://pfam.xfam.org/>) revealed the domain (2OG-Fell Oxy) region starting from 183 and ends at 252 having clan CL0029 and that domain was blast against *V. unguilata* v1.2 genome from Phytozome v 13 database since, *M. pinnata* genome is not available till date. The blast results show total 154 hits but maximum aligned sequences (> 60%) was selected (total 17 sequences) for further studies by a genome-wide HMM method by allotting a unique identification from *VuFLS*-1 to *VuFLS*-17 (*Vigna. unguilata* flavonol synthase). Table 1 depicts the physico-chemical properties of all *VuFLS* families including their transcript ID, chromosome location, strand, isoelectric point (pI), sequence length, number of exons, molecular weight also estimated subcellular localization. The length of the protein sequence varied from 185 amino acids (Vigun09g073700_ *VuFLS*-8, smallest among 17 sequences) to 393 amino acids (Vigun05g116600_ *VuFLS*-5, largest among 17 sequences). In addition to this, the molecular weight of the sequences also varied from 21086.15 kDa (Vigun09g073700_ *VuFLS*-8) to 44604.8 kDa (Vigun05g116600_ *VuFLS*-5). The result of sub-cellular localisation of all the genes depicts that maximum gene fall under category of cytosol followed by nucleus and very few of them are in Golgi and vacuole and has been represented by heatmap (Fig. 8).

3.4 Phylogenetic, motif analysis and gene structure of FLS proteins

The molecular evolutionary genetics analysis (MEGA X) software was used for the multiple sequence alignments of all the 17 protein sequences. The phylogenetic tree of these FLS sequences was generated for comparative analysis across the diverse species using maximum likelihood method with 100 bootstrap replications. The motifs conserved in each of the FLS protein sequences of cowpea were identified by the tool MEME (available at <https://meme-suite.org/meme/tools/meme>) by putting 10 as default number of motifs and Pfam database (<https://pfam.xfam.org/>) was subjected for analysing their annotations. The identification of the introns, exons, and UTR sections in each cowpea FLS gene, was done by the Gene Structure Display Server.

3.5 Chromosome localization, gene duplication and Synteny analysis

The information regarding the number of chromosomes, length and position of each *VuFLS* gene was obtained from a Phytozome v 13 database. The Phenogram plot was used for visualising the location of each *VuFLS* gene on the chromosome. The following criteria were used to compile the genes: The 80% of the longer gene was covered by the aligned coding nucleotide sequences and > 80% of the aligned sequences were identical. Tandemly duplicated genes are those that are physically separated by less than 50 Kilobases (kb) [33]. The general feature format (GFF) files and genome fasta files of the *V. unguilata* and *G. max* were taken from the phytozome. Then these files were put into the OneStep MCScan in TBtools for synteny analysis. The orthologous FLS genes located on syntenic chromosomal blocks were shown using the dual syteny visualisation in TBtools as shown in Fig. 4. and Fig. 5.

3.6 Promoter analysis and sub-cellular localisation

The PlantCARE database, was used for examining the *cis*-acting regulatory elements present in the promoter region of the identified FLS genes which regulated their expression. Different varieties of elements, such as, light-responsive, site-specific elements hormone-responsive, developmental-related, and promoter-related were discovered.

In the *VuFLS* promoter regions, the study explored an entire of 769 *cis*-regulatory elements. It includes 27 *cis*-acting elements, 4 AT-rich DNA binding protein (ABTP-1)-binding elements, 41 components of a conserved DNA module, 184 common *cis*-acting elements in promoter and enhancer regions, and 3, respectively, three 60K protein binding sites in promoter regions of *VuFLS* genes (Fig. 7). The predicted subcellular localization indicates that *VuFLS* proteins are present in almost all organelles. However, most of them (out of 17 *VuFLS*) are localized to the specific organelles. All of them (except *VuFLS*-5) are located in cytoplasm, all of them (except *VuFLS*-9 and *VuFLS*-11) are present in nucleus, while some of them are also located in mitochondria, plasma membrane and ER while very few are present in Golgi.

3.7 Gene ontology

All the discovered *VuFLS* genes translated protein sequences were annotated. The proteins were functionally grouped into three ontological groups using the BGI WEGO [34] and OmicsBox/Blast2GO tools: Cellular processes, biological processes, and molecular functions (Fig. 9). In biological process, proteins played major roles in biosynthetic and catabolic process, methylation, cellular metabolic process, response to abiotic, biotic and chemical stimulus and single-organism metabolic processes. Proteins were categorised as being involved in oxidoreductase activity, transferase activity, ion binding, carbohydrate, and small molecule binding under the heading of molecular function. Under cellular component, none of proteins of all the 17 identified *VuFLS* genes are involved in any kind of cellular process.

4. Discussion

Study on genes involved for encoding the enzyme protein involved in the synthesis of medicinally important has been of interest for many researchers. Pandey et al. (2016) [35] studied the genes involved in flavonoid biosynthesis pathway in banana. As the primary enzyme involved in flavonoid pathway, FLS plays a key function, and various FLS genes in high plants have undergone genetic, enzymatic and chemical, analysis [36]. Although, genome-wide analyses of various gene families have been carried out in many plants as well as other organisms. However, sequence information of FLS as well as the genome level identification and characterization of FLS in *M. pinnata* has not been studied yet. In the present study, isolation of total RNA, preparation of cDNA, PCR based amplification of complete ORF FLS gene, Sanger sequencing, sequence analysis was carried and analysed sequence has been submitted in the NCBI with the accession no. OM469017. The thorough analysis of the ORF FLS sequence reveals that this sequence shows 85% similarity with *V. unguilata* FLS gene. Hence, the genome wide study of the FLS gene family was done by using *V. unguilata* V1.2 genome available from Phytozome v 13 database. According to Owens (2008) [37], five sequences having maximum resemblance to FLAVONOL SYNTHASE1 (*AtFLS1*), a previously identified FLS gene that is essential for flavonoid metabolism, are found in the *Arabidopsis* (*Arabidopsis thaliana*) genome. Presence of the Fe(II)-dependent dioxygenases (2 ODDs) conserved domain (Fig. 1) in FLS gene family confirms that this helps in synthesis of flavones and also involved in a variety of critical metabolic pathways at various stages of plant development [38]. Chromosomal localization of 6 *A. thaliana* genes *AtFLS2*, -3, -4, and -5 are on 5 number of chromosomes arranged in a tandem array. In present study out of 17 FLS genes, 3 genes were found on chromosome number-1 and 8, while 2 genes on chromosomes 6 and 9 and all other chromosomes carry single gene each as shown in Fig. 6. An evolutionary study, expression, and genome wide analysis of ovate family proteins in cotton (*Gossypium* spp.) were studied by Zhang et al (2022) [39].

According to Stracke (2009) [40], *VuFLS* may be involved in controlling development and growth of plants, synthesis of flavonoids and many types of environmental challenges. In *A. thaliana* FLS1 promoter two *cis*-acting elements [namely MYB-recognition element (MRE) that bound by R2R3-MYB proteins and ACGT-containing element (ACE)] have been identified [41]. Also, the evolutionary study on genome scale of HSP 70 proteins and its role in response to drought and heat stress was studied [42]. However, the present findings imply that most *VuFLS*s play a significant role in flavonoid synthesis (Fig. 7).

Physico-chemical characteristics of each *VuFLS* gene and their encoded protein (such as the molecular weight, protein length, coding sequence, relevant chromosome having the start and end points of each gene, forward and reverse types of strands, GRAVY, PI and predicted subcellular location) has been studied given in Table-1. The subcellular localization study on *VuFLS* genes revealed that the cytoplasm contains the most genes, followed by mitochondria, the nucleus, chloroplasts, and others (Fig. 8). Although, previous findings stated that in *Arabidopsis*, flavonoids are produced and localised to the cytoplasm, while *MaFLS* appears to be located to the cell periphery and cytoplasm but not in the nucleus [43]. However, numerous flavonoid metabolic enzymes, including CHS, CHI, and FLS [44] have been found to have dual cytoplasmic/nuclear distribution. Such differences in *VuFLS* gene localisation suggest that FLS function has become more diverse with the biological processes because they are controlled by these genes. Based on the gene structure analysis, phylogeny study, and motif studies of the cowpea FLS proteins, we found that exon, intron, and UTR regions are same for the members of the same subgroup. But they vary with different sub-groups. These findings suggested that changes in the genes encoding the various *VuFLS* sub-groups may have been influenced during evolution processes including the acquisition or removal of introns. The quantity, order and type of the motifs in proteins of the same group were also comparable, but they vary from those in proteins of other groups.

The purpose of gene ontology was to describe genes and protein functions. By this database, genes can be divided in three categories like cellular, biological and molecular functions. As the result depicted (Fig. 9) that out of 17 genes, 12 of them are involved in biological processes showing functions catabolic and biosynthetic process, response to abiotic, biotic and chemical stimulus, methylation, cellular and primary cellular metabolic processes. Whereas, total 5 genes show molecular functions like carbohydrate binding, transferase and oxidoreductase activity, small molecule and ion binding. However, there is no chemical function shown by any of the genes listed above. But many studies showed various biological, chemical, and molecular functions of flavonol synthase genes [45]. Synteny analysis revealed the presence of four conserved regions/genes (*VuFLS12* to *VuFLS15*) on cowpea chromosome number 1 followed by 3 (*VuFLS9* to 11) on chromosome number 3, two (*VuFLS1*, *VuFLS2*) on chromosome number 6 and (*VuFLS5*, *VuFLS6*) on chromosome number 9 rest one on each chromosome. The comparative synteny study revealed that *VuFLS* genes share more than 80% identity with cowpea and soybean

delineating their orthologous relation within monocot plants indicating presence of similar FLS domains (Fig. 4 and Fig. 5). Since orthologous genes perform similar functions, the results also suggest analogous role of FLS genes in cowpea as those in the other selected dicots.

Conclusion

This is the first study depicting the isolation, amplification, sequencing of flavonol synthase gene in *M. pinnata* (*MpFLS*). The sequence analysis showed very much similarity with *V. unguilata* flavonol synthase (*VuFLS*). Therefore, *Vigna unguilata* v1.2 genome was taken for further genome scale characterization of *MpFLS*. Total 17 FLS genes that were localized on 10 chromosomes were highly conserved. Further, characterization of these genes may give fruitful evidence for *MpFLS* genes and cloning and expression purpose. Insight into the evolutionary history of FLS genes in cowpea was gained by chromosomal analysis and phylogenetic analysis. The FLS gene family was discovered to have undergone evolution, by phylogenetic comparison and synteny analysis. The distribution of genes throughout the genome, as well as their arrangement and structure, indicates a complicated evolutionary history for the selected FLS gene family. For further studies, the cloning and expression of the flavonol synthase, and its response to environmental stress needs to be experimentally validated in *M. pinnata*. Also, additional investigations on functional characterization of the cowpea FLS gene family may also be carried thoroughly.

Declarations

FUNDING

This work was supported by DST/INSPIRE Fellowship (as Junior Research Fellowship) awarded to GK by GOVERNMENT OF INDIA, MINISTRY OF SCIENCE and TECHNOLOGY, Department of Science and Technology Bhawan, New Mehrauli Road, New Delhi-110016, India having Grant No. [IF180386].

Conflict of interest: - The authors declare no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Krentz, A.J., Bailey, C.J., 2005. Oral Antidiabetic Agents. *Drugs* 65, 385–411. doi.org/10.2165/00003495-200565030-00005.
2. Gupta, Y.K. and Briyal, S., 2004. Animal models of cerebral ischemia for evaluation of drugs. *Indian J PhysiolPharmacol*, 48, 379-394.
3. Krishnamurthi A., 1969. The Wealth of India, vol. VIII. Publication and Information Directorate CSIR, New Delhi, India.
4. Abdullah, M., X. Cheng., A. Shakoar, Y. Cao., Y. Lin., Y. Cai and J. Shangao., 2018. New opinion of sugar and light crosstalk in the induction of anthocyanins biosynthesis in fruits. *Intl. J. Agric. Biol.*, 20: 2465–2474. doi.org/10.17957/IJAB/15.0790.
5. Stracke, R., De Vos, R.C., Bartelniewoehner, L., Ishihara, H., Sagasser, M., Martens, S. and Weisshaar, B., 2009. Metabolomic and genetic analyses of flavonol synthesis in *Arabidopsis thaliana* support the *in vivo* involvement of leucoanthocyanidin dioxygenase. *Planta*, 229(2), 427-445. doi.org/10.1007/s00425-008-0841-y.
6. Akita, Y., S. Kitamura., R. Mikami and H. Ishizaka., 2018. Identification of functional flavonol synthase genes from fragrant wild cyclamen. *J. Plant Biochem. Biotechnol.*, 27: 147–155. doi.org/10.1007/s13562-017-0423-9.
7. Takahashi, R., S.M. Githiri, K. Hatayama, E.G. Dubouzet, N. Shimada, T. Aoki, S. Ayabe, T. Iwashina, K. Toda and H. Matsumura, 2007. A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color. *Plant Mol. Biol.*, 63, 125–135. doi.org/10.1007/s11103-006-9077-z.
8. Ferreyra, M.L.F., S. Rius, J. Emiliani, L. Pourcel, A. Feller, K. Morohashi, P. Casati and E. Grotewold., 2010. Cloning and characterization of a UV-B-inducible maize flavonol synthase. *Plant J.*, 62: 77–91. doi.org/10.1111/j.1365-313X.2010.04133. x.
9. Preuß, A., R. Stracke., B. Weisshaar., A. Hillebrecht., U. Matern and S. Martens., 2009. *Arabidopsis thaliana* expresses a second functional flavonol synthase. *FEBS Lett.*, 583, 1981–1986. doi.org/10.1016/j.febslet.2009.05.006.
10. Kim, B.G., E.J. Joe and J.H. Ahn., 2010. Molecular characterization of flavonol synthase from poplar and its application to the synthesis of 3-O-methylkaempferol. *Biotechnol. Lett.*, 32, 579–584. doi.org/10.1007/s10529-009-0188-x.
11. Chandran, A.K.N., Bhatnagar, N., Kim, B. and Jung, K.H., 2016. Genome-wide identification and analysis of rice genes to elucidate morphological agronomic traits. *Journal of Plant Biology*, 59(6), 639-647. doi.org/10.1007/s12374-016-0407-7.
12. R. Scott-Moncrie, 1936. Biosynthesis of Flavonoids. *J. Genet.*, 32,117.
13. Britsch, L., Heller, W. and Grisebach, H., 1981. Conversion of flavanone to flavone, dihydroflavonol and flavonol with an enzyme system from cell cultures of parsley. *Z. Naturforsch.*, 36, 742-750. doi.org/10.1515/znc-1981-9-1009.

14. Oksana, S., Marian, B., Mahendra, R. and Bo, S.H., 2012. Plant phenolic compounds for food, pharmaceutical and cosmetics production. *J. Med. Plant Res.* 6(13), 2526-2539. doi.org/10.5897/JMPR11.1695.
15. T. A. Holton., F. Brugliera, and Y. Tanaka., 1993. Biosynthesis of Flavonoids. *Plant J.*, 4, 1003-1010. [doi: 10.3389/fpls.2012.00222](https://doi.org/10.3389/fpls.2012.00222).
16. Xu, F., Li, L., Zhang, W., Cheng, H., Sun, N., Cheng, S. and Wang, Y., 2012. Isolation, characterization, and function analysis of a flavonol synthase gene from *Ginkgo biloba*. *Molecular Biology Reports*, 39(3), pp.2285-2296. [DOI:10.1007/s11033-011-0978-9](https://doi.org/10.1007/s11033-011-0978-9).
17. Al Muqarrabun, L.M.R., Ahmat, N., Ruzaina, S.A.S., Ismail, N.H. and Sahidin, I., 2013. Medicinal uses, phytochemistry and pharmacology of *Pongamia pinnata* (L.) Pierre: A review. *Journal of ethnopharmacology*, 150(2), pp.395-420. [doi: 10.1016/j.jep.2013.08.041](https://doi.org/10.1016/j.jep.2013.08.041).
18. Kumari G., Nigam V.K. and Pandey D.M., 2022. The molecular docking and molecular dynamics study of flavonol synthase and flavonoid 3'-monooxygenase enzymes involved for the enrichment of kaempferol. *Journal of Biomolecular Structure and Dynamics*, pp.1-14. doi.org/10.1080/07391102.2022.2033324.
19. Rio, D.C., Ares, M., Hannon, G.J. and Nilsen, T.W., 2010. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harbor Protocols*, 2010(6), pp.pdb-prot5439. [doi: 10.1101/pdb.prot5439](https://doi.org/10.1101/pdb.prot5439).
20. Sambrook, J., Fritsch, E.F. and Maniatis, T., 1989. *Molecular cloning: a laboratory manual* (No. Ed. 2). Cold spring harbor laboratory press.
21. Koichiro Tamura, Glen Stecher, and Sudhir Kumar (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution* 38:3022-3027. [doi: 10.1093/molbev/msab120](https://doi.org/10.1093/molbev/msab120).
22. Goodstein DM., Shu S., Howson R., Neupane R., Hayes RD., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar DS., 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*, 40, 1178–1186. [doi.org/ 10.1093/nar/gkr944](https://doi.org/10.1093/nar/gkr944).
23. David M. Goodstein, Shengqiang Shu, Russell Howson, Rochak Neupane, Richard D. Hayes, Joni Fazo, Therese Mitros, William Dirks, UffeHellsten, Nicholas Putnam, and Daniel S. Rokhsar, Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Res.* 2012 40 (D1): D1178-D1186. [doi: 10.1093/nar/gkr944](https://doi.org/10.1093/nar/gkr944).
24. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press. 571-607. [doi.org 10.1385/1-59259-890-0:571](https://doi.org/10.1385/1-59259-890-0:571).
25. Chen C., Chen H., He Y., Xia R., 2018. TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. *BioRxiv*. [doi.org/ 10.1101/289660](https://doi.org/10.1101/289660).
26. Kumar S., Stecher G., Li M., Knyaz C., and Tamura K., 2018. *Molecular Biology and Evolution* 35:1547-1549. [doi: 10.1093/molbev/msy096](https://doi.org/10.1093/molbev/msy096).
27. Bailey TL., Williams N., Misleh C., Li WW., 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.*, 34: 369–373. doi.org/10.1093/nar/gkl198.
28. Punta M, Coghill PC., Eberhardt RY., Mistry J., Tate J., Boursnell C., Pang N., Forslund K., Ceric G., Clements J., Heger A., Holm L., Sonnhammer EL., Eddy SR., Bateman A., Finn RD., 2012. The Pfam protein families database. *Nucleic Acids Res.*, 40, D290–D301. [doi.org/ 10.1093/nar/gkr1065](https://doi.org/10.1093/nar/gkr1065).
29. Bo Hu, Jinpu Jin, An-Yuan Guo, He Zhang, Jingchu Luo and Ge Gao. 2015. GSDS 2.0: an upgraded gene features visualization server. doi.org/10.1093/bioinformatics/btu817.
30. Simpson MG., 2010. *Phylogenetic systematics*, 2nd edn. Academic Press, Cambridge. *Plant systematics*.
31. El Sappah A., Abbas M., Elyrs AS., Yadav V., El-Sappah HH., Zhu Y., Huang Q., Yu W., Soaud SA., Xianming Z., 2021. The Hsp70 gene family in *Solanum lycopersicum*; genome-wide identification and expression analysis under heavy metals stresses. doi.org/10.21203/rs.3.rs-546628/v1.
32. Cannon SB., Mitra A., Baumgarten A., Young ND., May G., 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.*, 4(1): 10. doi.org/10.1186/1471-2229-4-10.
33. Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., Xu, H., Huang, X., Li, S., Zhou, A. and Zhang, X., 2018. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic acids research*, 46(W1), W71-W75. [doi: 10.1093/nar/gky400](https://doi.org/10.1093/nar/gky400).
34. Pandey, A., Alok, A., Lakhwani, D. et al., 2016. Genome-wide Expression Analysis and Metabolite Profiling Elucidate Transcriptional Regulation of Flavonoid Biosynthesis and Modulation under Abiotic Stresses in Banana. *Sci Rep* 6, 31361. <https://doi.org/10.1038/srep31361>.
35. Nguyen, N.H., J.H. Kim, J. Kwon, C.Y. Jeong, W. Lee, D. Lee, S. Hong and H. Lee, 2016. Characterization of *Arabidopsis thaliana* flavonol synthase 1 (FLS1)- overexpression plants in response to abiotic stress. *Plant Physiol. Biochem.*, 103: 133–142. doi.org/10.1016/j.plaphy.2016.03.010.
36. Owens, D.K., Alerding, A.B., Crosby, K.C., Bandara, A.B., Westwood, J.H. and Winkel, B.S., 2008. Functional Analysis of a Predicted Flavonol Synthase Gene Family in *Arabidopsis*. *Plant Physiol.* 147 (3), 1046–1061. doi.org/10.1104/pp.108.117457.

37. Wei, S., Zhang, W., Fu, R. 2021. Genome-wide characterization of 2-oxoglutarate and Fe (II)-dependent dioxygenase family genes in tomato during growth cycle and their roles in metabolism. BMC Genomics 22,126. doi.org/10.1186/s12864-021-07434-3.
38. Zhang, L., Azhar, M.T., Che, J. and Shang, H., 2022. Genome-wide identification, expression and evolution analysis of OVATE family proteins in cotton (*Gossypium* spp.). Gene, 146653. doi.org/10.1016/j.gene.2022.146653.
39. Stracke, R., De Vos, R.C., Bartelniewoehner, L., Ishihara, H., Sagasser, M., Martens, S. and Weisshaar, B., 2009. Metabolomic and genetic analyses of flavonol synthesis in *Arabidopsis thaliana* support the *in vivo* involvement of leucoanthocyanidin dioxygenase. Planta, 229(2), 427-445. doi.org/10.1007/s00425-008-0841-y.
40. Hartmann U., Sagasser M., Mehrtens F., Stracke R., Weisshaar B., 2005. Di Verential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol Biol, 57, 155–171. doi.org/10.1007/s11103-004-6910-0.
41. Ahmad, M.Z., Li, P., She, G., Xia, E., Benedito, V.A., Wan, X.C. and Zhao, J., 2020. Genome-wide analysis of serine carboxypeptidase-like acyltransferase gene family for evolution and characterization of enzymes involved in the biosynthesis of galloylated catechins in the tea plant (*Camellia sinensis*). Front. Plant Sci., 11, 848. doi.org/10.3389/fpls.2020.00848.
42. Saslowsky D., Warek U., Winkel B., 2005. Nuclear localization of flavonoid enzymes in *Arabidopsis*. J. Biol. Chem. 280, 23735–23740. doi: 10.1074/jbc.M413506200.
43. Kuhn B., Geisler M., Bigler L., Ring C., 2011. Flavonols accumulate asymmetrically and affect auxin transport in *Arabidopsis*. Plant Physiol. 156, 585–595. doi.org/10.1104/pp.111.175976.
44. Fang, H., Qi, X., Li, Y., 2020. De novo transcriptomic analysis of light-induced flavonoid pathway, transcription factors in the flower buds of *Lonicera japonica*. Trees 34, 267–283 <https://doi.org/10.1007/s00468-019-01916-4>.

Tables

Table-1: The table depicting the physico-chemical properties of 17 *VuFLS* genes.

Transcript ID	Gene Name	Chromosome	Location start-End	Strand	CDS(bp)	Protein length (aa)	Protein Molecular weight (kDa)	pI	GRAVY
Vigun06g132700	<i>VuFLS-1</i>	6	25865687..25870712	Forward	987	328	36617.17	6.75	-0.1
Vigun06g132700	<i>VuFLS-2</i>	6	25865687..25870712	Forward	981	326	36426.96	6.75	-0.106
Vigun04g120100	<i>VuFLS-3</i>	4	30769279..30770936	Reverse	1089	362	40001.68	5.52	-0.277
Vigun08g121000	<i>VuFLS-4</i>	8	28932686..28935314	Forward	1038	345	39314.44	6.3	-0.507
Vigun05g116600	<i>VuFLS-5</i>	5	12409405..12412921	Forward	1182	393	44604.8	5.49	-0.47
Vigun10g186500	<i>VuFLS-6</i>	10	40279975..40282198	Forward	1113	370	41232.23	8.33	-0.184
Vigun09g016600	<i>VuFLS-7</i>	9	1212595..1215384	Reverse	1077	358	40979.04	6.57	-0.473
Vigun09g073700	<i>VuFLS-8</i>	9	8191868..8193683	Reverse	558	185	21086.15	7.83	-0.405
Vigun03g128100	<i>VuFLS-9</i>	3	12374708..12378021	Forward	1086	383	43048.11	5.71	-0.332
Vigun03g128100	<i>VuFLS-10</i>	3	12374708..12378021	Forward	1152	361	40267.89	5.69	-0.309
Vigun03g410500	<i>VuFLS-11</i>	3	61837958..61840763	Reverse	972	323	35651.14	8.19	-0.109
Vigun01g199300	<i>VuFLS-12</i>	1	37580493..37585301	Forward	1023	340	38140.74	7.06	-0.252
Vigun01g199300	<i>VuFLS-13</i>	1	37580493..37585302	Forward	1023	340	38140.74	7.06	-0.252
Vigun01g199300	<i>VuFLS-14</i>	1	37580493..37585303	Forward	1023	340	38140.74	7.06	-0.252
Vigun01g199300	<i>VuFLS-15</i>	1	37580493..37585304	Forward	1017	338	37892.42	6.67	-0.258
Vigun02g013700	<i>VuFLS-16</i>	2	5101550..5104851	Reverse	1086	361	40536.46	5.35	-0.254
Vigun07g046500	<i>VuFLS-17</i>	7	4776776..4780303	Forward	1059	352	40197.06	5.97	-0.329

Figures

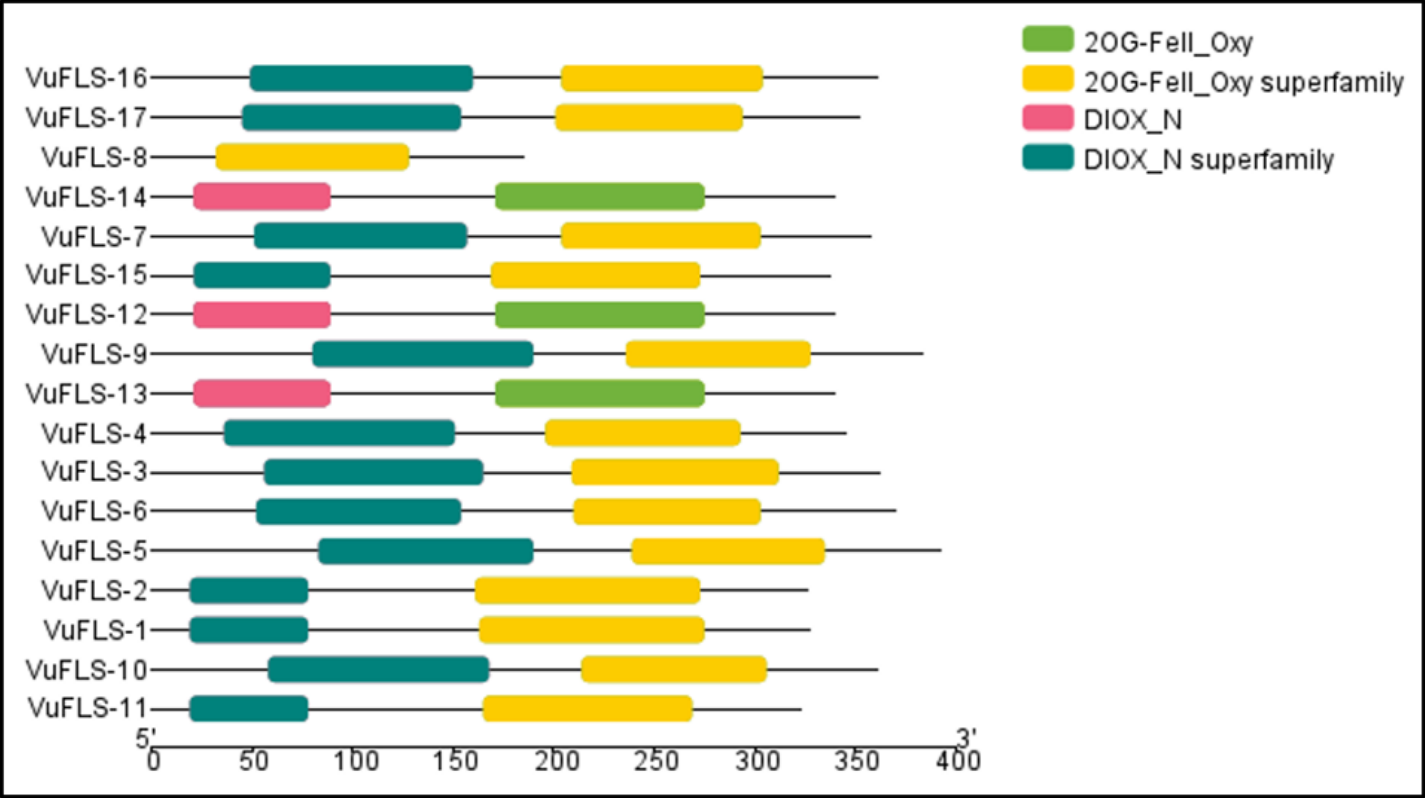


Figure 1

The conserved domain analysis of all *VuFLS* gene family members

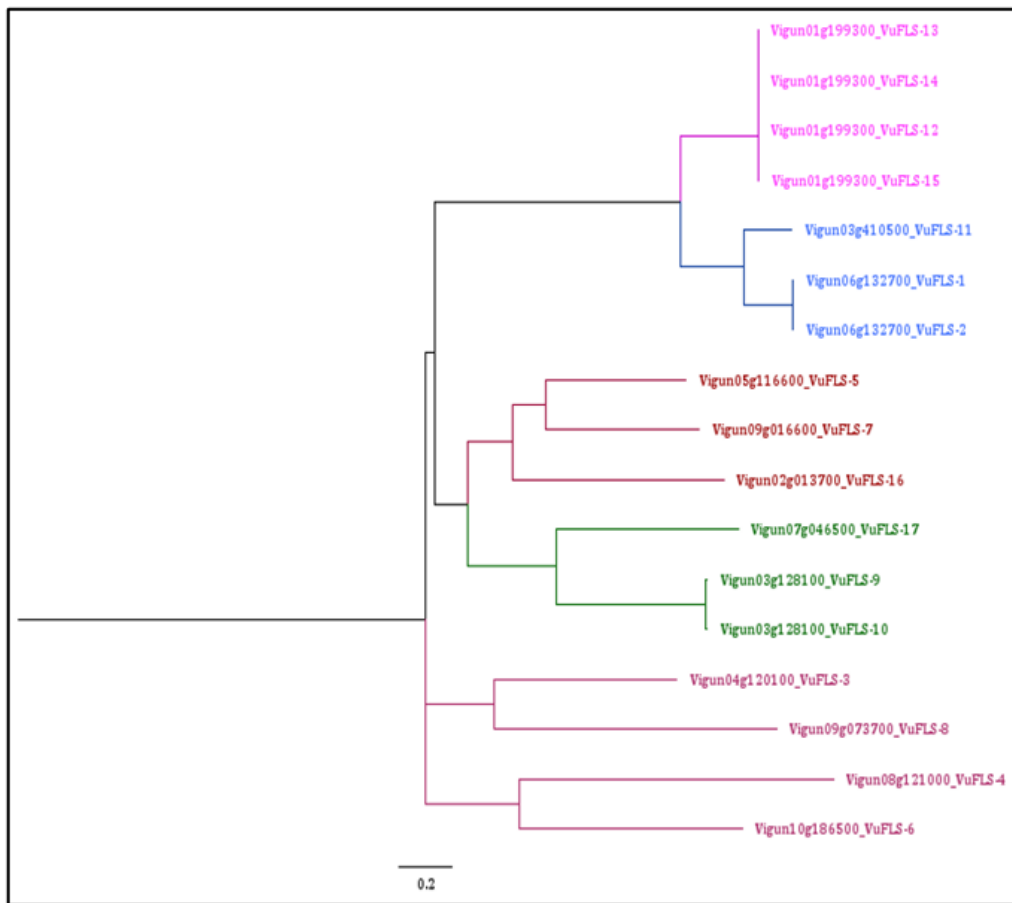


Figure 2

The Phylogenetic tree of the VuFLS gene family constructed by Maximum Likelihood method and JTT matrix-based model using MEGA X (Molecular Evolutionary Genetics Analysis) software.

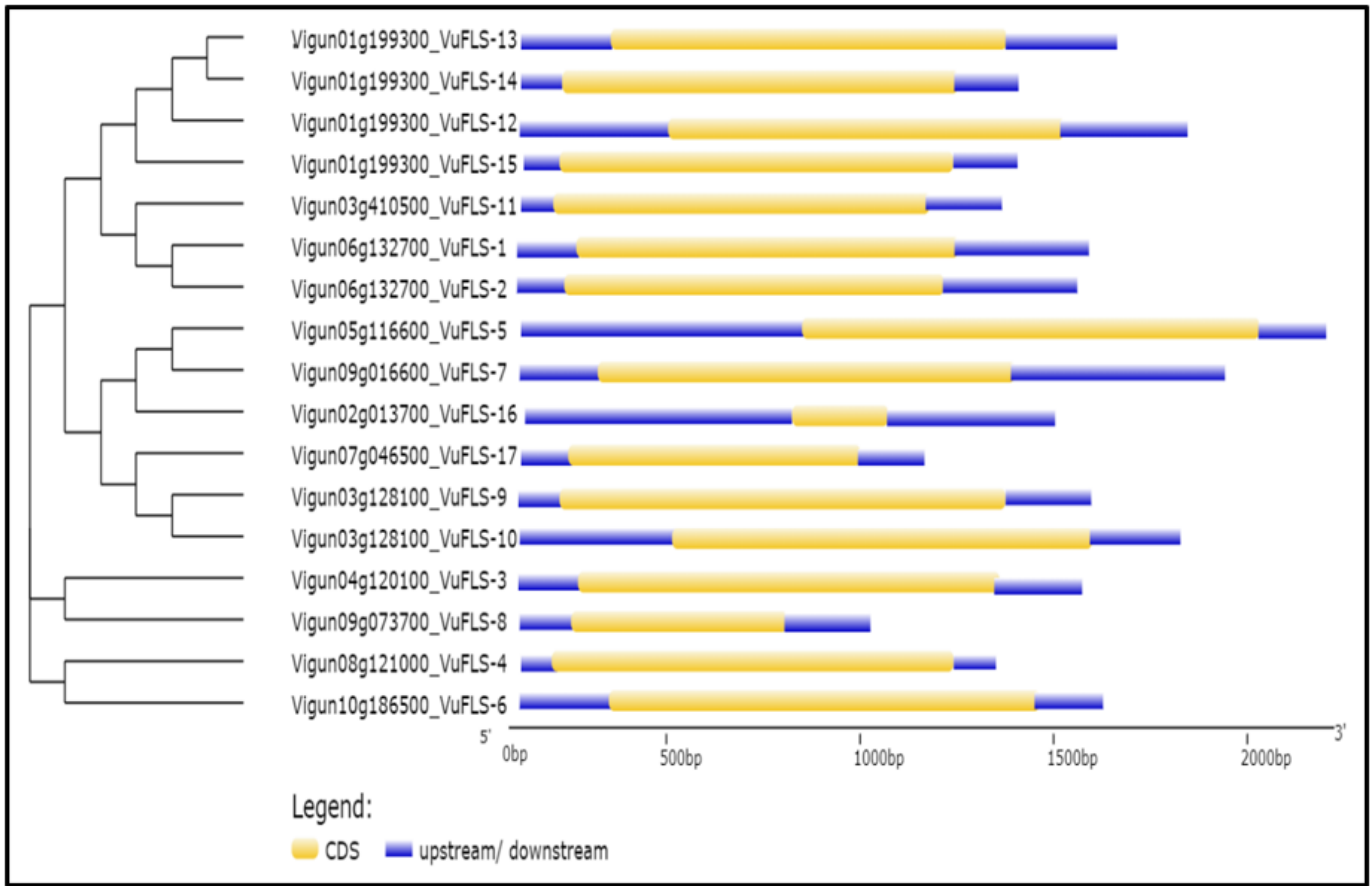


Figure 3

The exon-intron organization of *VuFLS* genes plotted using Gene Structure Display Server database where blue box in two terminals represent 5'-UTR and 3'-UTR, yellow box indicates exons.

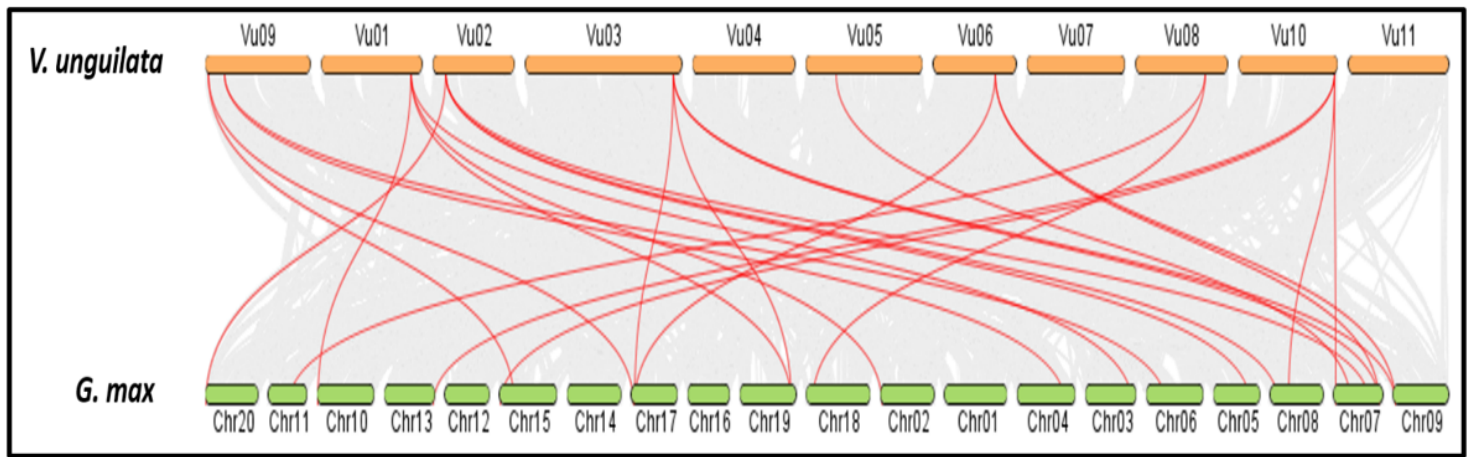


Figure 4

The Synteny study of *FLS* genes between *V. unguilata* and *G. max*. The collinear blocks between two genomes are represented by the grey lines in the backdrop. The syntenic areas between the *V. unguilata* and *G. max* *FLS* gene pairs are highlighted by the blue lines.

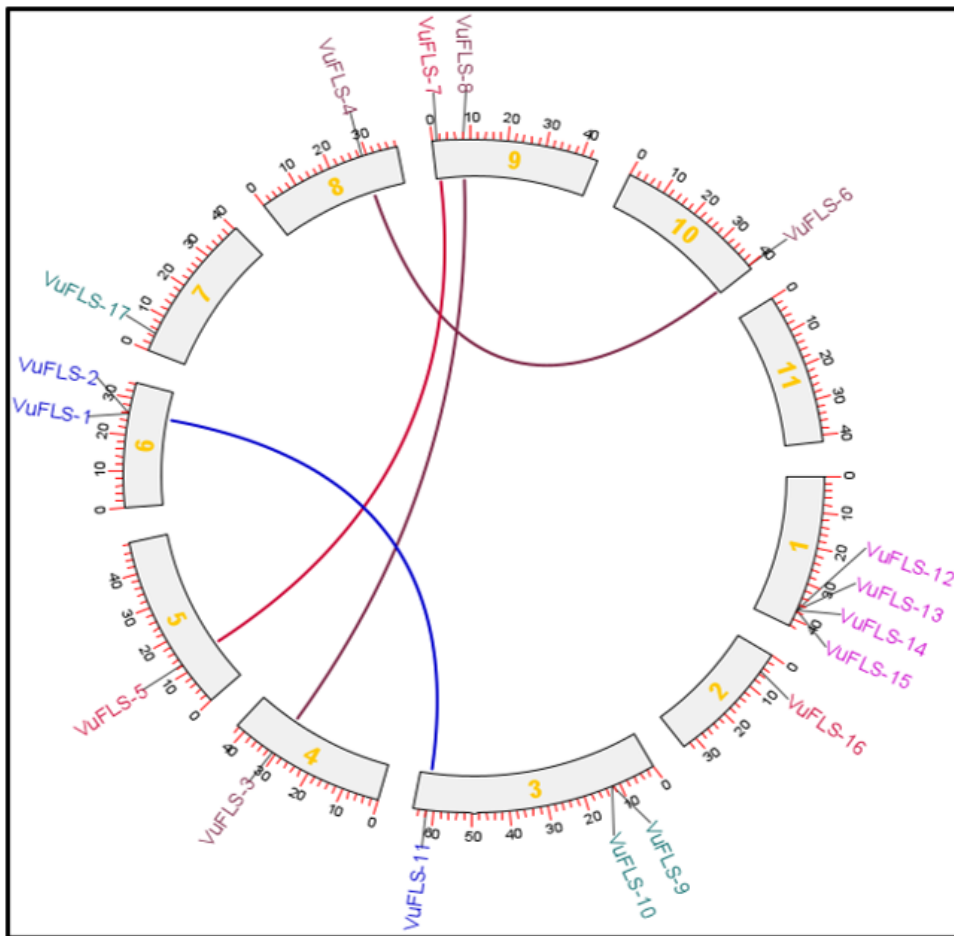


Figure 5

The circos visualization map showing synteny, collinearity and duplication of *VuFLS* genes and their chromosomal location.

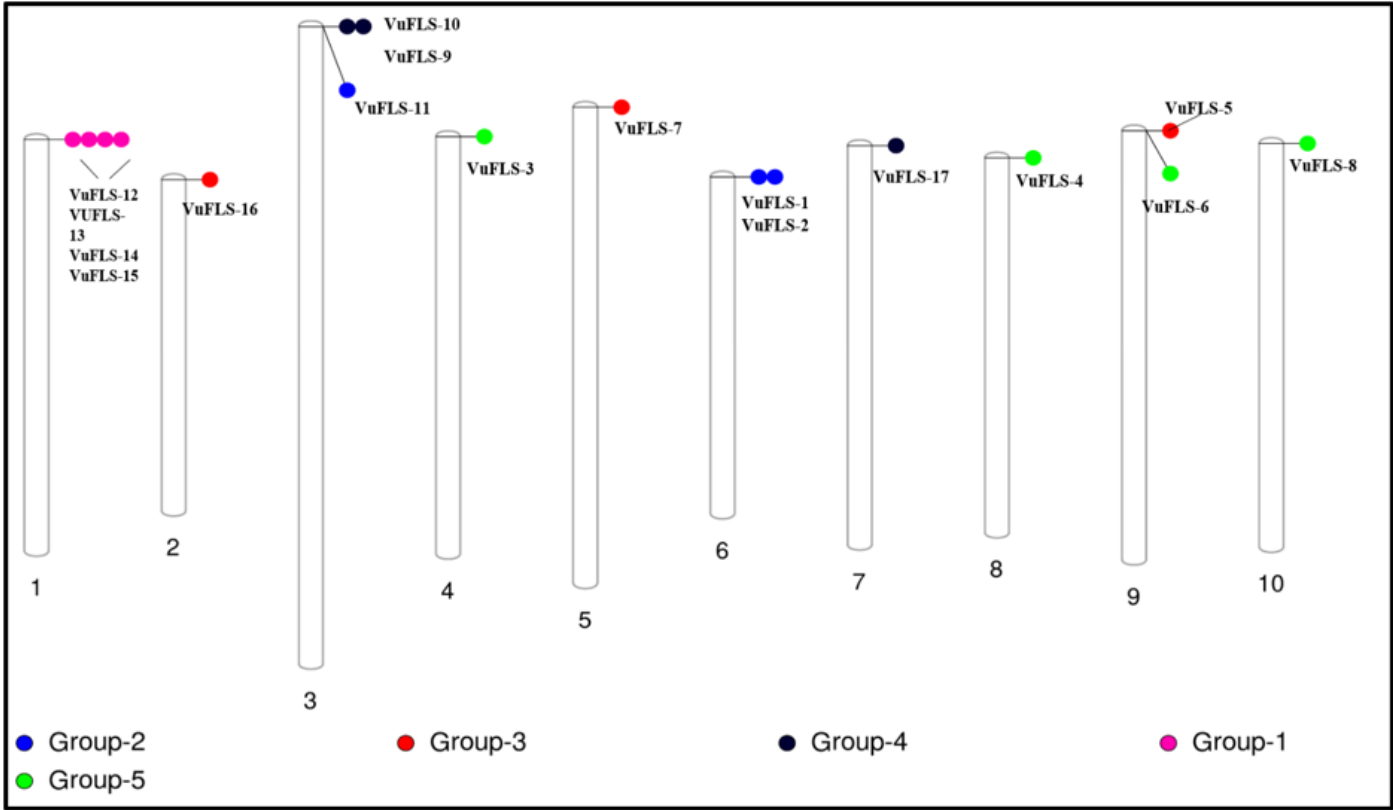


Figure 6

The figure showing the chromosomal localization of the *VuFLS* genes, the number of chromosomes is indicated below each bar. The various coloured circles reflect the genes on each chromosome that belong to the particular categories.

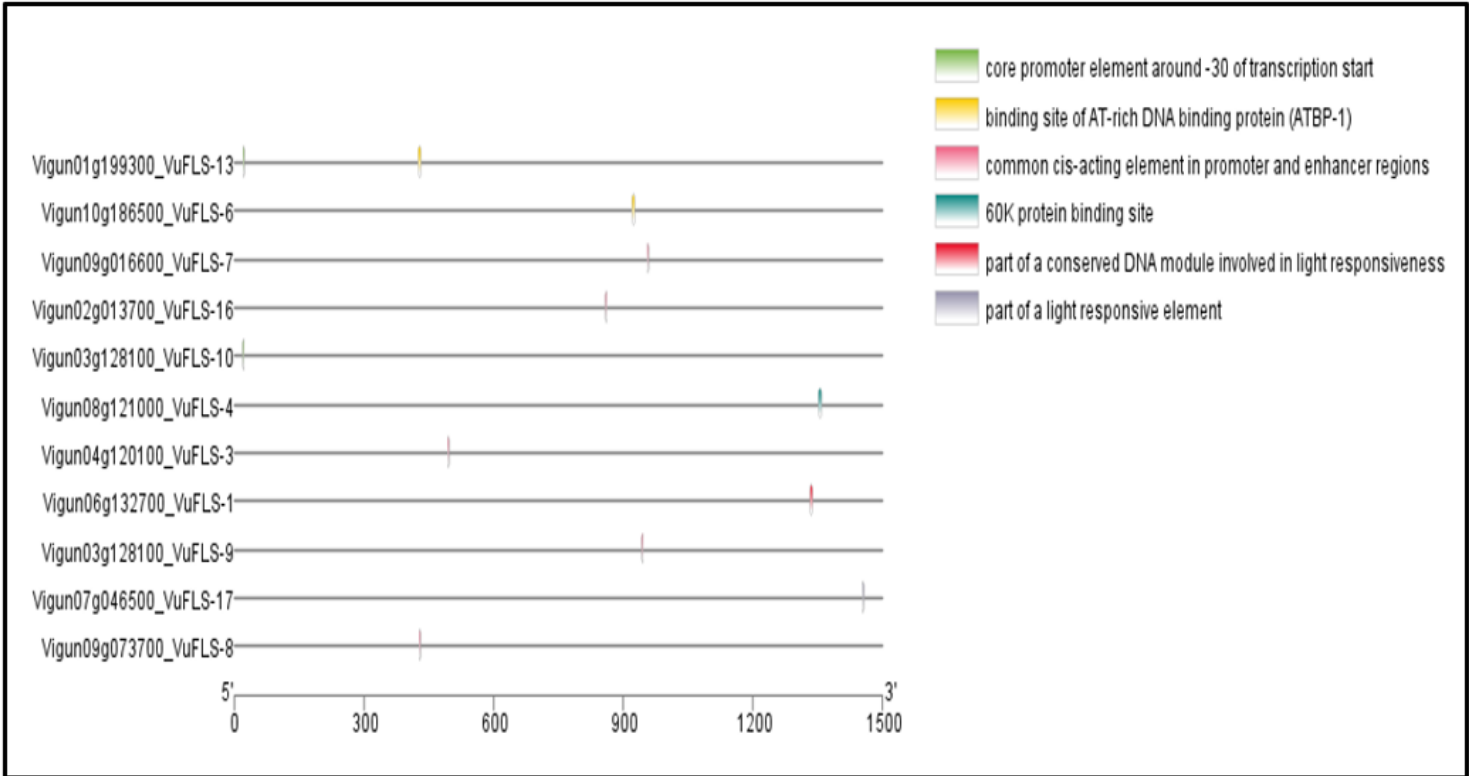


Figure 7

The figure shows the *cis*-regulatory element analysis of the *VuFLS* family genes, PlantCare database was used to examine the 1500 bp upstream region of each of the *VuFLS* genes. The core promoter elements, ATBP-1, *cis*-acting element in promoter and enhancer regions, protein binding sites and light responsive elements, in *VuFLS* family members.

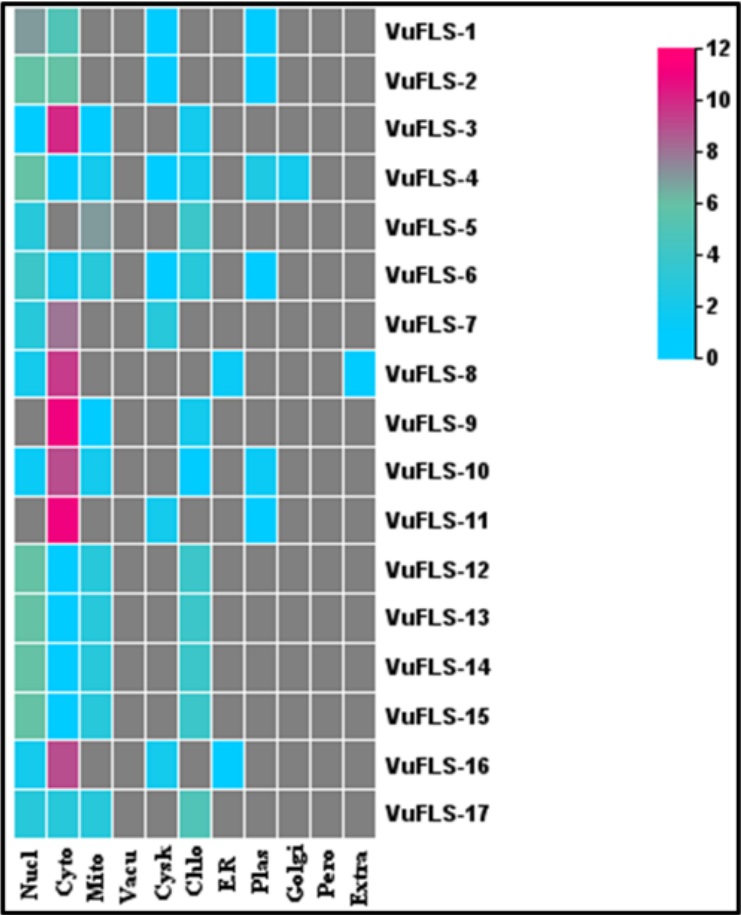


Figure 8

The figure showing the subcellular localisation of the *VuFLS* genes using cello online tool.

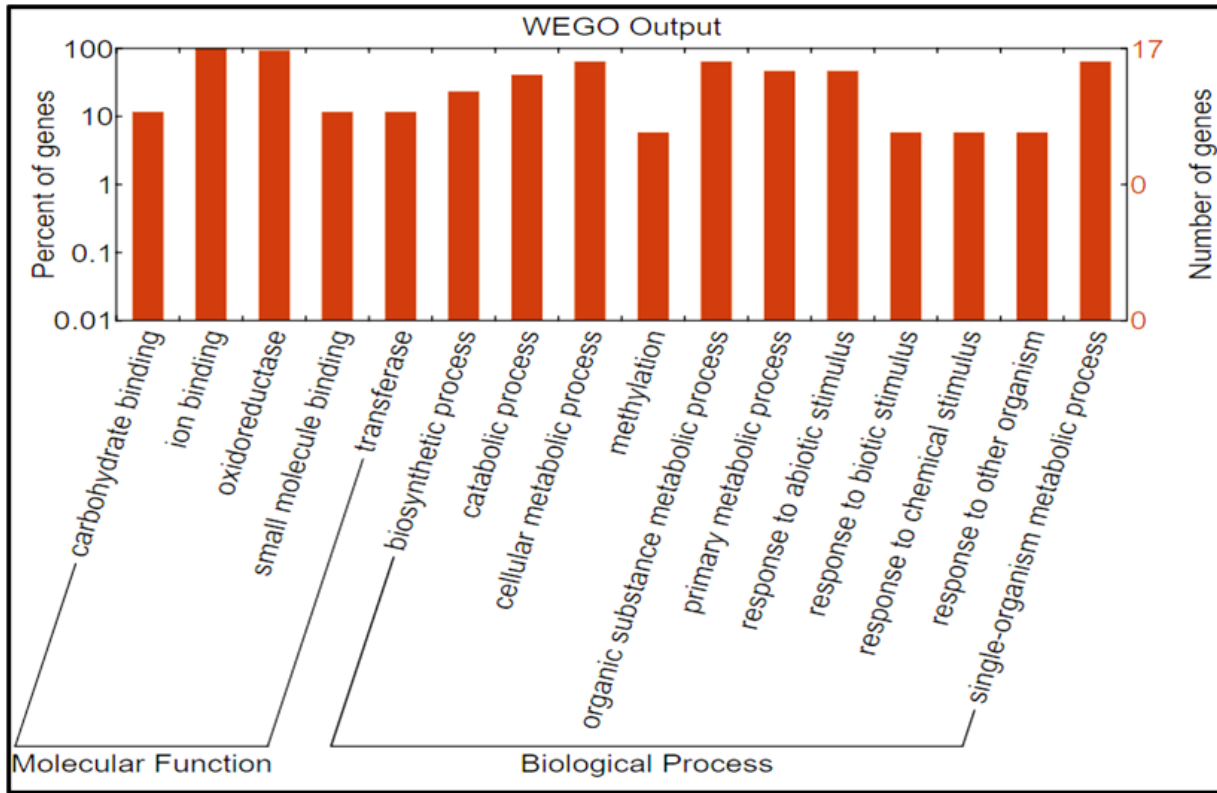


Figure 9

The Gene Ontological classification of identified *VuFLS* proteins a histogram representation using BGI WEGO into Molecular function and biological processes.

Supplementary Files

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

- [Supplementaryfile.docx](#)