**Identification of epi-transcriptomic methylation marker genes in soybeans (*Glycine max*), and their expression profiling in various organs and environmental stimuli.**

**Abstract**

**Background**: N6-methyladenosine (m6A) is one of the most frequently found mRNA modification in eukaryotes. Without altering the nucleotide sequence, the m6A modification can regulate various biological processes in organisms through the modulation of alternative splicing, alternative polyadenylation, folding, translation, localization, transport, and decay of multiple types of RNA. This dynamic complex has three crucial components namely the writer, eraser, and reader, collaborating by doing addition, subtraction and reading methyl groups, respectively. To date, investigation on m6A regulatory genes in plants remains obscured.

**Results:** In this study, 42 soybeans m6A regulatory genes were identified through a comprehensive genome-wide fashion. The structural analysis uncovered the architectural and functional diversity among different subgroups of soybean m6A regulatory genes. 18 Duplication events were pinpointed which evolved through purified selection. Cluster and hub gene identification revealed a lower strength of interaction of eraser proteins. Based on the interaction between miRNA and soybean m6A, it can be hypothesized that writer components might prefer to be regulated by miRNA than erasers and readers. Furthermore, Expression analysis showed varied responses of m6A gene sets in different stress and infection contexts.

**Conclusions:** This study may form a basement to explore the m6A regulatory genes in soybean. This comprehensive genome-wide approach might be beneficial for better understanding the insights and dynamic mechanisms of soybean m6A genes.

**Introduction**

Plants have always suffered from various biotic and abiotic stresses such as pathogens, nutritional imbalance, salt, light intensity, and drought. To cope with these environmental variations, plants must evolve multiple mechanisms[1]. Epigenetic modification is one of them which is associated with gene expression regulation under stress. Herein, Epigenetics refers to stimuli-triggered changes in gene expression without altering the underlying DNA sequence that can be passed to offspring[2]. This process is mainly involved in DNA/RNA methylation, histone modifications, chromatin remodelling, and noncoding RNAs.

To date, RNA modification is an emerging major epigenetic modification in the plant life cycle and plays a critical role in the regulation of gene expression, both at the transcriptional and post-transcriptional levels[3]. Like DNA and protein, RNA goes through 160 types of post-transcriptional modifications where mRNA modification is found to be less common in comparison to tRNA and rRNA[4][5]. Among mRNA modifications, N6-methyl-adenosine (m6A) is one of the most abundant inner, dynamic, and reversible posttranscriptional modifications in eukaryotes and serves as a novel epigenetic marker that is involved in various biological processes[6]. Furthermore, three enzyme complexes namely methyltransferases (writers), demethylases (erasers), and m6A-binding proteins (readers) are found to be involved in introducing, deleting, and interpreting specific methylation marks on mRNAs[7].

The writer complex first discovered in mammals that include methyltransferase-like 3 (METTL3), methyltransferase-like 14 (METTL14), wilms tumor 1-associated protein (WTAP), RNA binding motif protein 15/15B (RBM15/RBM15B), vir-like m6A methyltransferase associated (VIRMA or KIAA1429), zinc finger CCCH-type containing 13 (ZC3H13), methyl- transferase-like 16 (METTL16), methyltransferase-like 4 (METTL4), methyltransferase-like 5 (METTL5), and zinc finger CCHC-type containing 4 (ZCCHC4)[8]. The orthologs of these complexes have also been identified in Arabidopsis such as MTA (ortholog of METTL3), MTB (ortholog of METTL14), VIR (ortholog of VIRMA) and FIP37(ortholog of WTAP)[9][10][11][12]. Moreover, E3 ubiquitin ligase HAKAI, an additional m6A writer component has been identified in Arabidopsis [13]. METTL3 (S-adenosyl-methionine-binding protein) is the key catalytic component of the m6A methyltransferase complex. METTL14 and WTAP making complex with METTL3 are found to be involved in M6A installation and modification, respectively[14]. Besides this, VIRMA is a regulatory subunit of m6A methyltransferase recruiting the m6A complex to the special RNA site and facilitating m6A installation[15].

Conversely, the m6A demethylase erases the methyl group from the mRNA, indicating the reversibility of m6A modification. Herein, FTO (fat mass and obesity-associated protein) is the first eraser complex discovered in mammals that facilitates in restoration of the methylated base to the adenine base[16]. Subsequently, ALKBH5 (alkB homolog 5), the second mammalian m6A demethylase facilitating the removal of m6A potentially affects different subsets of target mRNAs[17]. Arabidopsis contains several putative m6A eraser ALKBH family proteins[12]. However, ALKBH9B and ALKBH10B are the only two eraser proteins found in Arabidopsis that have been functionally investigated concerning viral infection and floral transition[18].

While writers and erasers are actively engaged in their dynamic action, the reader family regulates mRNA nuclear export, splicing, degradation, translation, and stability. YT521-B homology (YTH) domain family is the first identified mammalian reader complex which includes YTHDF1, YTHDF2, YTHDF3, YTHDC1, and YTHDC2[19]. Different reader proteins have different functions on m6A-modified RNAs. For example, YTHDF2 regulates the degradation[20], YTHDF1 modulates translation[21], YTHDF3 cooperates with YTHDF1 and YTHDF2 thus modulating the translation and degradation[22], the insulin-like growth factor 2 mRNA-binding proteins (IGF2BP1/2/3) promote stability[23], FMRP enhances nuclear export and stability[24], and YTHDC1 modulates nuclear export and splicing[25]. Furthermore, YTHDC2 and Eukaryotic initiation factor 3 (eIF3) are also crucial in mRNA translation[21][26]. Notably, three evolutionarily conserved c-terminal region (ECT) family proteins have recently been functionally characterized in Arabidopsis as YTHD homologs[27][28].

Indeed, M6A modification is a crucial regulator performing three main functions: mRNA processing, plant growth and development, and stress response[29]. Researchers suggest that m6A regulatory writer proteins are essential for shoot and root growth, leaf, cotyledon, and floral development[10]. Partial loss of FIP37 causes huge over-proliferation of shoot meristems. Moreover, depletion of FIP37 results in embryo lethality[30]. The m6A modification associated genes are expressed differentially in response to biotic and abiotic stresses including drought[31], salt[32], cold[33], and bacterial infections[34]. Nevertheless, m6A modifications have been studied in limited plant species, highlighting the need for future investigation in other plants.

Soybean (Glycine max), a Leguminosae family member, has become one of the main economical oilseed beans[35]. Currently, researchers shed light on pinpointing stress-responsive writer genes as well as eraser genes in this plant[36][37]. Nevertheless, comprehensive, or systematic analysis of all the m6A gene families remains scarce. Therefore, we were encouraged to conduct genome-wide identification of writers, erasers, and readers genes of the m6A pathway in Soybeans. Here we identified 12 writers, 11 erasers, and 19 readers of m6A genes and conducted a comprehensive analysis of the classification, evolution, gene structure, and potential interaction network of m6A regulatory genes in soybeans. We also studied gene expression patterns in response to a range of biotic and abiotic stresses.

**2. Method and materials**

**2.1. Genome-Wide Identification of m6A Gene Family in soybean**

To identify m6A components and their protein families in soybeans, the amino acid sequences of m6A-related proteins reported in Arabidopsis thaliana[38], including writers, erasers, and readers, were used as queries to perform BLASTP against the soybean genomic sequences in the plant genomics resource database, phytozome[39]. Putative genes and corresponding protein sequences were retrieved from the same database and the presence of conserved domains was confirmed by NCBI Conserved Domain(<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and Pfam (<http://pfam.xfam.org/>). The identity of the “writer” family was confirmed by the presence of the MT-A70 domain (PF05063), Wtap domain (PF17098), and Vir-N domain (PF15912). The presence of 2-OG Fe (II) oxygenase superfamily (PF13532) and YTH domain (PF04146) confirmed the identity of the “eraser” and “reader” families, respectively. Candidates without the m6A conserved and typical domains were not further considered. The prefix “GM” for glycine max (soybean) was added to all members followed by subclass IDs (MT, FIP37, VIRILIZER, ECT, CPSF, and ALKBH). Chromosomal location, strand position, CDS coordinate (5’ to 3’), gene length, CDS length and protein length were retrieved from the Phytozome database[39]. Furthermore, the Physicochemical properties of the identified proteins such as (molecular weight, and theoretical isoelectric point) were calculated using the ExPASy-ProtParam tool [40]. Subcellular localization of proteins was predicted with BUSCA [41].

**2.2. Chromosomal localization and gene duplication events**

The GFF file of soybean was retrieved from the phytozome database [39]. Chromosomal locations were visualized by TBtool[42]. Using the Plant Genome Duplication Database (PGDD, <http://pdgd.njau.edu.cn:8080/>)[43], paralogous genes were found by calculating the number of duplication events among soybean genes. From the same database, the synonymous substitution rate (Ks) and non-synonymous substitution rate (Ka) were obtained. Using the Ka/Ks ratio, the selection pressure of duplicated genes was computed. Ka/Ks < 1 denotes purifying selection, Ka/Ks = 1 implies neutral drifting, and Ka/Ks > 1 signifies positive or Darwinian selection for a pair of genes[44]. Using T = Ks/ (2\*6.1\*10^-9) \*10^-6 formula, the approximate date (Mya, million years ago) of each duplication event was estimated [45].

**2.3. Gene Structure, Conserved Domain and Conserved Motif Analysis**

Genomic and CDS sequences were retrieved from the phytozome database[39]. The Newick trees of writer, eraser and reader protein family were plotted by MEGA11[46]. Subsequently, the physical mapping of the exon-intron sequence was plotted using GSDS (Gene Structure Display Server 2.0, <http://gsds.gao-lab.org/>). The conserved domains of m6A regulatory protein sequences were analyzed in the Batch CD-Search program[47] and then the output file was submitted to TBtools for visualized analyses[42]. The conserved motifs were analyzed using TBtools with a maximum number of motifs of 10[42].

**2.4. Phylogenetic Analysis**

Protein sequences of writer, eraser and reader of 13 species namely Linum usitatissimum, Acorus americanus, Anacardium occidentale, Aquilegia coerulea, Brachypodium hybridum, Coffea arabica, Gossypium barbadense, Oryza sativa, Panicum hallii, Solanum lycopersicum, Triticum aestivum, Zea mays, and Porphyra umbilicalis were retrieved from the phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>). Multiple sequence alignment was performed using the MUSCLE algorithm[48] and phylogenetic trees for the protein families were constructed using Molecular Evolutionary Genetics Analysis (MEGA-11) software[46]. The maximum likelihood method was employed with node reliability assessed through bootstrap analysis based on 1000 replicates. The resulting phylogenetic trees were further refined and graphically represented using the Interactive Tree of Life (iTOL) platform (<https://itol.embl.de>).

**2.5. Promoter analysis in soybean m6A regulatory genes.**

The 1000 bp upstream sequences from the transcription start site, were collected, and submitted to the PlantCare database[49] for cis-acting element analysis. The presence of identified cis-acting elements on the putative promoter of every gene had been depicted using TBtools[42].

**2.6. Functional enrichment visualization of m6A regulatory genes**

Gene ontology enrichment in three categories such as cellular component (CC), biological process (BP) and molecular function (MF) was executed using the STRING database[50]. FDR (False discovery rate) =<0.05 was set. Number of terms was shown as 10. The results were shown as a bar chart and a bubble chart.

**2.7. Protein-protein interactions, cluster prediction and hub gene identifications**

The STRING v11.5 database[50] was used to construct the protein-protein interaction network among m6A regulatory genes in soybeans. The output file from STRING was visualized by Cytoscape software[51]. The network's co-expressing network clusters (highly linked areas) were examined using the Molecular Complex Detection (MCODE) (v2.0.2) plugin[52]. Hub genes were recognized using the cytoHubba plugin[53] in Cytoscape[51] with shortest path and ranked by degree.

**2.8. Structure Construction by Homology Modeling**

The homology modelling approach was used to predict the three-dimensional (3D) structures of the m6A regulatory proteins. The amino acid sequence was queried against the SWISS-MODEL server (<https://swissmodel.expasy.org/> )[54] to search for templates, and the best templates with a similar amino acid sequence and known three-dimensional structures were used to Build the Models. These proteins' structures were examined with the UCSF ChimeraX [55] visualization tools.

**2.9. micro-RNA targets prediction**

Mature miRNAs were downloaded from the miRBase database[56]. The psRNAT-target database[57] was used to search the regulatory relationship between miRNAs and soybean m6A regulatory genes, with an Expectation threshold of < 5 and other parameters at their default values. Cytoscape[51] was used to visualize the interaction network.

**2.10. Expression analysis of m6A regulatory genes**

RNAseq datasets (GSE137263[58], GSE129509[59] and GSE186317[60]) were downloaded from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) of the NCBI (<https://www.ncbi.nlm.nih.gov/>) to observe the expression pattern in different organs, biotic and abiotic stress. Normalized FPKM (fragments per kilobase of exon per million) data were used to observe expression patterns in emerging nodules, mature nodules, emerging lateral roots, and mature lateral roots. To identify the expression patterns in abiotic stress including salt stress, heat stress, dehydration, water deficit, and combined water deficit and heat stress, LFC values were extracted using different R packages and a heatmap was plotted using TBtool[42]. The expression pattern of m6A regulatory genes in soybean mosaic virus (SMV) infection was also observed. Heatmap was plooted with LFC values that indicated upregulations and downregulation of these gene sets.

**3. Results**

**3.1. Genome‑wide identification and characterization of m6A regulatory genes in soybean**

After comprehensive screening of the soybean genome, 12 m6A writers, 11 m6A erasers and 19 m6A readers were finally identified. All m6A regulatory genes were named according to their subclass identifier. The amino acid sequence length, relative molecular weights (MWs), isoelectric points (pIs), subcellular localization and Gravy are listed in **Table 1**. It was found that the soybean writer's CDS length ranged from 624 bp to 6561 bp. GmVIRILIZER2 is the biggest member of the writer family and has a polypeptide length of 2187 aa and a molecular weight of 240878.49 kDa. Meanwhile, GmFIP37c is the smallest, having a polypeptide length of 208 aa and a molecular weight of 23126.92 Da. The pI values of the writer family differed widely, ranging from 5.04 (GmFIP37a) to 8.81 (GmMTC2). Amongst the twelve writer genes, ten genes exhibit acidic PI while the remaining two genes show basic features. Likewise, CDS length of m6A Eraser varies from 957 bp to 2052 bp although all the proteins belonged to ALKBH family. GmALKBH9B3 is the longest eraser protein with a 684 aa long and molecular weight of 73462.54 Da. On the other hand, the smallest one, GmALKBH10B4 is 957 aa long with a molecular weight of 35393.88 Da. In terms of PI, eight of eleven erasers have acidic PI whereas three others possess basic PI. The predicted CDS length of the m6A reader also varies from 387 bp (GmECT15) to 2127 bp (GmECT9). The largest one is 77175.5 Da in size and 709 aa long. The smallest one (GmECT15) is 129 aa long with a molecular weight of 14596.74 Da. In terms of localization, most m6A regulatory proteins were in the nucleus and only several proteins were in the plasma membrane (GmECT13, GmALKBH10B5, GmECT9) and chloroplast (GmMTC2, GmECT6). The Grand Average of Hydropathicity (GRAVY) values for all m6A proteins are less than zero, indicating that all soybean m6A proteins are hydrophilic.

**3.2. Chromosomal localization and gene duplication**

The identified putative 42 genes have been distributed throughout the 17 chromosomes of soybeans **(Table 1 and Figure 1)**. The distribution of genes across the chromosomes can be summarized as follows: Chromosomes 17 and 8 each have the highest count, with four genes. Following closely, chromosomes 16, 10, 14, 5, and 2 each have three genes. Chromosomes 20 and 19 have two genes each. Finally, chromosomes 4, 6, 3, 1, 12, 9, and 15 each have a single gene in their respective location. Different duplication events of m6A regulatory genes in soybeans were identified through PGDD[43], as gene duplication is an important mechanism for gene family expansion. A total of 18 duplication events were observed **(Table 2 and Figure 1)**. Six writers (GmCPSF30a-GmCPSF30b, GmMTA1-GmMTA2, GmMTB1-GmMTB2, GmFIP37a-GmFIP37c, GmFIP37b-GmFIP37d, GmVIRILIZER1-GmVIRILIZER2) sister gene pairs were observed to be duplicated. Alongside, five erasers were discovered to be duplicates: GmALKBH10B7-GmALKBH9B1, GmALKBH10B1-GmALKBH10B3, GmALKBH10B4-GmALKBH10B5, GmALKBH10B2-GmALKBH10B6, and GmALKBH9B2-GmALKBH9B3. In addition, sister gene pairs were shown to be duplicated in seven readers (GmECT4-GmECT11, GmECT15-GmECT16, GmECT5-GmECT12, GmECT9-GmECT13, GmECT8-GmECT17, GmECT1-GmECT3, and GmECT10-GmECT14). All the duplication events of m6A regulatory genes were segmental duplications **(Table 2)**. To explore the evolutionary history and direction of selection, non-synonymous to synonymous substitution (Ka/Ks) ratios were calculated. Purifying selection, neutral selection, and positive selection were indicated by Ka/Ks ratios below 1, equal to 1, and above 1, respectively. In this study all analyzed duplication events yielded Ka/Ks ratios were less than 1 indicating the influence of purifying selection on these genes throughout their evolutionary history. The time of divergence of these gene pairs varied from 5.71 to 15.50 Mya **(Table 2)**.

**3.3. Structural analysis of writer, eraser, and reader members**

Exon intron distribution indicates the diversity of gene functions. The profile of exon-introns of each identified putative gene was examined using GSDS (Gene Structure Display Server 2.0, <http://gsds.gao-lab.org/>). According to the gene structure, the exon numbers of soybean m6A writers, erasers, and readers varied from 6-27, 6-8 and 2-9, respectively **(Figure 2A)**. Amongst the m6A writers, the highest number of exons was observed in GmVIRILIZER1(27), and the lowest number was seen in GmMTB1(6) and GmMTB2(6). In terms of eraser, all the genes contain only 6 to 8 exons indicating the lowest number of exons among the three m6A regulatory proteins. Within readers, GmECT15 contains only 2 exons. Group-specific conserved domains were obtained using Batch CD-Search program[47] and Pfam (<http://pfam.xfam.org/>). MT-A70, Wtap, HAKAI and VIR\_N, indicate the m6A writer. MT-70 domain was found in MTA while MT-70 superfamily found in both MTB and MTC genes **(Figure 2B)**. In addition, GmFIP37a, GmFIP37b, GmFIP37c and GmFIP37d were characterized with WTAP superfamily. The VIR\_N superfamily was found in GmVIRILIZER1 and GmVIRILIZER2. According to the phytozome description, we found HAKAI-related protein. However, pfam analysis did not show any HAKAI domain for that protein. Therefore, this gmHAKAI was excluded from downstream analysis. The eraser family are characterized by the 2OG-FeII\_Oxy superfamily (Fe (II) dependent Oxygenase superfamily) and 2OG-FeII\_Oxy, all the identified putative eraser genes containing these conserved domains **(Figure 2B)**. YTH (YT-521-B-like domain) was used to identify reader proteins. All the readers possess that domain **(Figure 2B)**. In the motif analysis, MTC1 and MTC2 contain only one motif **(Figure 2C)**. Other members of the writer contain 4 to 6 motifs. All the eraser families contain 5-9 conserved motifs while 2-5 motifs were observed in the reader family.

**3.4. The phylogenetic classifications of m6A regulatory genes in soybean**

Demonstrating Phylogenetic relationships is crucial as it gives insight into deducing the gene origins, examining their molecular adaptation, learning about the evolution of morphological traits, and rebuilding demographic variations in recently differentiated species [61]. Phylogenetic trees for m6A writers, erasers, and readers, respectively, were constructed to study the evolutionary relationships of m6A modifiers. All the writer family, MT, FIP37(WTAP) VIR\_N and HAKAI formed 4 groups A, B, C, and D **(Figure 3A)**. The phylogeny revealed that GmVIRILIZER, GmMTA, GmMATB and GmMTC have 3 common closest orthologs of Gossypium barbadense, Linum usitatissimum and Anacardium occidentale. Another writer family GmFIP37 have the closest ortholog of Coffea arabica and Solanum lycopersicum while Linum usitatissimum is the distantly related ortholog. Zea mays is the most distantly related ortholog of GmVIRILIZER, demonstrating the plant kingdom's evolutionary divergence. On the other hand, the red algae species Porphyra umbilicalis has the most distant association with the orthologs of GmMTA, suggesting a wider evolutionary range. The orthologs of GmMTB and GmMTC are most distantly linked to Coffea arabica, highlighting the evolutionary connection among floral flowers. YTH domain-containing eraser protein orthologs also form 4 different groups A, B, C and D **(Figure 3B)**. The various species' readers are categorized according to their class **(Figure 3C).** Most of these readers' orthologs of Oryza sativa are out-grouped from their respective clads. While monocot-monocot and dicot-dicot couples are found in certain sister taxa, monocot-dicot pairs are found in others. The closely related species Acorus americanus and Aquilegia coerulea, for example, show strong evidence for their nodes and a recent shared ancestor. The implication here is that monocots and dicots may have split out from a common ancestor. These three evolutionary trees demonstrate how m6A dynamics are conserved across a variety of kingdoms. The presence and functional relevance of the m6A alteration in different species within these categories is demonstrated by this phylogeny.

**3.5. Identification of cis-regulatory elements in the putative promoter region**

The spatiotemporal specific expression of protein-coding genes can be uncovered by analyzing the cis-acting elements (CREs) in the promoter region. Using the PlantCare database[49], CREs in the promoter regions (1000 bp upstream of the transcription start site) of m6A regulatory genes were predicted to better understand the transcriptional regulation and potential biological functions of soybean m6A regulatory genes **(Figure 4)**. These CRE elements were divided into four groups including light-responsive, phytohormone-responsive, stress-responsive, and development-related. Light-responsive elements were predominant in all the m6A regulatory genes, suggesting that these gene sets may be regulated by light signalling. ABRE and TGA are predominant in phytohormone elements. These substantial number of phytohormone-responsive elements indicated the possible activation or inhibition of the expression level of the m6A regulatory gene by the hormone signals. In terms of development, CAT-box, GCN4-motif, ARE and O2-site were enriched in the promoter regions. These CREs were respectively associated with meristem expression, endosperm expression, anaerobic induction, and metabolism regulation. MBS (related to drought response), TC-rich repeat (related to defense and stress) and LTR (related to cold response) elements were also highly enriched.

**3.6. Functional analysis of m6A regulatory genes**

The GO analysis of the m6A regulatory genes revealed that the biological process was mostly engaged in the regulation of gene expression **(Figure 5A)**. Nucleic acid metabolic process, Regulation of mRNA metabolic process, Negative regulation of gene expression and mRNA metabolic process were the dominant molecular processes. Intracellular anatomical structure was seen in cellular component **(Figure 5B)**. mRNA binding, RNA binding, Nucleic acid binding and Binding were seen in molecular function **(Figure 5C)**. These results highlight the critical function of m6A regulatory genes in coordinating cellular and molecular processes that influence the dynamics of genes and RNA.

**3.7. Protein-protein interaction and cluster identification**

A multi-protein complex performs together in m6A modification, highlighting the prediction of a potential protein-protein interaction network in this regard. Therefore, a PPI network based on the STRING database[50] was constructed in this study. There were 192 edges and 42 nodes in the protein-protein interaction (PPI) network that was produced **(Figure 6)**. Two clusters were identified by using MCODE[52]. Cluster 1 contains writer proteins including MTA MTB and VIRILLIZER **(Figure 7A)**. Cluster 2 contains reader and writer proteins named GmFIP37a, GmFIP37b, GmFIP37c, GmFIP37d, MTC1, MTC 2, CPSF30a and ECT14**(Figure 7B)**. The cluster formation gives valuable insight into the biological mechanism of m6A regulatory proteins. These proteins are vital for plant growth and development as well as for how well plants adapt to their surroundings. These findings shed important light on the molecular mechanism and possible biological purposes of soybean m6A regulatory genes.

**3.6. Identification of hub genes and their 3D structure modelling**

The top ten hub genes were extracted by cytohubba plugin[53] in Cytoscape[51]. Notably, there was no eraser member in this hub genes **(Figure 7C)**. GmVIRILIZER1 may interact with GmVIRILIZER2, GmECT10, GmECT14, GmMTB1, and GmMTB2, according to the network. GmECT10 and GmECT14 interact with GmMTB1. Furthermore, GmMTA1 engages in interactions with every hub gene in this network. However, GmMTA2 has no connection with GmMTA1 or GmCPSF30a. Furthermore, GmMTB2 engages in interactions with GmMTB1, GmECT14, and GmECT10. No interaction was found between GmCPSF30a and reader hub genes. However, GmFIP37d interacts with two reader families including GmECT10 and GmECT14. GmVIRILIZER2 interacts with GmMTB1, GmECT10, GmECT14 and GmMTB2. The protein structure of 10 hub genes was predicted using SWISS-MODEL[54]. Protein structure has a polypeptide chain of amino acids specific to each protein type. The chain is folded into a secondary structure characterized by alpha-helices and beta-sheets[62]. All the predicted structures are depicted in **(Figure 8).**

**3.8. Micro RNA target prediction**

Potential miRNA binding sites of m6A regulatory genes were also examined by using psRNATarget[57]. 10 writers had 81 miRNAs, 10 erasers had 31 miRNAs, and 18 readers had 60 miRNAs after interactions with an Expectation penalty score less than 5 were filtered out **(Figure 9 and Supplementary Table 1)**. It is noteworthy that soybean m6A writers (n = 81) seemed to be more regulated by miRNA than m6A readers (n = 60) and erasers (n = 31). Different miRNA families have been shown to exhibit specific target genes frequently **(Figures 9A, 9B and 9C)**. Interestingly our research did not identify any miRNAs that regulate the two writers (GmMTC1 and GmMTC2), one eraser (ALKBH9B2), and one reader (ECT16) in soybeans. Given the context-dependent activation of miRNA-mediated regulation, more research is required to understand the complex interactions among m6A regulatory genes fully.

**3.9. Roots and Nodules expression analysis of m6A regulatory genes in soybean.**

Symbiotic legume nodules and lateral roots arise away from the root meristem via dedifferentiation events in soybean[59]. These organs share some morphological and developmental similarities. We explored the expression level of m6A in emerging nodules (EN), mature nodules (MN), emerging lateral roots (ELR) and young lateral roots (YLR) of soybeans. Most of the genes displayed a broad expression range across all the samples indicating that they were extensively involved in the growth and development of roots and nodules **(Figure 10)**. GmMTA1 displayed high expression levels in mature nodules. Notably, GmMTB1, GmFIP37a, GmFIP37d, GmALKBH10B6, GmECT1, and GmECT12 showed a medium level of expression in all the selected tissues (EN, MN, ELR and YLR). GmALKBH10B5 showed higher expression levels in mature nodules compared to EN, ELR and YLR. Additionally, GmECT9 was expressed highly in all the selected tissues while GmECT13 exhibited high expressions in EN and ELR compared to MN and YLR. Importantly, GmECT17 showed higher expression in ELR but displayed medium expression in the other three tissues.

**3.10. Expression patterns of m6A regulatory genes under various abiotic and biotic stress**

The expression pattern of m6A regulatory genes was observed under different abiotic and biotic stress **(Figure 11A and 11B)**. The majority of m6A regulatory genes remained unresponsive when subjected to salt stress and dehydration. GmMTB2 and GmECT9 exhibited significant upregulations in heat stress and combined water deficit and heat stress **(Figure 11A)**. On the contrary, GmALKBH9B2 showed downregulation in combined water deficit and heat stress. The remaining soybean's m6A genes displayed a medium level of expression except GmECT4 and GmECT2. These two genes did not show any changes across all the abiotic stresses **(Figure 11A)**.

The soybean mosaic virus (SMV) treatment resulted in the significantly higher up-regulation of multiple genes, such as GmMTB1, GmMTB2 and GmALKBH9B1 **(Figure 11B)**. In the writer's family GmMTA2, GmCPSF30a, GmVIRILIZER1, GmMTA1, GmVIRILIZER2, and GmCPSF30b showed upregulation compared to control. In the ALKBH group, three genes such as GmALKBH10B3 GmALKBH10B7 and GmALKBH10B6 upregulated in SMV treatment. Additionally, GmECT9, GmECT2, GmECT17, GmECT15 and GmECT7 displayed upregulation in the readers group. GmALKBH9B4 was the only member of the m6A regulatory gene that displayed significant downregulation in SVM treatment. Overall, the expression patterns of m6A regulatory genes in soybeans under different stress conditions indicated that they might play critical biological roles in various stress responses.

**4. Discussion**

RNA N6-methyladenosine operates a crucial regulatory role in plant growth and development. The methylation levels of target transcripts are dynamically regulated by three types of m6A regulatory genes. Herein we extracted a total of 42 candidate m6A regulatory genes in soybean, including 12 writers, 11 erasers, and 19 readers **(Table 1)**. The gene sets were confirmed by the conserved domain of writer, reader, and eraser. Notably, one of the writer's domains named HAKAI was not found in soybean according to Pfam. Therefore, this gene was excluded from downstream analysis. Soybean (42) has more m6A regulatory genes than others, such as Arabidopsis (33), tomato (25), grape (40), rice (33), S. moellendorffii (22), M. polymorpha (16), P. patens (18), and Chinese pine (36)[63]. It falls short of tobacco (52) and upland cotton (75), common wheat (85), and maize (55)[63]. A low number of m6A regulatory genes were found in Chinese pine (36) compared to species such as common wheat and upland cotton, though Chinese pine has the largest genome size (25.4 Gb). The plants which have experienced polyploidization events such as Common wheat (2n = 6x = 42), upland cotton (2n = 4x = 52), and tobacco (2n = 4x = 48) exhibit a higher abundance of m6A regulatory genes[64][65][66]. Likewise, Soybean (2n = 4x = 40)[67], a polyploid species also possesses a higher number of m6A regulatory genes. additionally, maize, a segmental allopolyploid, also shows a higher number of m6A regulatory genes[68]. Using PGDD, a total of 18 duplication events were identified in this study. All the duplication events were Segmental duplication **(Table 2) (Figure 1).** From the Ka and Ks analysis, duplicated gene pairs went through purifying selection, a finding consistent with studies on tomatoes[69], tea plants[70], and tobacco[63]. Structural analysis of the identified genes was also performed which revealed unique gene structures, domains and motifs of writer, reader, and eraser proteins **(Figure 2)**.

In order to explore the evolutionary patterns of m6A writer, eraser, and reader genes, phylogenetic trees were constructed across 13 plant species which included Rhodophyta (red algae), and Angiosperms (comprising eudicots and monocots) **(Figure 3)**. Monocot-dicot, monocot-monocot, and dicot-dicot pairs were spotted in this study. For instance, Acorus americanus and Aquilegia coerulea share a recent ancestor, implying shared monocot-dicot divergence. In the previous study on m6A genes of Oryza sativa, a similar evolutionary pattern was observed[71]. In another study on Arabidopsis, the MT domain was found to closely resemble among distinct plant species while other components of m6A were organized differently. Cis-regulatory elements in the promoter region were identified. According to this analysis, CREs were associated with light response, hormone response, plant growth and development, and stress response. Similar findings were also observed in other plant species such as tobacco[63], tea[70], tomato[69], and poplar[72]. Light responsive CREs were predominant in soybean m6A regulatory genes **(Figure 4)**.

GO terms in three categories including biological process, cellular components and molecular function were observed. These results indicate the involvement in RNA modification of writer eraser and reader genes of soybean **(Figure 5)**. The PPI network revealed that MTA1, MTA2, MTB1 and MTB2 are highly interconnected genes, indicating their importance as the catalytic core **(Figure 6)**. Most of the erasers did not interact with the writer and reader. On the other side, CPSF30a and CPSF30b two readers had stronger interaction than any other reader components **(Figure 6)**. Cluster analysis showed two cluster formations of m6A regulatory genes of soybeans **(Figure 7)**. 6 writer components formed cluster 1, highlighting their dynamic action in RNA modification **(Figure 7A)**. Writer components also predominated in cluster 2 nevertheless no erasers were found in any cluster **(Figures 7A and 7B)**. Previous studies also found similar results that there were no interactions among eraser proteins in tobacco[63]. Furthermore, hub gene identification was performed. This exhibited that there were no eraser genes in the top 10 hub genes **(Figure 7C)**. Using homology modelling, the 3D structure of 10 hub genes was predicted as these genes might be the core catalytic component in m6A methylation **(Figure 8)**. The secondary structures such as alpha helices and beta sheath are important to maintain protein stability and functionality[73]. To explore the mutual interaction of miRNA and m6A soybean, miRNA targets prediction was performed. Interestingly writer component had greater miRNA regulation compared to readers and erasers **(Figures 9A, 9B and 9C)**. These findings indicate that soybean writers may prefer to be regulated by miRNA.

Researchers have demonstrated that altering the m6A system may cause irregular growth and development. Many previous studies have mentioned the association of m6A modification with reproductive cell development and fruit ripening. In this study, soybean m6A expression was observed in roots and nodules. Readers and erasers have a higher expression than writers regarding roots and nodules **(Figure 10)**. These findings pinpoint the dynamic action of m6A regulatory genes of soybeans. Two reader genes named GmECT9 and GmECT13 are highly expressed in the respective samples. According to prior studies, CPSF30L in Arabidopsis might regulate APA by binding to m6A-modified RNAs that are involved in nitrate signaling, particularly WRKY1 and NRT1.1[74]. As root and nodule have a role in nitrogen absorption, we hypothesized that GmECT9 and GmECT13 have similar mechanisms. However further investigation is needed in this regard.

CREs elements such as MBS (related to drought response), TC-rich repeat (related to defense and stress) and LTR (related to cold response) were highly enriched in soybean m6A genes according to Promoter analysis **(Figure 4)**. This result indicates the involvement of soybeans' m6A genes in environmental stimuli. Salt stress has been shown to have a major impact on the m6A methylation levels on mRNA in Arabidopsis[32]. The expression patterns of GmMTAs and GmMTBs under various abiotic stressors have been documented in earlier research, suggesting their possible role in stress tolerance, particularly in the reaction to darkness or alkalinity[36]. According to another research, GmALKB10B genes were probably triggered by alkalinity, cold, and dryness[37]. In this study, five abiotic stress responses such as salt stress, dehydration, heat stress, Water deficit and Combined water and heat had been reported. Most of the soybean's m6A genes did not show any changes in salt and dehydration stress, indicating their stress tolerance possibility **(Figure 11A)**. Notably two genes named GmMTB2 and GmALKBH9B2 showed upregulation and downregulation in Combined water and heat stresses. Interestingly GmECT4 and GmECT2 did not show any changes across all the abiotic stresses. Therefore, we assumed that these two genes might be stress tolerance m6A genes in soybeans. Validating these findings require a wet lab inquiry. Distinct biotic stresses have also a great impact on the plant life cycle. NtALKBH10 and NtVIR1 of tobacco exhibited robust responses to specific biotic stresses (R. solanacearum and black shank) but showed insensitivity to Cucumber Mosaic Virus (CMV)[63]. In this study, biotic stress responses (soybean mosaic virus) have been documented. Most of the soybean's m6A genes displayed upregulation in SMV treatment. Nevertheless, one member called GmALKBH9B4 downregulated in this regard. In conclusion, the m6A regulatory genes of soybeans showed a variety of expression patterns under stress, nodules, and roots, suggesting that these genes have a multitude of roles. The precise processes of the putative m6A regulatory genes in soybeans required further investigation.

**5. Conclusion**

This study presents a comprehensive and systematic investigation of m6A regulatory genes in soybeans and a total of 42 m6A regulatory genes were identified. we analyzed the features of these genes in terms of gene structure, conserved domains, and motifs. Phylogenetic study and duplication events were also examined. Functional analysis of m6A regulatory genes in cis-elements, and interaction networks, demonstrated their crucial roles in soybeans. The altered expression pattern of these genes had been profiled in roots, nodules, abiotic and biotic stresses. This study gives a reference framework for investigating the functional variety of soybean m6A regulatory genes at the epigenetic level.

**6. References**

[1] Z. Tao, F. Yan, M. Hahn, and Z. Ma, “Regulatory roles of epigenetic modifications in plant ‑ phytopathogen interactions,” *Crop Heal.*, pp. 1–20, 2023, doi: 10.1007/s44297-023-00003-y.

[2] C. Deans and K. A. Maggert, “What do you mean, ‘Epigenetic’?,” *Genetics*, vol. 199, no. 4, pp. 887–896, 2015, doi: 10.1534/genetics.114.173492.

[3] J. Tang, S. Chen, and G. Jia, “Detection, regulation, and functions of RNA N6-methyladenosine modification in plants,” *Plant Commun.*, vol. 4, no. 3, p. 100546, 2023, doi: 10.1016/j.xplc.2023.100546.

[4] Z. Liang, A. Riaz, S. Chachar, Y. Ding, H. Du, and X. Gu, “Epigenetic Modifications of mRNA and DNA in Plants,” *Mol. Plant*, vol. 13, no. 1, pp. 14–30, 2020, doi: 10.1016/j.molp.2019.12.007.

[5] P. Boccaletto *et al.*, “MODOMICS: A database of RNA modification pathways. 2017 update,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D303–D307, 2018, doi: 10.1093/nar/gkx1030.

[6] G. Bhatia, W. Prall, B. Sharma, and B. D. Gregory, “Covalent RNA modifications and their budding crosstalk with plant epigenetic processes,” *Curr. Opin. Plant Biol.*, vol. 69, no. 215, 2022, doi: 10.1016/j.pbi.2022.102287.

[7] X. Fang, M. Li, T. Yu, G. Liu, and J. Wang, “Reversible N6-methyladenosine of RNA: The regulatory mechanisms on gene expression and implications in physiology and pathology,” *Genes Dis.*, vol. 7, no. 4, pp. 585–597, 2020, doi: https://doi.org/10.1016/j.gendis.2020.06.011.

[8] Y. Wang *et al.*, “Epigenetic modification of m 6 A regulator proteins in cancer,” pp. 1–19, 2023.

[9] L. Shen *et al.*, “N6-Methyladenosine RNA Modification Regulates Shoot Stem Cell Fate in Arabidopsis,” *Dev. Cell*, vol. 38, no. 2, pp. 186–200, 2016, doi: 10.1016/j.devcel.2016.06.008.

[10] Z. Bodi *et al.*, “Adenosine methylation in Arabidopsis mRNA is associated with the 3′ end and reduced levels cause developmental defects,” *Front. Plant Sci.*, vol. 3, no. MAR, pp. 1–10, 2012, doi: 10.3389/fpls.2012.00048.

[11] S. Zhong *et al.*, “MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor,” *Plant Cell*, vol. 20, no. 5, pp. 1278–1288, 2008, doi: 10.1105/tpc.108.058883.

[12] J. Hu, S. Manduzio, and H. Kang, “Epitranscriptomic RNA Methylation in Plant Development and Abiotic Stress Responses,” vol. 10, no. April, pp. 1–11, 2019, doi: 10.3389/fpls.2019.00500.

[13] K. Růžička *et al.*, “Identification of factors required for m6A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI,” *New Phytol.*, vol. 215, no. 1, pp. 157–172, 2017, doi: 10.1111/nph.14586.

[14] J. Liu *et al.*, “A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation.,” *Nat. Chem. Biol.*, vol. 10, no. 2, pp. 93–95, 2014, doi: 10.1038/nchembio.1432.

[15] J. Liu *et al.*, “VIRMA mediates preferential m6A mRNA methylation in 3′UTR and near stop codon and associates with alternative polyadenylation,” *Cell Discov.*, vol. 4, no. 1, 2018, doi: 10.1038/s41421-018-0019-0.

[16] G. Jia *et al.*, “N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO,” *Nat. Chem. Biol.*, vol. 7, no. 12, pp. 885–887, 2011, doi: 10.1038/nchembio.687.

[17] G. Zheng *et al.*, “ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility,” *Mol. Cell*, vol. 49, no. 1, pp. 18–29, 2013, doi: https://doi.org/10.1016/j.molcel.2012.10.015.

[18] H. C. Duan *et al.*, “ALKBH10B is an RNA N6-methyladenosine demethylase affecting arabidopsis floral transition,” *Plant Cell*, vol. 29, no. 12, pp. 2995–3011, 2017, doi: 10.1105/tpc.16.00912.

[19] C. Xu *et al.*, “Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain,” *Nat. Chem. Biol.*, vol. 10, no. 11, pp. 927–929, 2014, doi: 10.1038/nchembio.1654.

[20] X. Wang *et al.*, “N 6-methyladenosine-dependent regulation of messenger RNA stability,” *Nature*, vol. 505, no. 7481, pp. 117–120, 2014, doi: 10.1038/nature12730.

[21] X. Wang *et al.*, “N6-methyladenosine Modulates Messenger RNA Translation Efficiency,” *Cell*, vol. 161, no. 6, pp. 1388–1399, 2015, doi: https://doi.org/10.1016/j.cell.2015.05.014.

[22] H. Shi *et al.*, “YTHDF3 facilitates translation and decay of N 6-methyladenosine-modified RNA,” *Cell Res.*, vol. 27, no. 3, pp. 315–328, 2017, doi: 10.1038/cr.2017.15.

[23] H. Huang *et al.*, “Recognition of RNA N 6 -methyladenosine by IGF2BP proteins enhances mRNA stability and translation,” *Nat. Cell Biol.*, vol. 20, no. 3, pp. 285–295, 2018, doi: 10.1038/s41556-018-0045-z.

[24] B. M. Edens *et al.*, “FMRP Modulates Neural Differentiation through m6A-Dependent mRNA Nuclear Export,” *Cell Rep.*, vol. 28, no. 4, pp. 845-854.e5, 2019, doi: https://doi.org/10.1016/j.celrep.2019.06.072.

[25] I. A. Roundtree *et al.*, “YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs,” *Elife*, vol. 6, p. e31311, 2017, doi: 10.7554/eLife.31311.

[26] P. J. Hsu *et al.*, “Ythdc2 is an N6 -methyladenosine binding protein that regulates mammalian spermatogenesis,” *Cell Res.*, vol. 27, no. 9, pp. 1115–1127, 2017, doi: 10.1038/cr.2017.99.

[27] J. Scutenaire *et al.*, “The YTH domain protein ECT2 is an m6A reader required for normal trichome branching in arabidopsis,” *Plant Cell*, vol. 30, no. 5, pp. 986–1005, 2018, doi: 10.1105/tpc.17.00854.

[28] L. Arribas-Hernández, S. Bressendorff, M. H. Hansen, C. Poulsen, S. Erdmann, and P. Brodersen, “An m6A-YTH module controls developmental timing and morphogenesis in arabidopsis,” *Plant Cell*, vol. 30, no. 5, pp. 952–967, 2018, doi: 10.1105/tpc.17.00833.

[29] J. Hu, J. Cai, T. Xu, and H. Kang, “Epitranscriptomic mRNA modifications governing plant stress responses: underlying mechanism and potential application,” *Plant Biotechnol. J.*, vol. 20, no. 12, pp. 2245–2257, 2022, doi: 10.1111/pbi.13913.

[30] L. Shen *et al.*, “N6-Methyladenosine RNA Modification Regulates Shoot Stem Cell Fate in Arabidopsis,” *Dev. Cell*, vol. 38, no. 2, pp. 186–200, 2016, doi: https://doi.org/10.1016/j.devcel.2016.06.008.

[31] N. Hou *et al.*, “MdMTA-mediated m6A modification enhances drought tolerance by promoting mRNA stability and translation efficiency of genes involved in lignin deposition and oxidative stress,” *New Phytol.*, vol. 234, no. 4, pp. 1294–1314, 2022, doi: 10.1111/nph.18069.

[32] J. Hu *et al.*, “N6-Methyladenosine mRNA methylation is important for salt stress tolerance in Arabidopsis,” *Plant J.*, vol. 106, no. 6, pp. 1759–1775, 2021, doi: 10.1111/tpj.15270.

[33] D. Yang *et al.*, “RNA N6-Methyladenosine Responds to Low-Temperature Stress in Tomato Anthers,” *Front. Plant Sci.*, vol. 12, no. June, pp. 1–17, 2021, doi: 10.3389/fpls.2021.687826.

[34] T. Y. Zhang *et al.*, “Transcriptome-Wide N6-Methyladenosine (m6A) Profiling of Susceptible and Resistant Wheat Varieties Reveals the Involvement of Variety-Specific m6A Modification Involved in Virus-Host Interaction Pathways,” *Front. Microbiol.*, vol. 12, no. May, pp. 1–13, 2021, doi: 10.3389/fmicb.2021.656302.

[35] R. Dilawari *et al.*, “Soybean: A Key Player for Global Food Security BT - Soybean Improvement: Physiological, Molecular and Genetic Perspectives,” S. H. Wani, N. ul R. Sofi, M. A. Bhat, and F. Lin, Eds. Cham: Springer International Publishing, 2022, pp. 1–46. doi: 10.1007/978-3-031-12232-3\_1.

[36] P. Liu *et al.*, “Genome-wide identification and functional analysis of mRNA m6A writers in soybean under abiotic stress,” *Front. Plant Sci.*, vol. 15, no. July, pp. 1–16, 2024, doi: 10.3389/fpls.2024.1446591.

[37] J. Zhao *et al.*, “Genome-Wide Identification of the Soybean AlkB Homologue Gene Family and Functional Characterization of GmALKBH10Bs as RNA m6A Demethylases and Expression Patterns under Abiotic Stress,” *Plants*, vol. 13, no. 17, p. 2491, 2024, doi: 10.3390/plants13172491.

[38] Z. S. Nishat, S. Hasan, S. Islam, T. Hossain, and A. Ghosh, “Current Plant Biology Identification of epitranscriptomic methylation marker genes in Arabidopsis and their expression profiling in response to developmental , anatomical , and environmental modulations,” *Curr. Plant Biol.*, vol. 30, no. May, p. 100247, 2022, doi: 10.1016/j.cpb.2022.100247.

[39] D. M. Goodstein *et al.*, “Phytozome: A comparative platform for green plant genomics,” *Nucleic Acids Res.*, vol. 40, no. D1, pp. 1178–1186, 2012, doi: 10.1093/nar/gkr944.

[40] M. R. Wilkins *et al.*, “Protein Identification and Analysis Tools in the ExPASy Server BT - 2-D Proteome Analysis Protocols,” A. J. Link, Ed. Totowa, NJ: Humana Press, 1999, pp. 531–552. doi: 10.1385/1-59259-584-7:531.

[41] C. Savojardo, P. L. Martelli, P. Fariselli, G. Profiti, and R. Casadio, “BUSCA: An integrative web server to predict subcellular localization of proteins,” *Nucleic Acids Res.*, vol. 46, no. W1, pp. W459–W466, 2018, doi: 10.1093/nar/gky320.

[42] C. Chen *et al.*, “TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data,” *Mol. Plant*, vol. 13, no. 8, pp. 1194–1202, 2020, doi: 10.1016/j.molp.2020.06.009.

[43] T. H. Lee, H. Tang, X. Wang, and A. H. Paterson, “PGDD: A database of gene and genome duplication in plants,” *Nucleic Acids Res.*, vol. 41, no. D1, pp. 1152–1158, 2013, doi: 10.1093/nar/gks1104.

[44] N. Juretic, D. R. Hoen, M. L. Huynh, P. M. Harrison, and T. E. Bureau, “The evolutionary fate of MULE-mediated duplications of host gene fragments in rice,” *Genome Res.*, vol. 15, no. 9, pp. 1292–1297, 2005.

[45] A. H. Eren, “Genome-Wide Identi fi cation of ALDH Gene Family under Salt and Drought Stress in Phaseolus vulgaris,” 2024, doi: 10.32604/phyton.2024.058627.

[46] K. Tamura, G. Stecher, and S. Kumar, “MEGA11: Molecular Evolutionary Genetics Analysis Version 11,” *Mol. Biol. Evol.*, vol. 38, no. 7, pp. 3022–3027, 2021, doi: 10.1093/molbev/msab120.

[47] S. Lu *et al.*, “CDD/SPARCLE: The conserved domain database in 2020,” *Nucleic Acids Res.*, vol. 48, no. D1, pp. D265–D268, 2020, doi: 10.1093/nar/gkz991.

[48] R. C. Edgar, “MUSCLE: Multiple sequence alignment with high accuracy and high throughput,” *Nucleic Acids Res.*, vol. 32, no. 5, pp. 1792–1797, 2004, doi: 10.1093/nar/gkh340.

[49] M. Lescot *et al.*, “PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 325–327, 2002, doi: 10.1093/nar/30.1.325.

[50] D. Szklarczyk *et al.*, “STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, 2019, doi: 10.1093/nar/gky1131.

[51] P. Shannon *et al.*, “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, 2003.

[52] C. W. Hogue and M. Groll, “An automated method for finding molecular complexes in large protein interaction networks,” *BMC Bioinformatics*, vol. 29, no. 1, pp. 137–140, 2001, [Online]. Available: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/29.1.137

[53] C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, and C. Y. Lin, “cytoHubba: Identifying hub objects and sub-networks from complex interactome,” *BMC Syst. Biol.*, vol. 8, no. 4, pp. 1–7, 2014, doi: 10.1186/1752-0509-8-S4-S11.

[54] A. Waterhouse *et al.*, “SWISS-MODEL: Homology modelling of protein structures and complexes,” *Nucleic Acids Res.*, vol. 46, no. W1, pp. W296–W303, 2018, doi: 10.1093/nar/gky427.

[55] E. C. Meng *et al.*, “UCSF ChimeraX: Tools for structure building and analysis,” *Protein Sci.*, vol. 32, no. 11, pp. 1–13, 2023, doi: 10.1002/pro.4792.

[56] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, “MiRBase: From microRNA sequences to function,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D155–D162, 2019, doi: 10.1093/nar/gky1141.

[57] X. Dai, Z. Zhuang, and P. X. Zhao, “PsRNATarget: A plant small RNA target analysis server (2017 release),” *Nucleic Acids Res.*, vol. 46, no. W1, pp. W49–W54, 2018, doi: 10.1093/nar/gky316.

[58] L. C. DeMers, N. R. Redekar, A. Kachroo, S. A. Tolin, S. Li, and M. A. Saghai Maroof, “A transcriptional regulatory network of Rsv3-mediated extreme resistance against Soybean mosaic virus,” *PLoS One*, vol. 15, no. 4, pp. 1–20, 2020, doi: 10.1371/journal.pone.0231658.

[59] S. Adhikari, S. Damodaran, and S. Subramanian, “Lateral root and nodule transcriptomes of soybean,” *Data*, vol. 4, no. 2, pp. 1–11, 2019, doi: 10.3390/data4020064.

[60] R. Sinha *et al.*, “The transcriptome of soybean reproductive tissues subjected to water deficit, heat stress, and a combination of water deficit and heat stress,” *Plant J.*, vol. 116, no. 4, pp. 1064–1080, 2023, doi: 10.1111/tpj.16222.

[61] P. Kapli, Z. Yang, and M. J. Telford, “Phylogenetic tree building in the genomic age,” *Nat. Rev. Genet.*, vol. 21, no. 7, pp. 428–444, 2020, doi: 10.1038/s41576-020-0233-0.

[62] A. Al Masum, M. S. Arman, and A. Ghosh, “Methylglyoxal detoxifying gene families in tomato: Genome-wide identification, evolution, functional prediction, and transcript profiling,” *PLoS One*, vol. 19, no. 6 June, pp. 1–27, 2024, doi: 10.1371/journal.pone.0304039.

[63] H. Su *et al.*, “Genome ‑ wide identification of the N 6 ‑ methyladenosine regulatory genes reveals NtFIP37B increases drought resistance of tobacco ( Nicotiana tabacum L .),” *BMC Plant Biol.*, pp. 1–18, 2024, doi: 10.1186/s12870-024-04813-2.

[64] F. Li *et al.*, “Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution,” *Nat. Biotechnol.*, vol. 33, no. 5, pp. 524–530, 2015, doi: 10.1038/nbt.3208.

[65] I. J. LEITCH and M. D. BENNETT, “Biological relevance of polyploidy : ecology to genomics Polyploidy in arctic plants,” *Society*, vol. 82, no. 4, pp. 521–536, 2004, [Online]. Available: http://www.mendeley.com/research/biological-relevance-polyploidy-ecology-genomics-polyploidy-arctic-plants/

[66] L. J. Kelly, A. R. Leitch, J. J. Clarkson, R. B. Hunter, S. Knapp, and M. W. Chase, “Intragenic recombination events and evidence for hybrid speciation in nicotiana (solanaceae),” *Mol. Biol. Evol.*, vol. 27, no. 4, pp. 781–799, 2010, doi: 10.1093/molbev/msp267.

[67] W. R. Clarindo, C. R. De Carvalho, and B. M. G. Alves, “Mitotic evidence for the tetraploid nature of Glycine max provided by high quality karyograms,” *Plant Syst. Evol.*, vol. 265, no. 1–2, pp. 101–107, 2007, doi: 10.1007/s00606-007-0522-5.

[68] P. S. Schnable *et al.*, “The B73 Maize Genome: Complexity, Diversity, and Dynamics,” *Science (80-. ).*, vol. 326, no. 5956, pp. 1112–1115, Nov. 2009, doi: 10.1126/science.1178534.

[69] H. Shen, B. Luo, Y. Wang, J. Li, Z. Hu, and Q. Xie, “Genome-Wide Identification , Classification and Expression Analysis of m 6 A Gene Family in Solanum lycopersicum,” 2022.

[70] C. Zhu and S. Xie, “Genome-Wide Investigation of N6-Methyladenosine Regulatory Genes and Their Roles in Tea ( Camellia sinensis ) Leaves During Withering Process,” vol. 12, no. June, 2021, doi: 10.3389/fpls.2021.702303.

[71] M. Hasan, Z. S. Nishat, T. Hossain, and A. Ghosh, “Biochemistry and Biophysics Reports,” no. xxxx.

[72] X. Sun, W. Wu, Y. Yang, I. Wilson, F. Shao, and D. Qiu, “Genome-Wide Identification of m6A Writers, Erasers and Readers in Poplar 84K,” *Genes (Basel).*, vol. 13, no. 6, 2022, doi: 10.3390/genes13061018.

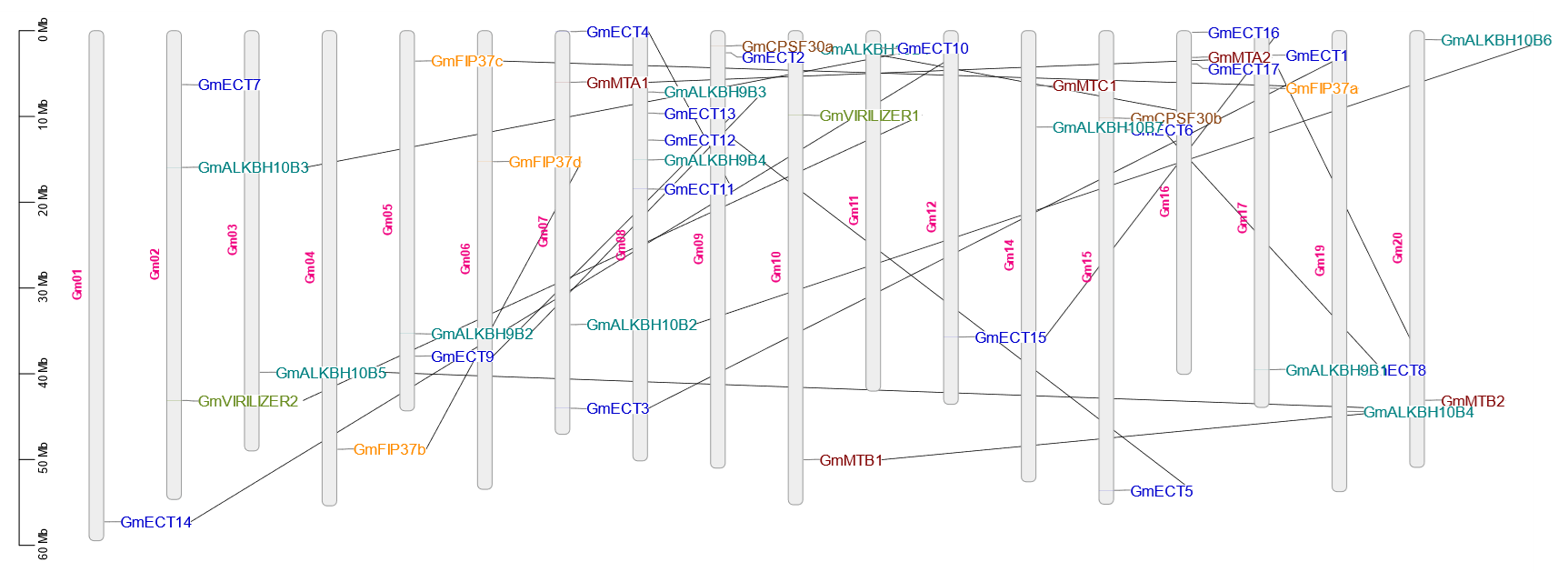
[73] D. Eisenberg, “The discovery of the α-helix and β-sheet, the principal structural features of proteins,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 20, pp. 11207–11210, 2003, doi: 10.1073/pnas.2034522100.

[74] Y. Hou *et al.*, “CPSF30-L-mediated recognition of mRNA m6A modification controls alternative polyadenylation of nitrate signaling-related gene transcripts in Arabidopsis,” *Mol. Plant*, vol. 14, no. 4, pp. 688–699, 2021, doi: 10.1016/j.molp.2021.01.013.

**Table 1**: List of identified RNA m6A genes in *Glycine max* along with their detailed information and subcellular localization.



**Abbreviations:** SN, serial number; S, subcellular; CDS, coding DNA Sequence; PP, Polypeptide; MW, Molecular Weight; pI, Isoelectric point; bp, base pair; aa, amino acid; kDa, kilodalton; Cp, Chloroplast; Ec, Extracellular; Nu, Nucleus; Pm, Plasma-membrane.



**Figure 1:** Chromosomal distribution of m6A regulatory genes. All m6A regulatory genes are found to be in different chromosomes of soybeans. The relative size of the corresponding chromosomes and the position of the respective genes could be estimated by using the scale provided left side of the figure. The chromosome numbers are provided at the middle of each bar. The straight lines connect the duplicated gene pairs **(Table 2)**.

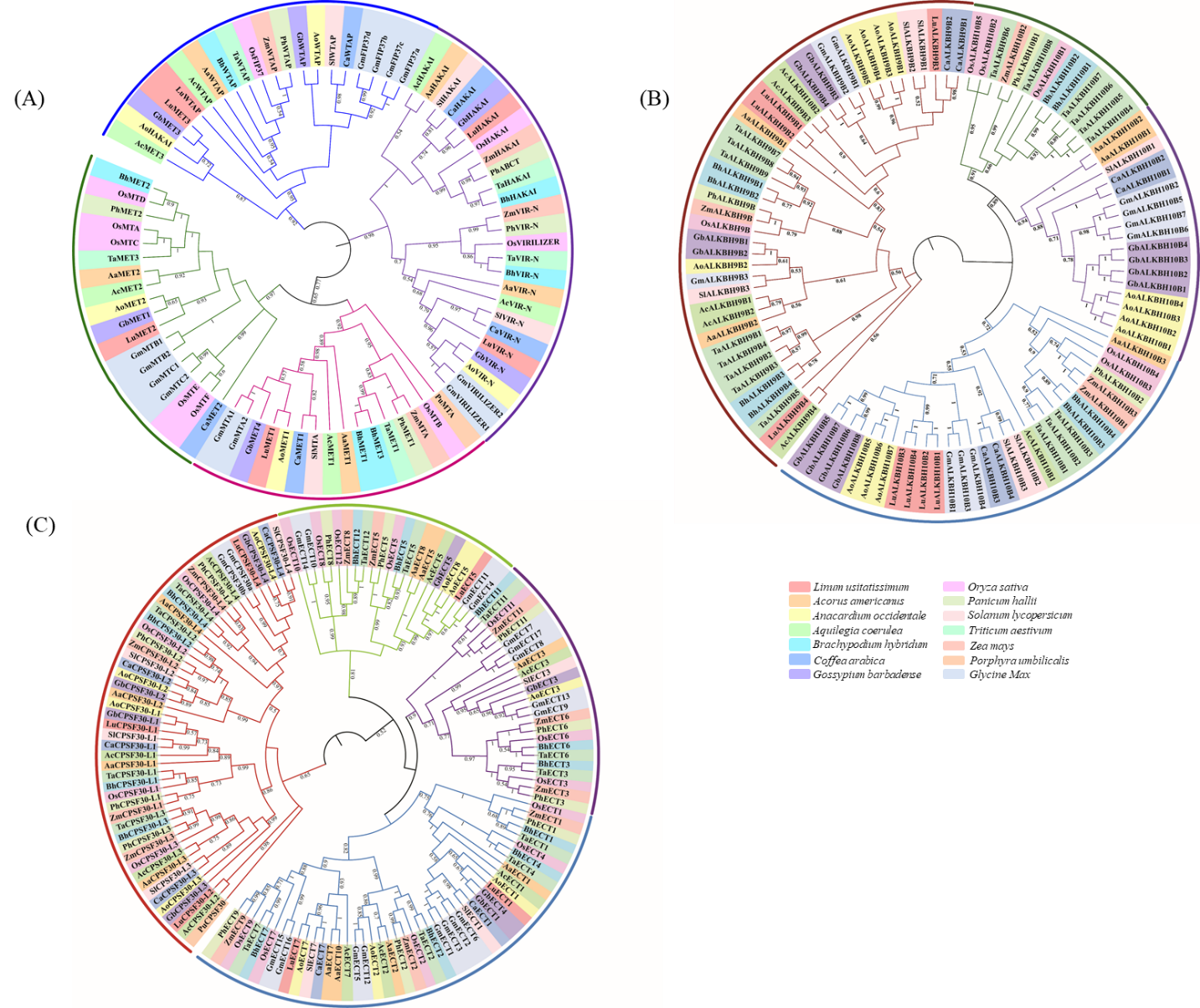
**Table 2:** List of duplicated genes identified in soybean (*Glycine max*)



A screenshot of a computer

Description automatically generated

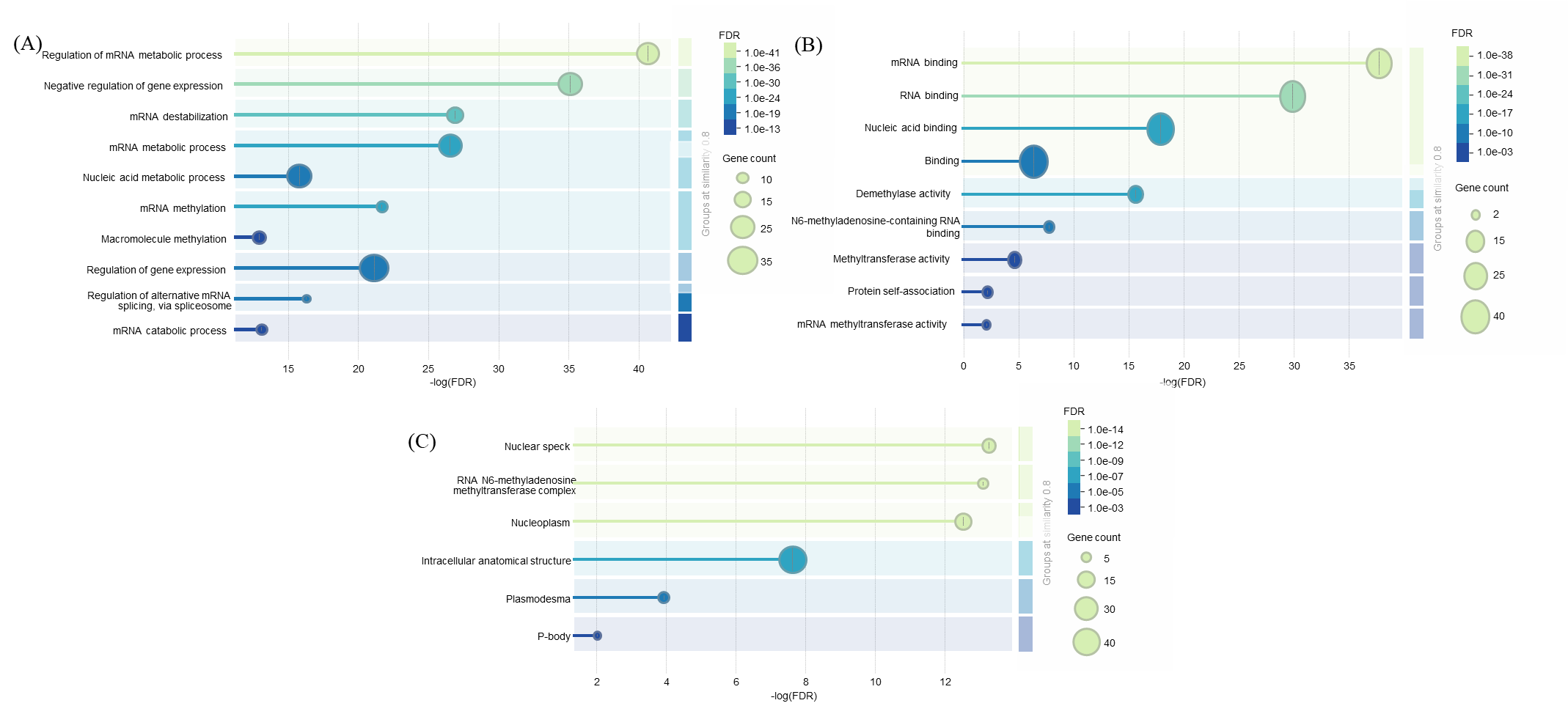
**Figure 2:** The evolutionary relationship together with the gene structure, the appropriate domain architecture and motif of the m6A regulating genes in soybean. **A.** Using MEGA-X and GSDS, respectively, three groups' phylogenetic trees and exon-intron architectures were constructed. The exon-intron length was proportionately shown and could be approximated using the scale below. **B.** Functional domains of m6A regulatory genes in soybean. The relative size of the protein and the respective domains' position could be estimated using the scale provided below. **C.** The organisation and distribution of the conserved motifs in the m6A writer genes. The squares in the motif represent the positions of conserved domains.



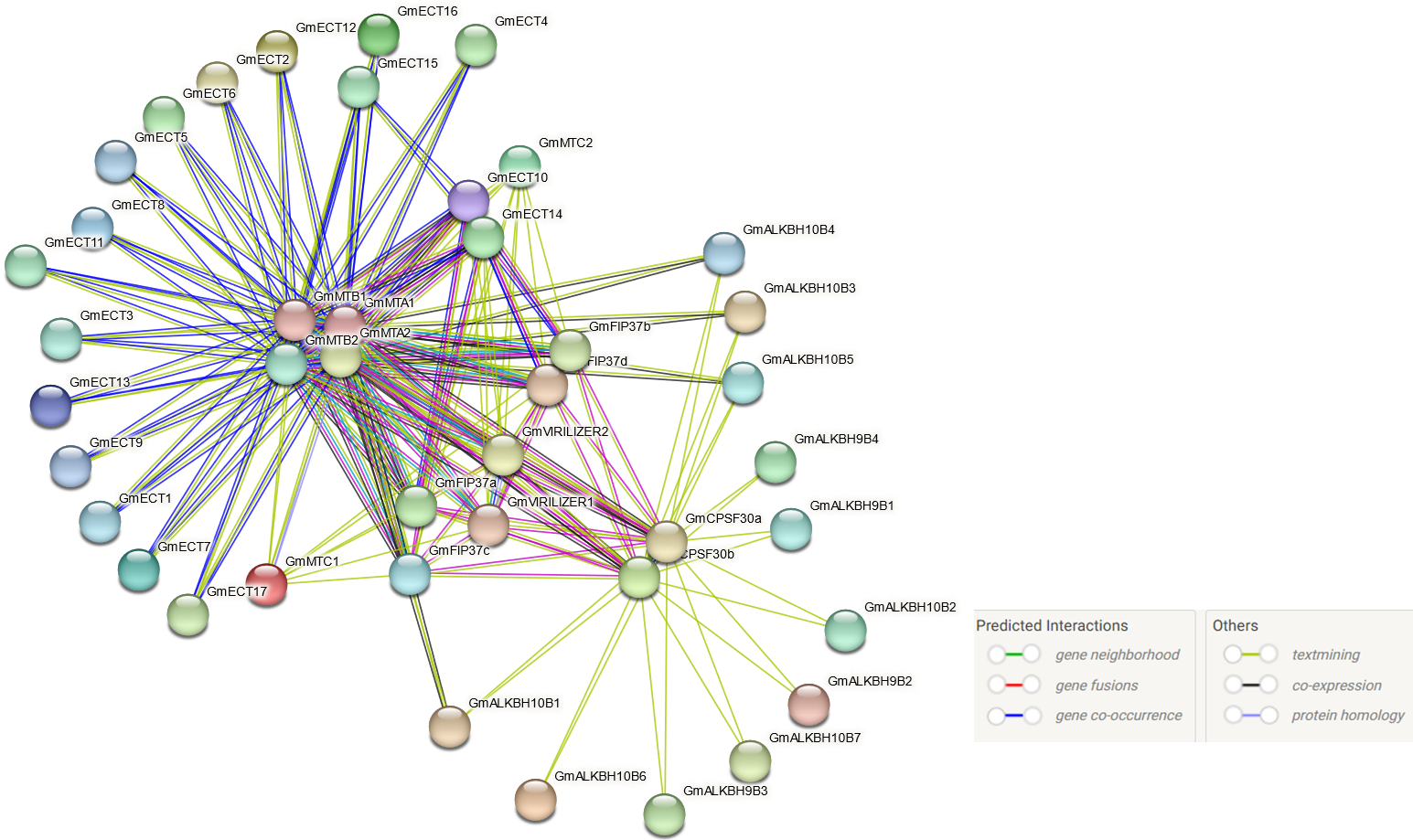
**Figure 3:** Phylogenetic analysis of m6A regulatory genes from six dicotyledons, six monocotyledons, and one Red Algae. **A.** Phylogenetic tree of writer genes. **B.** Phylogenetic tree of m6A eraser genes. **C.** Phylogenetic tree of m6A reader genes.



**Figure 4:** The distribution of cis-acting elements in the promoter regions of m6A regulatory genes. The 1000 bp 5′ upstream region of all the identified m6A genes was retrieved and analyzed using the PlantCARE database to identify the presence and number of cis-acting regulatory elements. The presence of selected motifs (x-axis) in the promoter of corresponding genes (Y-axis) was represented by the number of motifs in the heatmap.

****

**Figure 5:** **A.** Bubble diagram illustrating the enrichment of the gene ontology for the category of biological processes. **B.** A bubble chart for the Gene Ontology enrichment analysis term for the category of cellular components **C.** Bubble diagram for the term in the molecular function category from the analysis of Gene Ontology enrichment.

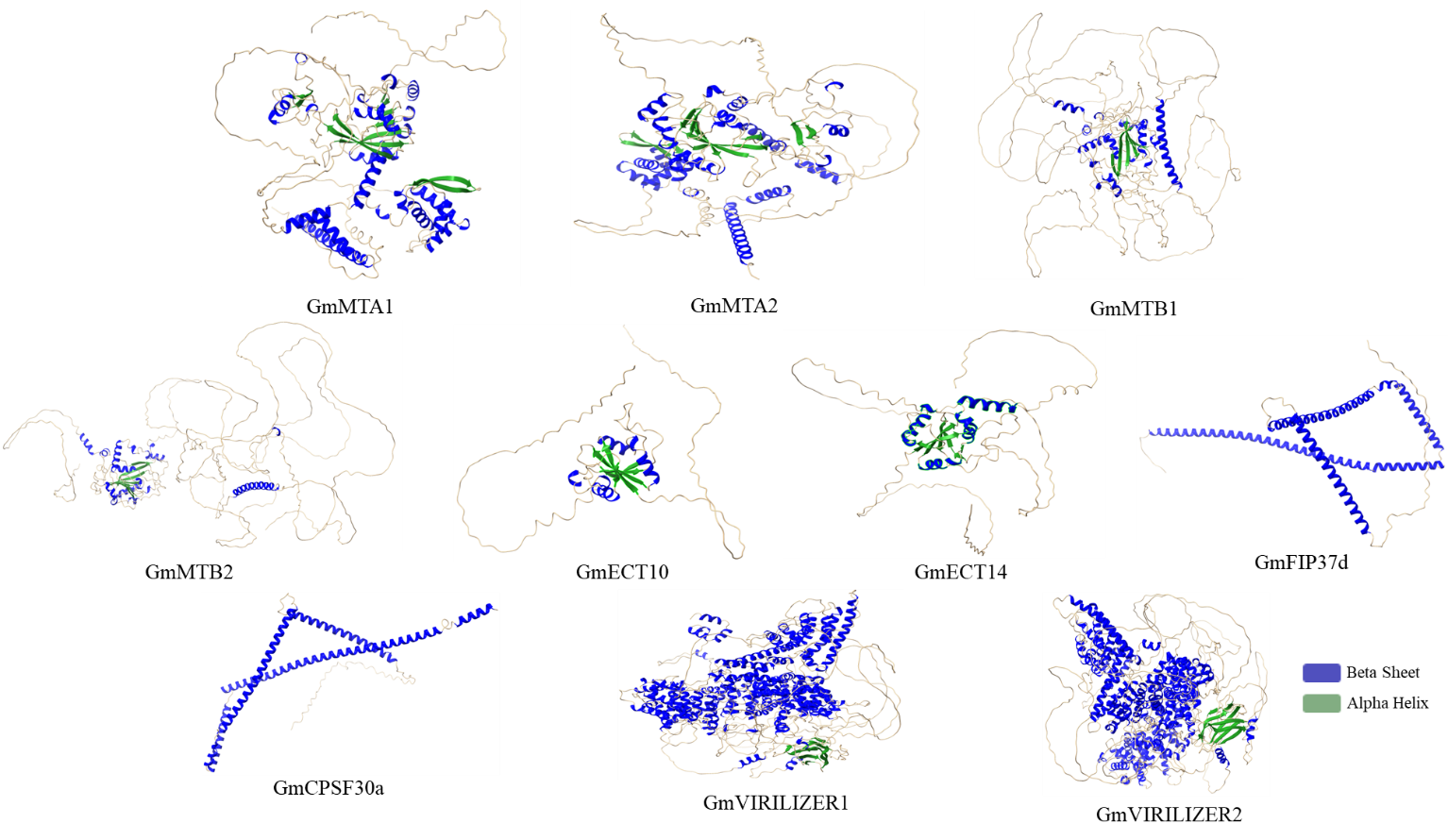
****

**Figure 6:** PPI network of 6A regulatory genes. PPI networks of writer eraser and reader genes were generated using the STRING database. Nodes represent the proteins and edges represent the protein-protein associations.

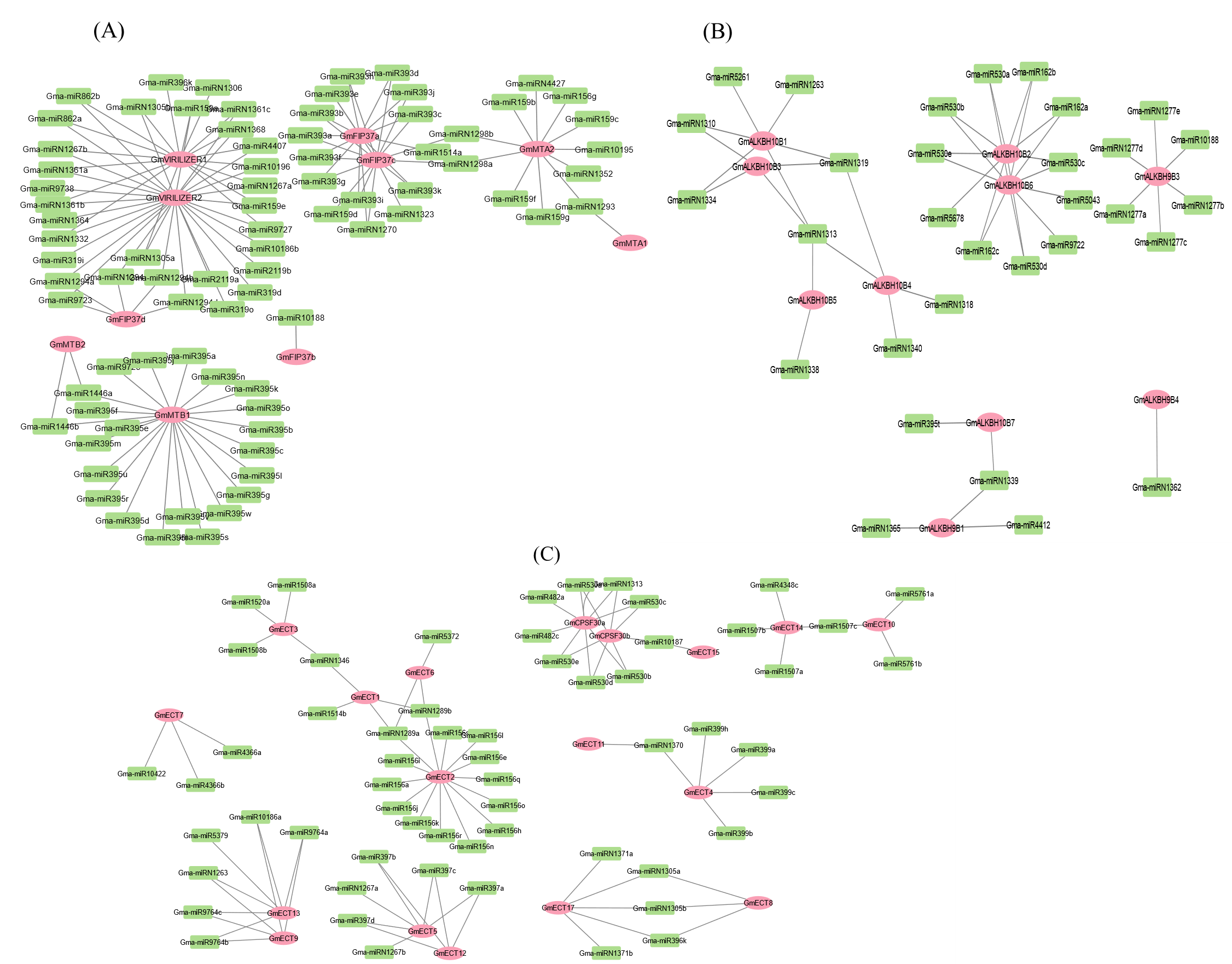
A computer screen shot of a diagram

Description automatically generated

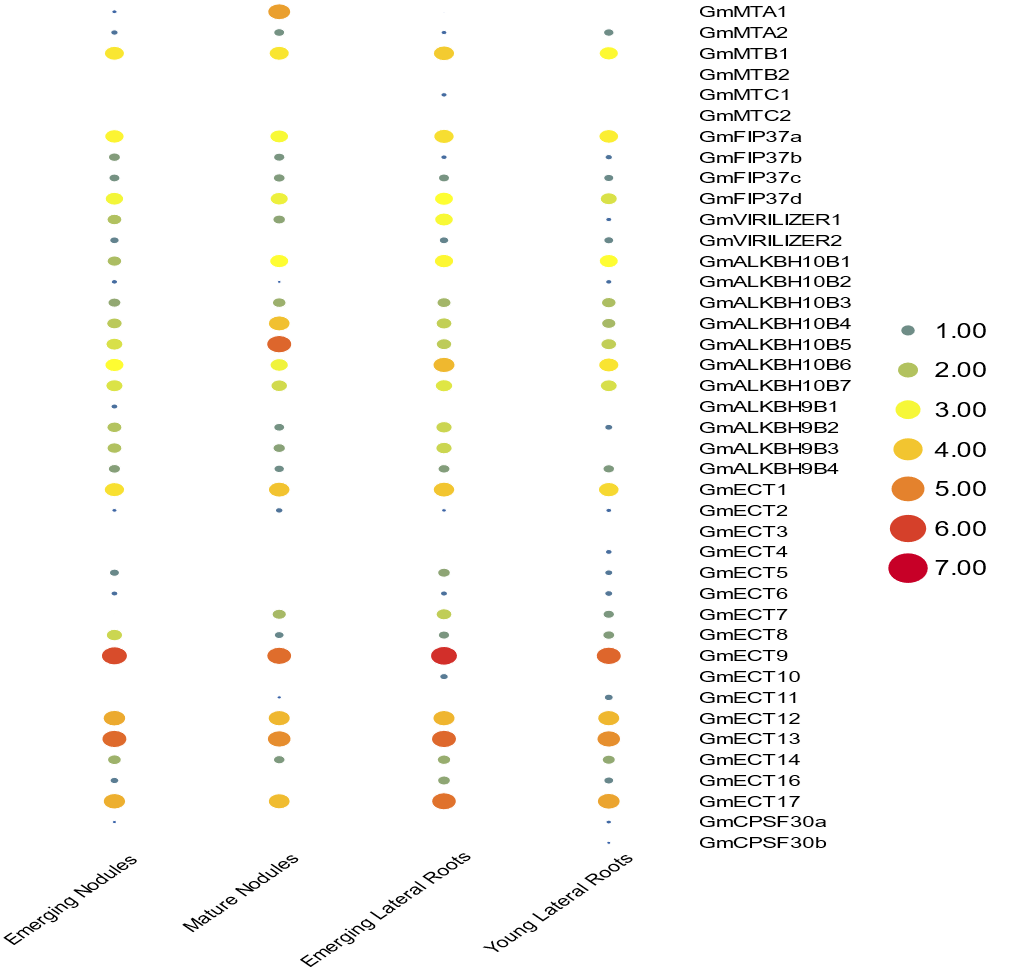
**Figure 7:** Cluster and hub genes identification in soybeans. Two clusters **(A and B)** were identified using MCODE. **C.** The top 10 hub genes were extracted by the Density of Maximum Neighborhood Component method in the cytohubba plug-in in Cytoscape.

****

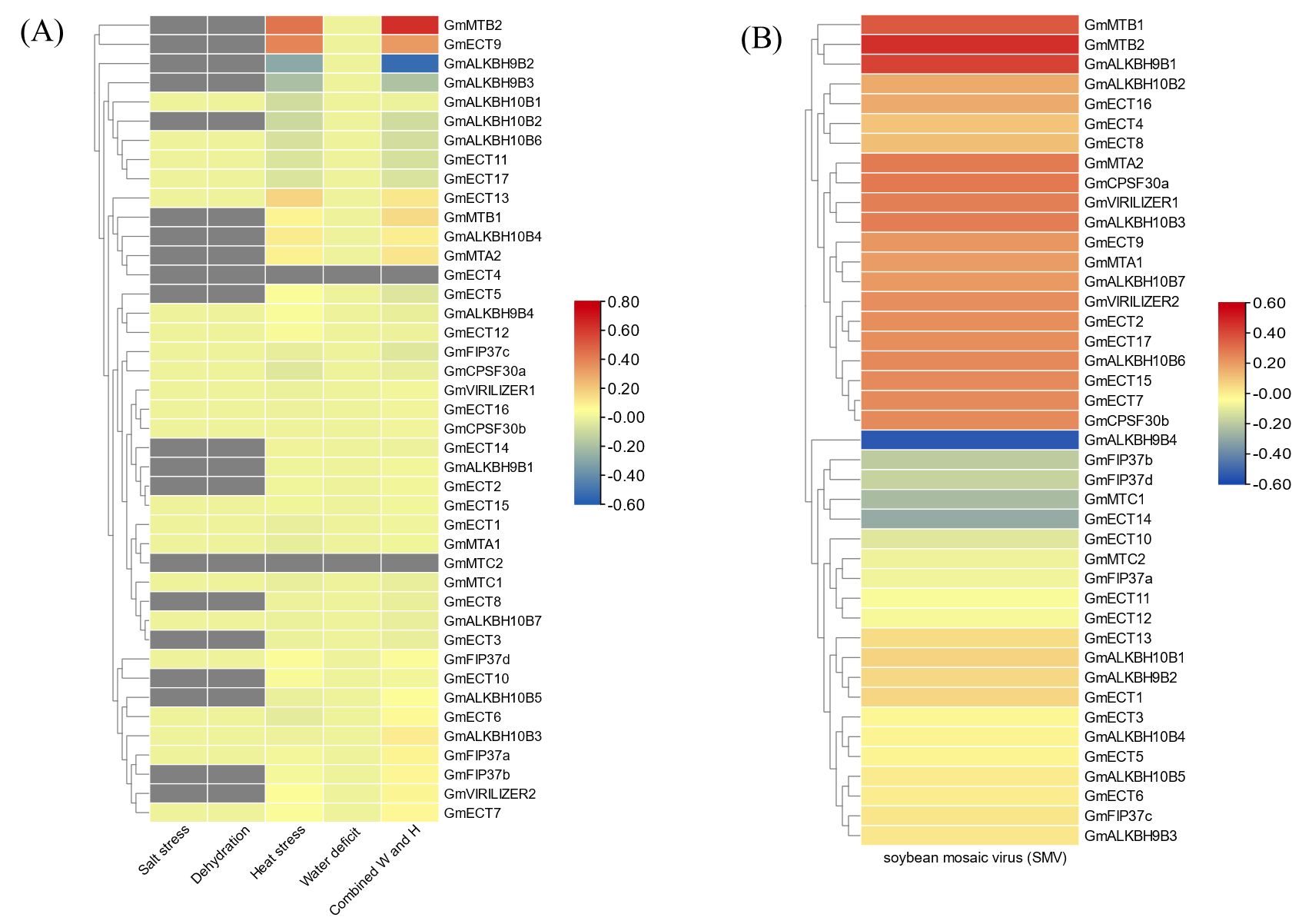
**Figure 8:** 3D structure prediction of 10 hub genes using SWISS-MODEL. Blue and green colors represent alpha helices and beta sheath respectively.

****

**Figure 9:** Genome-wide miRNA-regulated networks of the writer**(A)** eraser**(B)** and reader**(C)**. green nodes: miRNAs. pink nodes: genes that may be miRNA targets. black edges: correlations.



**Figure 10:** expression profiling of m6A regulatory genes in roots and nodules of soybean. The colour scale provided at the right of the figure represents the level of expression.



**Figure 11:** expression profiling of m6A regulatory genes under abiotic **(A)** and biotic stress**(B).** The colour scale provided at the right of the figure represents the level of expression. The stress-induced upregulation and down-regulation of all the transcripts are indicated by the red and blue colour, respectively.