



RESEARCH ARTICLE

Genome-wide analysis of MADS-box gene family in sesame

M Sandhya¹, E Priyadharshini¹, G Senthilraja^{1*}, T Anand¹, M Sudha², G Karthikeyan¹, M Raveendran²,
N Manivannan³ & M Jayakanthan²

¹Department of Plant Pathology, Centre for Plant Protection Studies, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

²Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

³Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

*Correspondence email - gsr.path@gmail.com

Received: 17 February 2026; Accepted: 11 April 2026; Available online: Version 1.0: 08 May 2026; Version 2.0: 19 May 2026

Cite this article: Sandhya M, Priyadharshini E, Senthilraja G, Anand T, Sudha M, Karthikeyan G, Raveendran M, Manivannan N, Jayakanthan M. Genome-wide analysis of MADS-box gene family in sesame. Plant Science Today. 2026; 13(2): 1-12. <https://doi.org/10.14719/pst.14136>

Abstract

Sesame (*Sesamum indicum* L.) is a major oilseed crop that plays a vital role in meeting the nutritional and edible oil demands of the global population. MADS-box genes represent an important class of transcription factors that regulate gene expression by modulating transcription initiation and interacting with specific cis-regulatory elements, thereby influencing plant growth, development and reproductive processes. However, detailed information on the MADS-box gene family in sesame remains limited, highlighting the need for in-depth genomic investigation. The present study aimed to identify and characterise MADS-box genes in sesame at the genome level. Multiple bioinformatics approaches were applied, including gene identification, structural characterisation, chromosomal mapping, domain prediction, conserved motif analysis, phylogenetic classification and cis-regulatory element exploration. The analysis led to the identification of 42 MADS-box genes in the sesame genome, which were classified into four groups: 25 MIKCC-type, 3 MIKC*-type, 12 Mα-type and 2 MY-type, indicating the predominance of type II genes. Functional annotation suggested that these genes are involved in diverse biological processes, particularly in plant growth, floral development and regulatory pathways.

Keywords: MADS-box genes; Mα-type; MY-type; MIKC*-type; MIKC^c-type; *Sesamum indicum*; transcription factors

Introduction

Sesame (*Sesamum indicum* L.) is an important oilseed crop belonging to the Pedaliaceae family. The genus *Sesamum* comprises several species, most of which are wild relatives distributed predominantly in sub-Saharan Africa, which is considered as the primary centre of diversity for the genus. However, the cultivated species *S. indicum* is believed to have originated in the Indian subcontinent. Sesame is considered the oldest oil crop, which is mainly grown in the tropics (1). It contributes significantly to the daily oil and protein needs of nearly half the global population. Sesame is often referred to as the 'queen of oilseeds' due to its high oil content, superior oil quality and oxidative stability (2). The proportion of PUFA (polyunsaturated fatty acids) in sesame oil is around 80 %, which is beneficial for human health, thereby surpassing all other major oil crops (3). Sesame is recognised for its short life cycle, high propagation rate, adaptability to diverse growing environments and substantial seed to oil ratio (4).

Genome sequencing of sesame led to the identification of 24148 genes, ushering in a new era of functional genomics research (4). Despite this advancement, the roles of many sesame genes remain unclear and few gene families have been thoroughly examined. This MADS-box gene family, which is found in animals, fungi and plants, is of particular importance (5). The acronym MADS

is derived from its founding members: MCM1 from yeast, AG from *Arabidopsis thaliana*, DEF from snapdragon and SRF from humans (6–9). MADS-box proteins are distinguished by a DNA-binding domain of about 60 amino acids located at the N-terminus. These proteins are recognised as transcription factors involved in developmental processes, particularly in floral development (10). These MADS-box genes are known to play a crucial role in the "ABCDE model" of floral organ development, with different functional genes contributing to the formation of sepals, petals, stamens, carpels and ovules (11–13). MADS-box genes are divided into two types: types I and II (14). Most type II plant genes are present, which are in turn classified into three additional plant-specific domains, namely the I (Intervening), K (Keratin-like) and C (C-terminal) domains. These domains contribute to the structural organisation and functional specificity of the encoded proteins. (15). Type I genes, also known as M-type, lack the K-domain (16). Type II genes are further divided into MIKC^c-type and MIKC*-type. Among type II genes, MIKC* have longer domains and more exons (17). MIKC^c-type genes have been more extensively studied than other types and are classified into 11 subfamilies based on structural differences (5). Type I genes are subdivided into Mα, Mβ and MY subgroups and generally experience more rapid natality and mortality when compared to type II genes in most plants (5, 18).

Although the sesame genome sequence is available and the crop holds significant importance as an oilseed, the comprehensive characterisation of the MADS-box gene family in *S. indicum* remains limited. This study presents a genome-wide analysis of the MADS-box gene family in *S. indicum*, with a total of 42 genes identified and distributed across 13 chromosomes. These genes were classified into type I and type II groups, followed by phylogenetic comparison with *A. thaliana* to determine evolutionary relationships. Gene structure analysis revealed that type II genes possess more complex exon-intron organisation, while motif and domain analyses indicated functional conservation, particularly among MIKCC-type genes, whereas type I genes exhibited simpler structures, suggesting functional divergence. We hypothesise that the structural organisation, phylogenetic relationships and regulatory features of MADS-box genes in *S. indicum* contribute to their functional roles in plant growth and development, while also reflecting species-specific evolutionary adaptations. These findings highlight the important roles of MADS-box genes in plant growth and development and provide a valuable foundation for future functional characterisation and crop improvement strategies in sesame.

Materials and Methods

Identification of MADS-box genes

The sesame genome and proteome were obtained from the NCBI website (4). provided the hidden Markov model (HMM) profiles for the SRF (type I) domain (PF00319). Using HMMER v. 3.0, MADS-box genes were identified in the sesame proteome by employing the HMM profile corresponding to the Pfam MADS-box family PF00319 (19). This process yielded 42 MADS-box proteins as potential MADS-box genes. To verify the accuracy of these predicted genes, their amino acid sequences were analysed using the simple modular architecture research tool (SMART) to confirm the presence of the MADS-box domain in all genes (20).

Assigning the location of MADS-box genes to the sesame genome

Using a GFF file obtained from the NCBI database, the physical locations of MADS-box genes were mapped to the 16 chromosomes of the sesame genome (4). An online tool called MG2C was used for visualising genetic maps, utilising PERL and SVG languages. This tool is compatible with various operating systems, with MG2C2.1 being the most recent version (21).

Alignment and phylogenetic analysis of MADS-box genes

Sesamum indicum, *A. thaliana* MADS-box proteins were downloaded from Ensembl plant and TAIR databases respectively. Sesame MADS-box proteins were aligned using ClustalW with *A. thaliana* MADS-box proteins (22). Unrooted maximum-likelihood (ML) trees of *Sesamum* and *Arabidopsis* MADS-box proteins were constructed using the MEGA11 software (23). iTOL software was used for further visualisation. The conserved domain analysis was also performed using TBtools to identify the domains associated with the MADS-box gene family of *S. indicum*.

Comparative synteny gene analysis of *Sesamum indicum* and *Arabidopsis thaliana*

To investigate evolutionary relationships and similarities between *S. indicum* and *A. thaliana*, a synteny analysis was conducted using the Circoletto web server (24, 25).

Conserved motif analysis

The MEME software was used to predict conserved motifs present within 42 MADS-box genes of *S. indicum* with the following parameters: number of repetitions -any, maximum number of motifs -20, optimum motif width set to ≥ 6 and ≤ 200 (26).

Cis-regulatory elements identification in the promoter region

In *S. indicum*, the promoter regions (1 kb) were retrieved from Ensembl plants for every single gene of the *S. indicum* SRF domain. For *in silico* analysis of promoter sequences to find cis-regulatory elements in the *S. indicum* genome, the PlantCARE database was used, after retrieving 1000 bp of promoter regions (27). Cis-regulatory element data for all genes were compiled and used to generate a heatmap, which was visualised using TBtools (28).

Subcellular localisation of *Sesamum indicum* MADS-box gene family

Subcellular localisation was predicted using the WoLFPSORT (28). To generate a heatmap, all subcellular localisation data were noted for each gene. The data was visualised as a heatmap by using the TB tool (28).

Results

Identification of MADS-box genes in the sesame genome

Using SRF domain protein sequences as queries in HMM profiling against the sesame protein dataset, a total of 42 MADS-box proteins were identified. Detailed information, including locus IDs, transcripts, chromosomal positions, protein length, molecular weight, isoelectric point and GRAVY values, is provided in Tables 1 and 2. The protein length ranged from 186 to 13280 bp, molecular weight ranged from 6896.09–49784.67 k Da, isoelectric point varied between 4.79 and 10.64 and GRAVY values ranged from -0.21 to -0.929, indicating variability in physicochemical properties.

Genomic localisation and mapping of SRF genes in *Sesamum indicum* by using the Mg2C tool

The identified MADS-box genes were mapped onto 13 chromosomes of the *S. indicum* genome (Fig. 1). The distribution was uneven, with the highest number of genes located on chromosome 8, followed by chromosomes 2 and 6. Chromosomes 5, 7, 10 and 14 contained a single gene each, while chromosomes 1, 4, 12, 13 and 15 contained two genes each. No genes were identified on chromosomes 9, 11 and 16. In several chromosomes, genes were observed in clusters, suggesting possible duplication events.

Conserved domain analysis

Domain analysis using NCBI CDD and TBtools identified a total of 18 domains among the 42 MADS-box proteins. Among these, the SRF-TF domain and K domain superfamily were predominantly present, indicating their importance in transcriptional regulation (Fig. 2).

Comparative phylogenetic tree of *Sesamum indicum* and *Arabidopsis thaliana*

To investigate evolutionary relationships, a maximum likelihood phylogenetic tree was constructed using MADS-box protein sequences from *S. indicum* and *A. thaliana* (Fig. 3). The analysis revealed a higher number of type II genes compared to type I genes in sesame. A total of 14 type I genes were identified, including 12 Ma and 2 My subgroups, while M β -type genes were absent. Additionally, 28 type II genes were identified, comprising 25 MIKCC and 3 MIKC* members.

Table 1. List of MADS-box genes identified in the sesame genome

Gene Name	Chromosome	Exon no	ORF length (bp)	Gene length (bp)	Group	Subfamily
SIN_1011302	3	6	579	1297	MIKC ^c	ANR1
SIN_1019617	8	2	741	832	Mα	NA
SIN_1001798	12	2	765	948	Mα	NA
SIN_1023144	8	1	1110	1110	Mα	NA
SIN_1004920	15	7	657	3940	MIKC ^c	E
SIN_1023145	8	1	654	654	Mα	NA
SIN_1020687	6	7	639	1956	MIKC ^c	B-PI
SIN_1001401	3	9	1086	4194	MIKC [*]	NA
SIN_1015610	6	1	189	189	MIKC ^c	SOC1
SIN_1023147	8	1	909	909	Mα	NA
SIN_1009098	3	7	705	1880	MIKC ^c	B-AP3
SIN_1026469	8	2	251	676	MIKC ^c	E
SIN_1022712	8	2	981	2270	Mα	NA
SIN_1024006	2	10	918	2861	MIKC [*]	NA
SIN_1026116	10	8	690	6574	MIKC ^c	SVP
SIN_1025213	15	8	690	6574	MIKC ^c	SVP
SIN_1020623	6	1	1368	1368	Mα	NA
SIN_1004662	13	1	654	654	Mα	NA
SIN_1017763	1	1	186	186	MIKC ^c	SOC1
SIN_1015232	3	1	192	192	MIKC ^c	ANR1
SIN_1018628	4	3	333	1700	MIKC ^c	ANR1
SIN_1020624	6	1	1146	11466	Mα	NA
SIN_1026472	8	2	267	591	MIKC ^c	FLC
SIN_1022802	8	1	714	714	Mγ	NA
SIN_1017236	2	22	2154	11023	MIKC ^c	AGL12
SIN_1017990	2	7	594	13280	MIKC ^c	FLC
SIN_1022876	8	1	741	741	Mα	NA
SIN_1011564	7	7	654	2080	MIKC ^c	B-AP3
SIN_1019911	8	5	513	5693	MIKC ^c	A
SIN_1004098	13	1	306	306	MIKC ^c	SVP
SIN_1026382	8	8	774	8555	MIKC ^c	A
SIN_1022147	3	6	546	1848	MIKC ^c	C/D
SIN_1001800	12	1	648	648	Mα	NA
SIN_1018255	2	6	645	4150	MIKC ^c	A
SIN_1019978	5	1	216	216	MIKC [*]	NA
SIN_1022400	6	1	234	234	MIKC ^c	C/D
SIN_1022844	8	2	615	1056	Mα	NA
SIN_1019318	14	2	243	420	MIKC ^c	C/D
SIN_1013997	1	7	726	6932	MIKC ^c	SOC1
SIN_1018254	2	8	717	3906	MIKC ^c	E
SIN_1003343	4	1	738	738	Mγ	NA
SIN_1013259	2	6	786	2056	MIKC ^c	BS

Table 2. Detailed properties of identified *Sesamum indicum* MADS-box genes

Transcript ID	Gene Name	Chromosome	Location start-end	Strand	Protein molecular weight (K Da)	pI	GRAVY
SIN_1011302.t	SIN_1011302	3	20931768-20931908	Forward	21829.91	9.3	-0.64
SIN_1019617.t	SIN_1019617	8	559875-560015	Reverse	27672.08	4.79	-0.61
SIN_1001798.t	SIN_1001798	12	5001128-5001268	Forward	28406.72	4.9	-0.586
SIN_1023144.t.cds1	SIN_1023144	8	18883412-18883552	Reverse	40637.62	8.33	-0.929
SIN_1004920.t	SIN_1004920	15	2339948-2340088	Forward	25216.66	8.78	-0.71
SIN_1023145.t.cds1	SIN_1023145	8	18886199-18886339	Reverse	24503.91	9.1	-0.465
SIN_1020687.t	SIN_1020687	6	17833556-17833696	Reverse	24845.44	6.46	-0.744
SIN_1001401.t	SIN_1001401	3	2637594-2637701	Reverse	41320.92	6.03	-0.557
SIN_1015610.t.cds1	SIN_1015610	6	9069730-9069870	Forward	7131.35	10.21	-0.306
SIN_1023147.t.cds1	SIN_1023147	8	18892925-18893065	Reverse	32915.15	6.14	-0.296
SIN_1009098.t	SIN_1009098	3	23434270-23434410	Forward	27095.85	9.4	-0.778
SIN_1026469.t	SIN_1026469	8	12855864-12856004	Reverse	9493.08	10.02	-0.21
SIN_1022712.t	SIN_1022712	8	15888837-15888977	Reverse	36055.41	9.046	-0.82
SIN_1024006.t	SIN_1024006	2	7661744-7661884	Reverse	34826.46	5.22	-0.561
SIN_1026116.t	SIN_1026116	10	15027474-15027596	Forward	26008.85	9.25	-0.554
SIN_1025213.t	SIN_1025213	15	8474571-8474711	Reverse	25974.37	5.69	-0.73
SIN_1020623.t.cds1	SIN_1020623	6	18247510-18247629	Reverse	49784.67	4.93	-0.86
SIN_1004662.t.cds1	SIN_1004662	13	2472516-2472656	Reverse	24412.81	8.94	-0.444
SIN_1017763.t	SIN_1017763	1	5787417-578755	Reverse	6896.09	10.37	-0.28
SIN_1015232.t.cds1	SIN_1015232	3	15990328-15990468	Forward	7067.24	10.21	-0.256
SIN_1018628.t	SIN_1018628	4	2653797-2653937	Reverse	12477.28	8.76	-0.506
SIN_1020624.t.cds1	SIN_1020624	6	18243482-18243601	Reverse	42211.83	6.39	-0.698
SIN_1026472.t	SIN_1026472	8	12841668-12841808	Reverse	9988.65	9.8	-0.341
SIN_1022802.t.cds1	SIN_1022802	8	16549895-16550014	Reverse	26804.67	8.99	-0.651
SIN_1017236.t	SIN_1017236	2	400583-400717	Reverse	80866.97	7.15	-0.318
SIN_1017990.t	SIN_1017990	2	18414228-18414368	Reverse	22175.59	6.51	-0.314
SIN_1022876.t.cds1	SIN_1022876	8	17167560-17167700	Forward	27314.21	5.32	-0.439
SIN_1011564.t	SIN_1011564	7	3533865-3534005	Forward	25136.7	9.3	-0.743
SIN_1019911.t	SIN_1019911	8	3081373-3081513	Reverse	19785.92	9.72	-0.665
SIN_1004098.t.cds1	SIN_1004098	13	3957211-3957351	Reverse	11708.82	8.93	0.284
SIN_1026382.t	SIN_1026382	8	13738633-13738773	Forward	29810.67	8.33	-0.878
SIN_1022147.t	SIN_1022147	3	12168993-12169133	Reverse	21188.6	9.88	-0.634
SIN_1001800.t.cds1	SIN_1001800	12	5010173-5010295	Forward	24321.89	9.15	-0.706
SIN_1018255.t	SIN_1018255	2	16807121-16807261	Reverse	24619.3	8.89	-0.689
SIN_1019978.t.cds1	SIN_1019978	5	14575782-14575922	Forward	7936.38	10.64	-0.218
SIN_1022400.t.cds1	SIN_1022400	6	23113548-23113688	Reverse	8884.13	9.93	-0.727
SIN_1022844.t	SIN_1022844	8	16869170-16869283	Reverse	22236.8	5.74	-0.548
SIN_1019318.t	SIN_1019318	14	2855413-2855553	Forward	9119.32	9.34	-0.639
SIN_1013997.t	SIN_1013997	1	17297021-17297161	Reverse	27550.32	6.57	-0.794
SIN_1018254.t	SIN_1018254	2	16815702-16815842	Reverse	27316.13	9.13	-0.695
SIN_1003343.t.cds1	SIN_1003343	4	14269538-14269657	Reverse	26900.88	9.34	-0.589
SIN_1013259.t	SIN_1013259	2	4935406-4935537	Forward	30572.92	8.43	-0.814

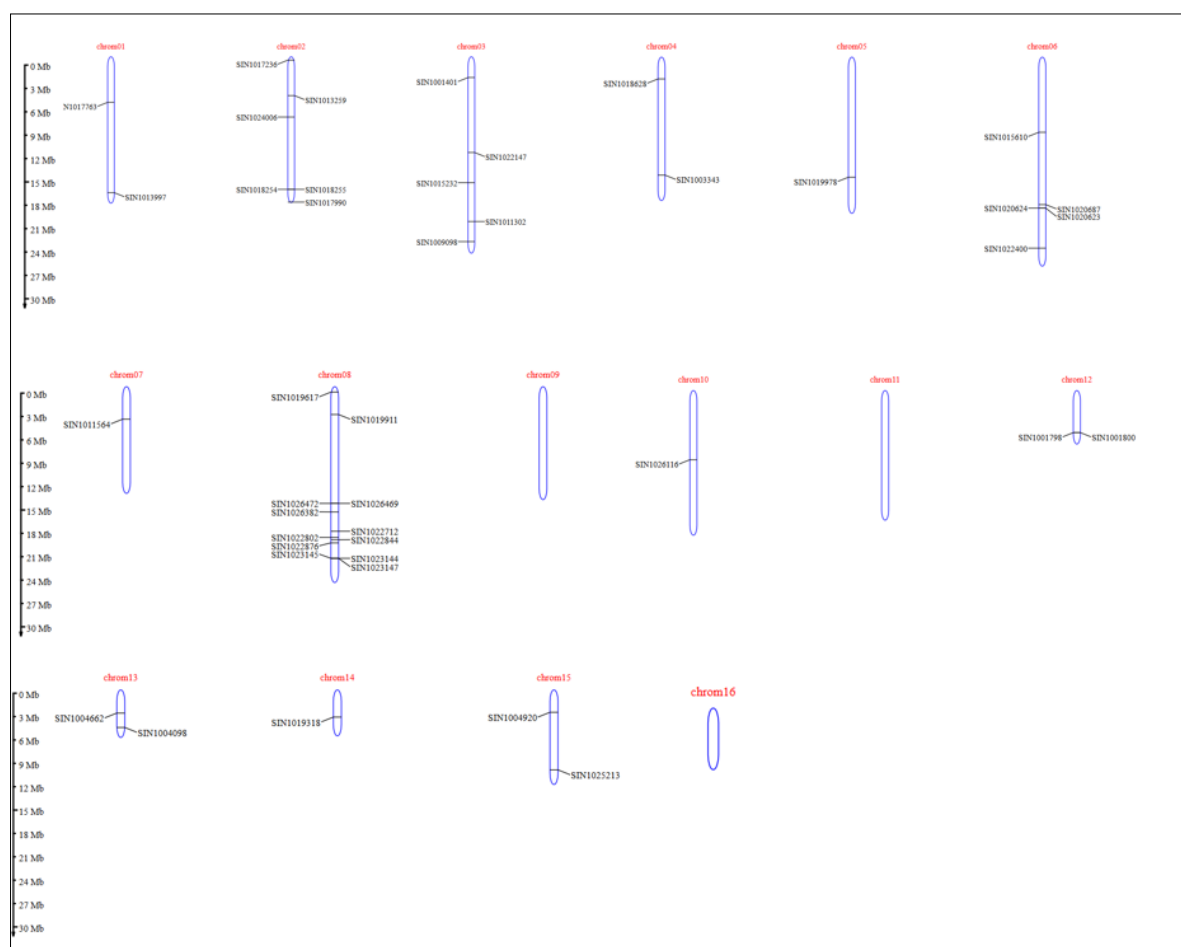


Fig. 1. Distribution of MADS-box genes in sesame. 01–16 represent chromosomes of the sesame genome. Blue bars represent chromosomes. Black lines on blue bars indicate the locations of MADS-box genes on linkage groups.

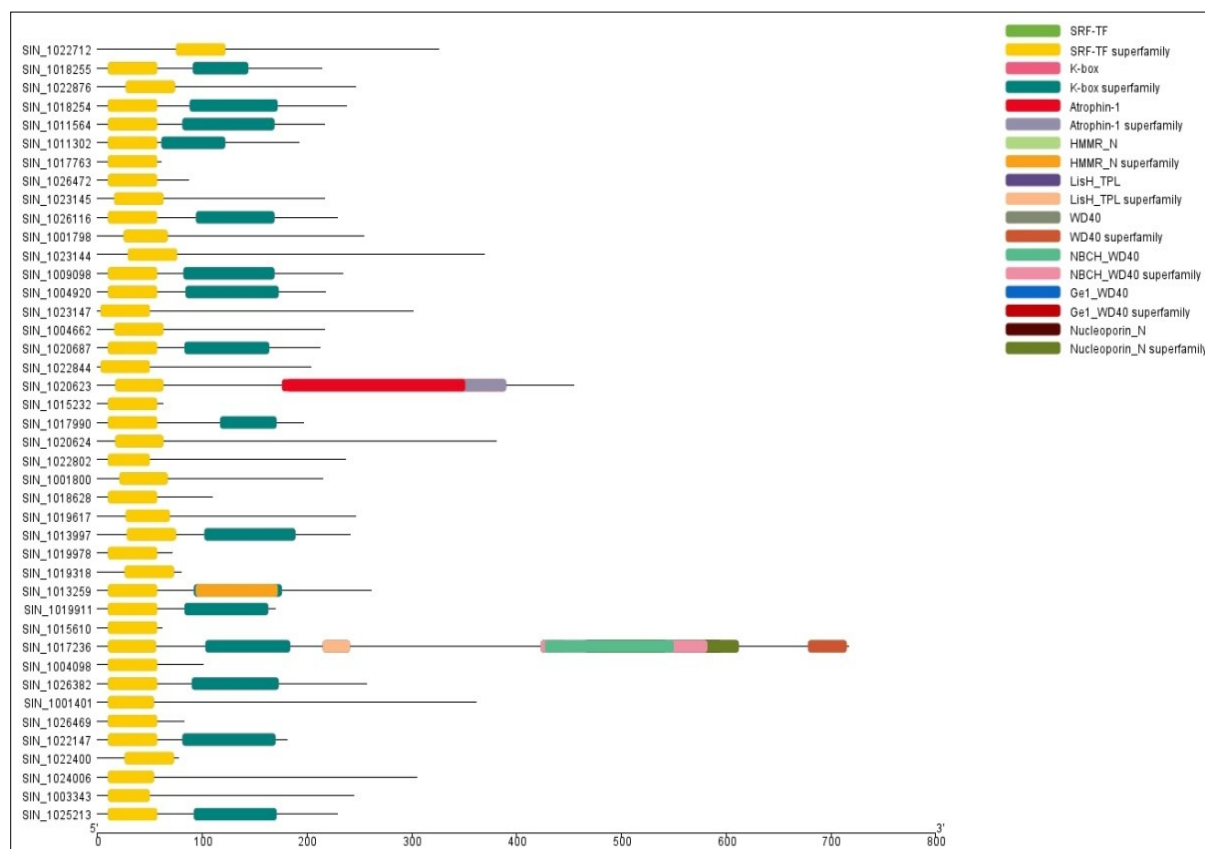


Fig. 2. Conserved domain regions of *Sesamum* SRF proteins. Yellow and green colours represent the SRF-TF and K-box superfamily respectively.

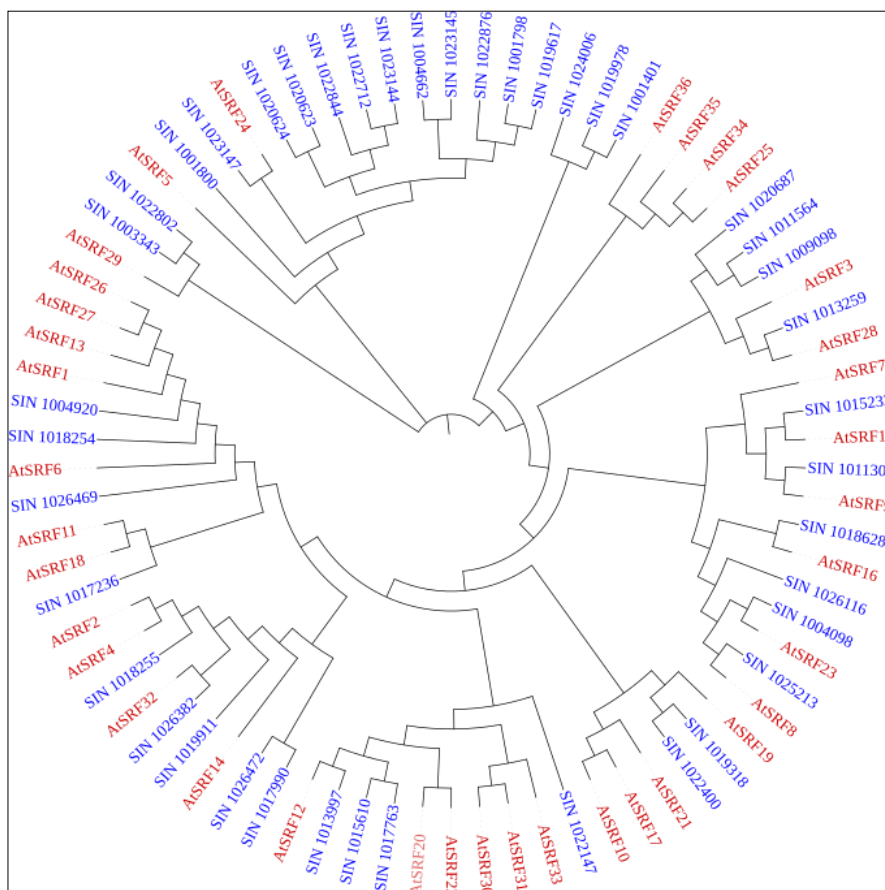


Fig. 3. Phylogenetic analysis of MADS-box proteins of *Sesamum indicum* and *Arabidopsis thaliana*.

Conserved motifs investigations of *Sesamum indicum* and *Arabidopsis thaliana*

Motif analysis using MEME identified 20 conserved motifs among the 42 MADS-box proteins (Fig. 4). Genes within the same phylogenetic group exhibited similar motif compositions. Motif 1 corresponded to the conserved MADS domain and was present in all MIK-type proteins. Motif 2 represented the K domain, which was predominantly found in type II proteins and is associated with protein-protein interactions. Motif 7 was identified as an alternative MADS domain in Ma, My and MIK* proteins.

Synteny analysis

Synteny analysis between *S. indicum* and *A. thaliana* revealed conserved gene pairs (Fig. 5). Several sesame genes showed syntenic relationships with corresponding *Arabidopsis* genes, indicating conserved genomic regions. Multiple gene duplication events were observed, suggesting their role in gene family expansion. Some gene pairs exhibited evidence of selection pressure following duplication.

Cis-regulatory elements identification in *Sesamum indicum* MADS-box genes

Promoter analysis of 1000 bp upstream regions identified 14 cis-regulatory elements across the MADS-box genes (Fig. 6 and Supplementary Table 1). These elements were categorised into stress-responsive, growth and developmental and basal regulatory elements. Certain genes, such as SIN_1023147 and SIN_1022513, showed a higher number of cis-elements, whereas others exhibited fewer regulatory elements. The distribution of these elements suggests variability in regulatory complexity among the genes (Table 3).

Sub-cellular localisation of *Sesamum indicum* MADS-box gene family

Subcellular localisation prediction indicated that MADS-box proteins are distributed across multiple cellular compartments, including the nucleus, cytoplasm, mitochondria, chloroplasts, endoplasmic reticulum, Golgi apparatus and other organelles (Fig. 7 and Supplementary Table 2). Several genes, such as SIN_1024006 and SIN_1019617, were predominantly localised in the nucleus, suggesting roles in transcriptional regulation, while others were distributed in different cellular compartments, indicating diverse functional roles.

Discussion

MADS-box genes have been identified in a wide range of plant species, including *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*, *Solanum melongena*, *Vitis vinifera*, *Malus domestica*, *Zea mays*, *Cucumis sativus*, *Glycine max*, *Solanum lycopersicum*, *Brachypodium distachyon*, *Prunus mume*, *Citrullus lanatus*, chrysanthemum, eggplant, cocoa and maize (18, 29–45). The number and diversity of MADS-box genes vary considerably across species, ranging from 43 in cucumber to over 200 in maize, reflecting differences in genome size, duplication events and evolutionary history. In *S. indicum*, a comparatively smaller number of MADS-box genes were identified, including 14 type I members belonging to the Ma and My subfamilies, while the Mβ subfamily was absent. This number is lower than that reported in several species such as rice, maize, poplar, *Prunus mume*, apple, *Arabidopsis*, chrysanthemum, cocoa and soybean, but higher than in *B. distachyon*, watermelon and cucumber. Such variation is likely associated with genome expansion events, including gene duplication and transposon

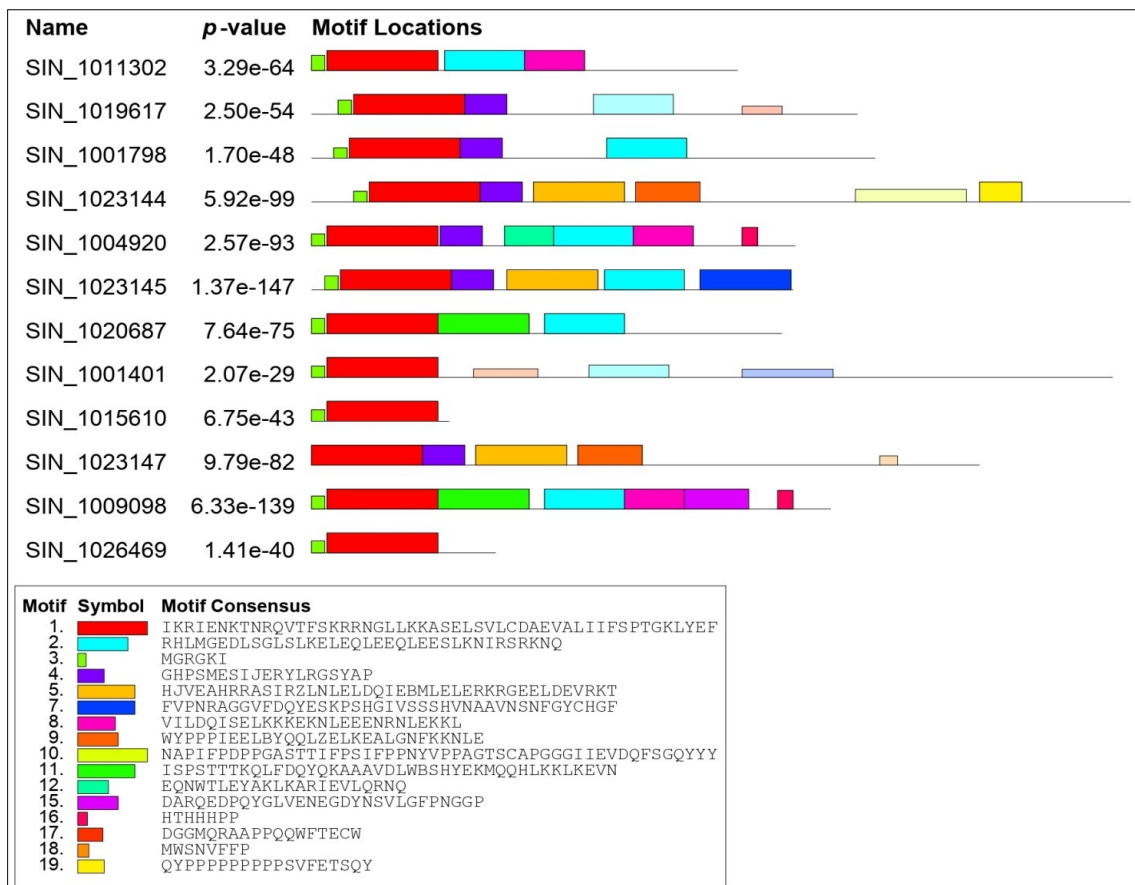


Fig. 4. Conserved motif analysis of sesame MADS-box proteins. The genes are on the left of the motif bars. Each motif is represented by a number in a colored box. Box length corresponds to motif length.

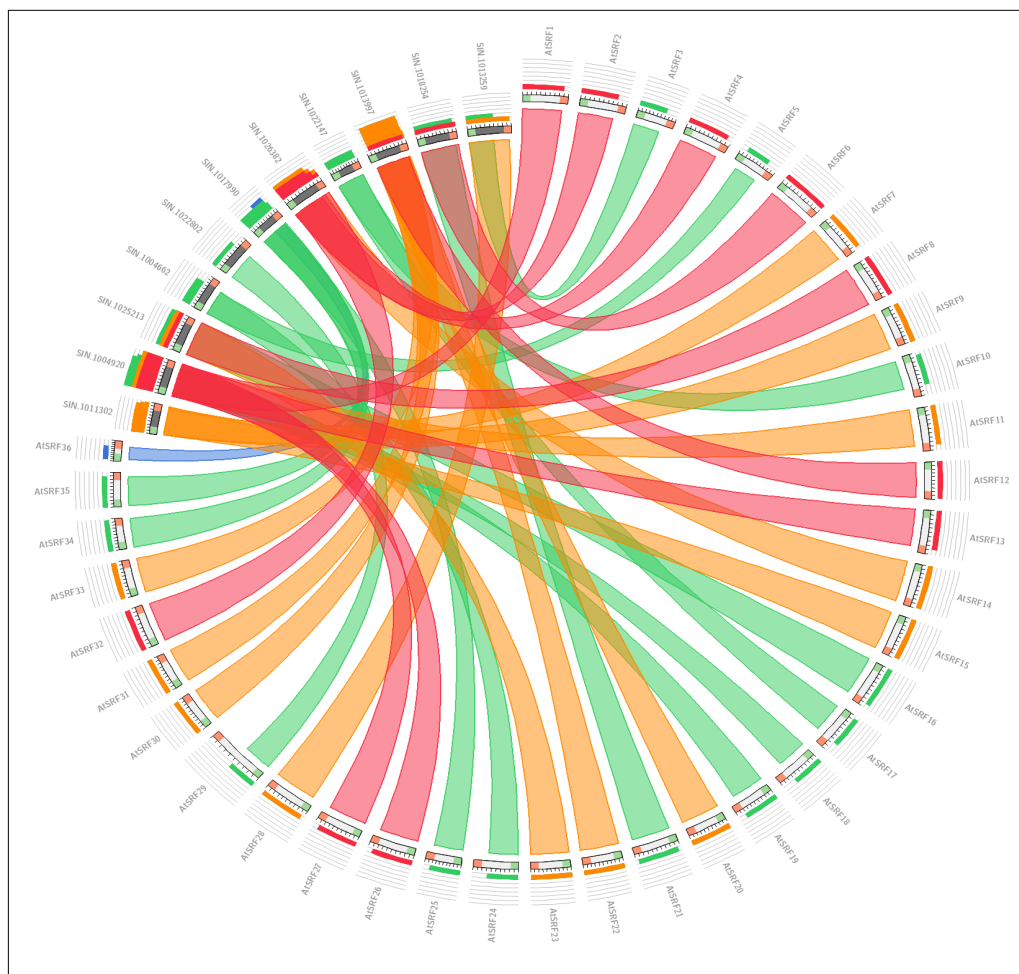


Fig. 5. The syntenic regions in selected *S.indicum* SRF proteins by taking the *A.thaliana* SRF proteins as the query sequence. Blue, green, orange and red show 50, 75, 99.99 and 100 % similarity respectively.

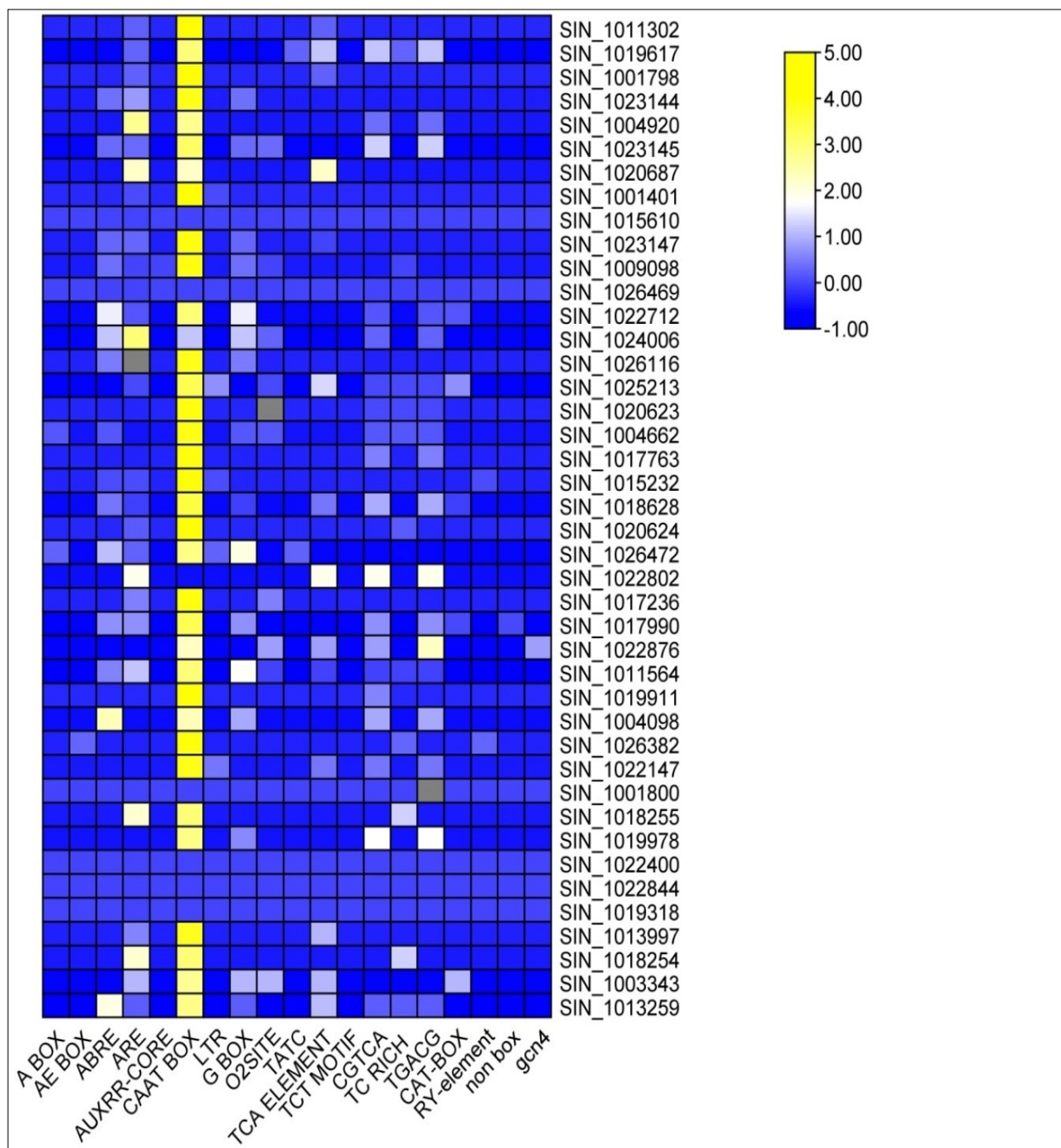


Fig. 6. The heatmap showing the presence or absence of cis-regulatory elements on the upstream promoter region of a gene.

Table 3. List of cis elements with corresponding regulatory functions

Cis element	Function
A BOX	Cis-acting regulatory element
ABRE	Cis-acting element involved in the abscisic acid responsiveness
ARE	Cis-acting regulatory element essential for the anaerobic induction
LTR	Cis-acting element involved in low-temperature responsiveness
G BOX	Cis-acting regulatory element involved in light responsiveness
O2 SITE	Cis-acting regulatory element involved in zein metabolism regulation
TATC	Cis-acting element involved in gibberellin-responsiveness
TCA ELEMENT	Cis-acting element involved in salicylic acid responsiveness
TCT MOTIF	Part of a light-responsive element
CGTCA	Cis-acting regulatory element involved in the MeJA-responsiveness
TC RICH	Cis-acting element involved in defence and stress responsiveness
TGACG	Cis-acting regulatory element involved in the MeJA-responsiveness
CAT BOX	Cis-acting regulatory element related to meristem expression
RY ELEMENT	Cis-acting regulatory element involved in seed-specific regulation
GCN4	Cis-regulatory element involved in endosperm expression

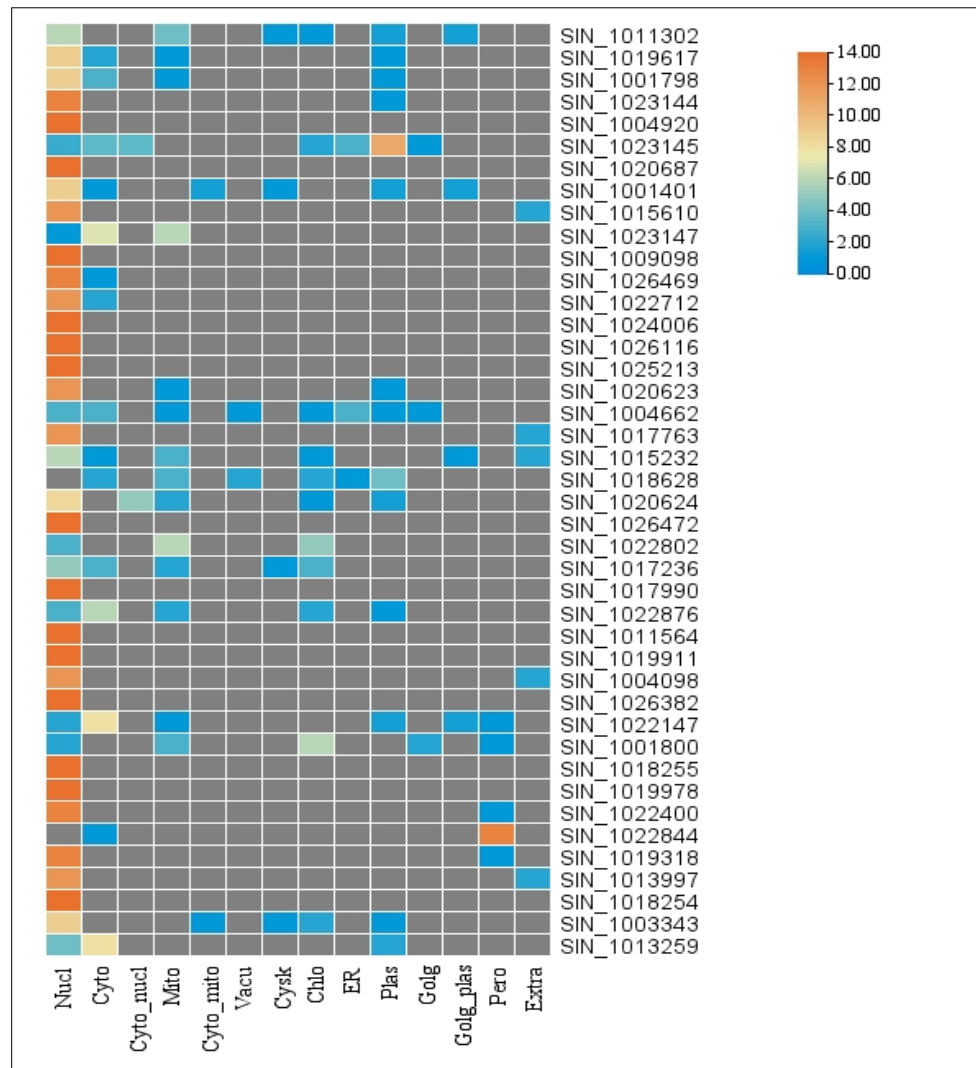


Fig. 7. Heat map representation of sesame MADS-box gene expression. The genes are on the right of the expression bars with domains represented at the bottom.

activity, whereas the comparatively smaller genome size of sesame may account for the reduced number of both type I and type II MADS-box genes (14).

The classification of sesame MADS-box genes into type I and type II groups is consistent with observations in other plant species, indicating evolutionary conservation of this gene family (46). However, the absence of specific subfamilies, particularly M β and AGL6, is noteworthy. M β -type genes, which are present in *Arabidopsis* and *S. lycopersicum* (18, 29–32), have been implicated in female gametophyte development, although their functions remain only partially understood (47). Sesame contains 28 type II MADS-box genes comprising 25 MIKCc and 3 MIKC* members, which are also reported in other species such as *Arabidopsis* and *S. lycopersicum*, although their numbers are comparatively lower than in *Paulownia fortunei* (48). Similarly, AGL6-clade genes, which play important roles in floral organ identity and the transition from vegetative to reproductive growth through interaction with FLC and FT pathways (49–51), were not identified in sesame. The absence of these subfamilies suggests lineage-specific gene loss or functional divergence and indicates that sesame may utilise alternative regulatory pathways to control floral development and flowering time.

Gene structure analysis revealed that type II MADS-box genes in sesame possess a higher number of exons compared to type I genes, a pattern consistent with previous reports in

Arabidopsis, rice, maize and other plant species (18, 29, 37). The conserved exon-intron organisation and motif composition observed among type II genes suggest strong functional conservation, particularly in domains responsible for DNA binding and protein-protein interactions. In contrast, type I genes, which typically contain fewer or no introns, may have undergone rapid evolutionary changes, contributing to their functional diversification. These structural differences indicate that type II genes are likely under stronger evolutionary constraints, whereas type I genes exhibit greater plasticity. Multiple replication events observed among gene pairs suggest their origin through duplication or tandem duplication, while selection pressure appears to have influenced their subsequent evolution (34).

In addition to gene structure, the analysis of conserved motifs and cis-regulatory elements provides further insights into the regulatory potential of these genes. The presence of diverse cis-regulatory elements associated with stress responsiveness and developmental processes suggests that sesame MADS-box genes are subject to complex transcriptional control. Genes with a higher density and diversity of cis-elements may exhibit more dynamic expression patterns, whereas those with fewer elements may have more specialised roles. The co-occurrence of stress-related and development-related cis-elements further implies potential cross-talk between environmental signalling pathways and developmental regulation, particularly in floral organ formation.

Furthermore, the predicted nuclear localisation of MADS-box proteins supports their role as transcription factors involved in regulating downstream target genes. The integration of gene structure, conserved motifs, cis-regulatory elements and expression profiles suggests that sesame MADS-box genes have undergone functional specialisation while retaining core regulatory functions. Overall, although the general characteristics of MADS-box genes in sesame are consistent with those reported in other plant species, notable differences, such as the reduced gene number and the absence of specific subfamilies (Mβ and AGL6), underscore species-specific evolutionary adaptations. These findings provide valuable insights into the regulatory mechanisms underlying floral development and stress responses in sesame and contribute to a broader understanding of MADS-box gene family evolution in plants.

Conclusion

In the present study, a total of 42 MADS-box genes were identified and mapped across 13 chromosomes of the sesame genome, comprising both type I and type II groups. Gene structure and motif analyses revealed that type II MADS-box genes are more abundant, structurally complex and likely functionally conserved compared to type I genes, which exhibit simpler architectures suggestive of evolutionary divergence. Phylogenetic analysis with *Arabidopsis* further supported the classification and evolutionary relationships of these genes. Notably, the absence of specific subfamilies such as Mβ and AGL6 indicates possible lineage-specific gene loss and highlights differences in regulatory mechanisms governing floral development in sesame. Additionally, promoter analysis suggests that these genes are regulated by diverse cis-elements, indicating their involvement in integrating developmental and stress-responsive pathways. Overall, this study provides a comprehensive framework for understanding the organisation, evolution and regulatory potential of MADS-box genes in sesame. Future research focusing on functional validation through gene expression studies, genome editing approaches and protein interaction analyses will be essential to elucidate their precise roles in plant growth, development and stress adaptation.

Acknowledgements

MS¹, EP and GS conceived and wrote the manuscript. TA, MS², GK, MR, NM and MJ contributed to improving the manuscript. All authors have read and approved the final manuscript. [MS¹ stands for M Sandhya and MS² stands for M Sudha].

Compliance with ethical standards

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

Ethical issues: None

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly to improve the language fluency and correct grammatical issues. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- Nayar NM, Mehra KL. Sesame: its uses, botany, cytogenetics and origin. *Econ Bot.* 1970;24:20–31. <https://doi.org/10.1007/BF02860629>
- Wei X, Wang L, Zhang Y, Ding X, Zhang J, Zhang X. Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. *Molecules.* 2014;19(4):5150–62. <https://doi.org/10.3390/molecules19045150>
- Wei W, Zhang Y, Lv H, Li D, Wang L, Zhang X. Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum* L.) germplasm. *J Integr Plant Biol.* 2013;55:745–58. <https://doi.org/10.1111/jipb.12049>
- Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, et al. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 2014;15:R39. <https://doi.org/10.1186/gb-2014-15-2-r39>
- Nam J, Kim J, Lee S, An G, Ma H, Nei M. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc Natl Acad Sci U S A.* 2004;101(7):1910–5. <https://doi.org/10.1073/pnas.0308430100>
- Passmore S, Maine GT, Elble R, Christ C, Tye BK. *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MATα cells. *J Mol Biol.* 1988;204(3):593–606. [https://doi.org/10.1016/0022-2836\(88\)90358-0](https://doi.org/10.1016/0022-2836(88)90358-0)
- Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM. The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature.* 1990;346(6279):35–9. <https://doi.org/10.1038/346035a0>
- Sommer H, Beltran JP, Huijser P, Pape H, Lonnig WE, Saedler H, et al. *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* 1990;9(3):605–13. <https://doi.org/10.1002/j.1460-2075.1990.tb08152.x>
- Norman C, Runswick M, Pollock R, Treisman R. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell.* 1988;55(6):989–1003. [https://doi.org/10.1016/0092-8674\(88\)90244-9](https://doi.org/10.1016/0092-8674(88)90244-9)
- Causier B, Kieffer M, Davies B. MADS-box genes reach maturity. *Science.* 2002;296:275–6. <https://doi.org/10.1126/science.1071401>
- Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature.* 2000;405:200–3. <https://doi.org/10.1038/35012103>
- Ma H, dePamphilis C. The ABCs of floral evolution. *Cell.* 2000;101:5–8. [https://doi.org/10.1016/S0092-8674\(00\)80618-2](https://doi.org/10.1016/S0092-8674(00)80618-2)
- Kaufmann K, Melzer R, Theissen G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene.* 2005;347:183–98. <https://doi.org/10.1016/j.gene.2004.12.014>
- Alvarez-Buylla ER, Pelaz S, Liljgren SJ, Gold SE, Burgeff C, Ditta GS, et al. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci U.S.A.* 2000;97:5328–33. <https://doi.org/10.1073/pnas.97.10.5328>
- Ma H, Yanofsky MF, Meyerowitz EM. AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* 1991;5:484–95. <https://doi.org/10.1101/gad.5.3.484>
- Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, et al. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol.* 2003;20:1963–77. <https://doi.org/10.1093/molbev/msg216>
- Becker A, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol.* 2003;29(3):464–89. [https://doi.org/10.1016/S1055-7903\(03\)00207-0](https://doi.org/10.1016/S1055-7903(03)00207-0)

18. Parenicova L, De Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell*. 2003;15:1538–51. <https://doi.org/10.1105/tpc.011544>
19. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–37. <https://doi.org/10.1093/nar/gkr367>
20. Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res*. 2012;40:D302–5. <https://doi.org/10.1093/nar/gkr931>
21. Chao J, Li Z, Sun Y, Aluko OO, Wu X, Wang Q, et al. MG2C: A user-friendly online tool for drawing genetic maps. *Mol Horticulture*. 2021;1:16. <https://doi.org/10.1186/s43897-021-00020-x>
22. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673–80. <https://doi.org/10.1093/nar/22.22.4673>
23. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 2021;38(7):3022–7. <https://doi.org/10.1093/molbev/msab120>
24. Darzentas N, Circoletto: visualizing sequence similarity with Circos. *Bioinformatics*. 2010;26(20):2620–1. <https://doi.org/10.1093/bioinformatics/btq484>
25. Restrepo-Montoya D, McClean PE, Osorno JM. Orthology and synteny analysis of receptor-like kinases RLK and receptor-like proteins RLP in legumes. *BMC Genomics*. 2021;22:1–17. <https://doi.org/10.1186/s12864-021-07384-w>
26. Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*. 2011;27(12):1653–9. <https://doi.org/10.1093/bioinformatics/btr261>
27. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, 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*. 2002;30(1):325–7. <https://doi.org/10.1093/nar/30.1.325>
28. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13:1194–202. <https://doi.org/10.1016/j.molp.2020.06.009>
29. Leseberg CH, Li A, Kang H, Duvall M, Mao L. Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene*. 2006;378:84–94. <https://doi.org/10.1016/j.gene.2006.05.022>
30. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics*. 2007;8:242. <https://doi.org/10.1186/1471-2164-8-242>
31. Chen Q, Li J, Yang F. Genome-wide analysis of the MADS-box transcription factor family in *Solanum melongena*. *Int J Mol Sci*. 2023;24(1):826. <https://doi.org/10.3390/ijms24010826>
32. Diaz-Riquelme J, Lijavetzky D, Martinez-Zapater JM, Carmona MJ. Genome-wide analysis of MIKCC-type MADS-box genes in grapevine. *Plant Physiol*. 2009;149:354–69. <https://doi.org/10.1104/pp.108.131052>
33. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet*. 2010;42:833–9. <https://doi.org/10.1038/ng.654>
34. Zhao Y, Li X, Chen W, Peng X, Cheng X, Zhu S, et al. Whole-genome survey and characterization of MADS-box gene family in maize and sorghum. *Plant Cell Tiss Org*. 2011;105(2):159–73. <https://doi.org/10.1007/s11240-010-9848-8>
35. Hu L, Liu S. Genome-wide analysis of the MADS-box gene family in cucumber. *Genome*. 2012;55:245–56. <https://doi.org/10.1139/g2012-009>
36. Fan CM, Wang X, Wang YW, Hu RB, Zhang XM, Chen JX, et al. Genome-wide expression analysis of soybean MADS genes showing potential function in the seed development. *PLoS One*. 2013;8:e62288. <https://doi.org/10.1371/journal.pone.0062288>
37. Hileman LC, Sundstrom JF, Litt A, Chen M, Shumba T, Irish VF. Molecular and phylogenetic analyses of the MADS-box gene family in tomato. *Mol Biol Evol*. 2006;23(11):2245–58. <https://doi.org/10.1093/molbev/msl095>
38. Gramzow L, Theißen G. Phylogenomics of MADS-box genes in plants -two opposing life styles in one gene family. *Biology*. 2013;2:1150–64. <https://doi.org/10.3390/biology2031150>
39. Wei B, Zhong RZ, Guo JJ, Liu DM, Li AL, Fan RC, et al. Genome-wide analysis of the MADS-box family in *Brachypodium distachyon*. *PLoS One*. 2014;9(1):e84781. <https://doi.org/10.1371/journal.pone.0084781>
40. Xu Z, Zhang Q, Sun L, Du D, Cheng T, Pan H, et al. Genome-wide identification, characterization and expression analysis of the MADS-box gene family in *Prunus mume*. *Mol Genet Genomics*. 2014;289(5):903–20. <https://doi.org/10.1007/s00438-014-0863-z>
41. Wang P, Wang S, Chen Y, Xu X, Guang X, Zhang Y. Genome-wide analysis of the MADS-box gene family in watermelon. *Comput Biol Chem*. 2019;80:341–50. <https://doi.org/10.1016/j.compbiolchem.2019.04.013>
42. Won SY, Jung JA, Kim JS. Genome-wide analysis of the MADS-box gene family in chrysanthemum. *Comput Biol Chem*. 2021;90:107424. <https://doi.org/10.1016/j.compbiolchem.2020.107424>
43. Chen Y, Yang J. Historic neighborhood design based on facility heatmap and pedestrian simulation: case study in China. *J Urban Plan Dev*. 2020;146(2):04020001. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000554](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000554)
44. Zhang Q, Hou S, Sun Z, Chen J, Meng J, Liang D, et al. Genome-wide identification and analysis of the MADS-box gene family in *Theobroma cacao*. *Genes (Basel)*. 2021;12(11):1799. <https://doi.org/10.3390/genes12111799>
45. Zhao D, Chen Z, Xu L, Zhang L, Zou Q. Genome-wide analysis of the MADS-box gene family in maize: gene structure, evolution and relationships. *Genes (Basel)*. 2021;12(12):1956. <https://doi.org/10.3390/genes12121956>
46. Tian Y, Dong Q, Ji Z, Chi F, Cong P, Zhou Z. Genome-wide identification and analysis of the MADS-box gene family in apple. *Gene*. 2015;555:277–90. <https://doi.org/10.1016/j.gene.2014.11.018>
47. Berner M, Heijmans K, Airoldi C, Davies B, Angenent GC. An atlas of type I MADS-box gene expression during female gametophyte and seed development in *Arabidopsis*. *Plant Physiol*. 2010;154(1):287–300. <https://doi.org/10.1104/pp.110.160770>
48. Deng M, Dong Y, Xu S, Huang S, Zhai X, Fan G. Genome-wide identification and expression of the *Paulownia fortune* MADS-box gene family in response to phytoplasma infection. *Genes*. 2023;14:696. <https://doi.org/10.3390/genes14030696>
49. Hecht V, Foucher F, Ferrandiz C, Macknight R, Navarro C, Morin J, et al. Conservation of *Arabidopsis* flowering genes in model legumes. *Plant Physiol*. 2005;137(4):1420–34. <https://doi.org/10.1104/pp.104.057018>
50. Ruelens P, Geuten K. When paleontology and molecular genetics meet: a genetic context for the evolution of conifer ovuliferous scales. *New Phytol*. 2013;200(1):10–12. <https://doi.org/10.1111/nph.12417>
51. Yoo SK, Wu X, Lee JS, Ahn JH. AGAMOUS-LIKE 6 is a floral promoter that negatively regulates the FLC/MAF clade genes and positively regulates FT in *Arabidopsis*. *Plant J*. 2011;65(1):62–72. <https://doi.org/10.1111/j.1365-3113.2010.04402.xf>

Additional information

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

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

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

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc

See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.