BMC Genomics

RESEARCH

Open Access

Genome-wide identification, characterization, and expression analysis of the MADS-box gene family in grass pea (*Lathyrus sativus*) under salt stress conditions

Mohamed Abdelsattar[1], Ahmed E. Nassar[2], Khaled H. Mousa[2], Ahmed Hussein[1], Manal M. S. El-Baghdady[1], Khaled H. Radwan[1,3], Manar S. Ibrahim[1], Achraf El Allali[4], Aladdin Hamwieh[2], Alsamman M. Alsamman[1,2*] and Zakaria Kehel[5]

Abstract

Background The MADS-box gene family possesses significant potential to improve crop production under harsh conditions by regulating growth, development, and the expression of floral organs. The grass pea (*Lathyrus sativus*), a crop grown predominantly in arid and semi-arid regions, could benefit greatly from the functions of MADS-box genes, which are not yet well characterized in this promising plant.

Results In this study, a comprehensive analysis of all MADS-box genes in grass pea was performed at both the genomic and transcriptomic levels. A total of 46 genes were identified and classified based on their MADS-box domains. A comparative phylogenetic analysis with apple, Arabidopsis, and rice categorized the grass pea genes into 31 type I genes (Mα, Mβ, Mγ) and 15 type II genes (MIKCc, MIKC*). Annotation analysis revealed variations in the intron-exon structures of the genes, with most type I genes being intronless. Ten distinct conserved motifs were identified across the genes. Structural analysis revealed the presence of MEF2-like and SRF-TF domains in the grass pea proteins. Protein-protein interaction analysis revealed extensive interactions among type II MADS-box genes, while enrichment analysis showed their involvement in various aspects of plant life, particularly floral organ development. Examination of the *cis*-elements in the promoter regions of the genes revealed up to 76 potential *cis*-elements, which were categorized into four groups based on their putative role in transcriptional regulation. RNA-seq was used to profile gene expression under different conditions to gain insights into their potential functional significance. Quantitative PCR (qPCR) analysis validated the expression levels of eight selected genes (*LSMADS_D1*, *LSMADS_R5*, *LSMADS_R7*, *LSMADS_R9*, *LSMADS_D11*, *LSMADS_D13*, *LSMADS_R13*, and *LSMADS_D29*) under salt stress conditions and confirmed their involvement in stress responses.

Conclusion This study represents the first genome-wide exploration of the MADS-box gene family in grass pea. Our results provide valuable insights that could improve our understanding of the plant's genomics, contribute to strengthening its resilience to challenging conditions, and help position it as an important crop in arid regions.

Keywords Grass pea, LSMADS, MADS-box, Expression profiles, RNA-seq, Salt stress

*Correspondence:
Alsamman M. Alsamman
smahmoud@ageri.sci.eg
Full list of author information is available at the end of the article

Background

The grass pea (*Lathyrus sativus*) is emerging as a promising crop with the potential to thrive in drought-prone regions such as India, Bangladesh, Pakistan, Nepal, and Ethiopia [1, 2]. It is highly valued by growers for its adaptability to a wide range of soil conditions and its ability to withstand poor soil quality and dense clay, owing to its flexible genome and deep root system [3]. Among the extensive genus *Lathyrus*, the grass pea stands out as the most economically important member [4]. The nutritional profile of grass pea is particularly appealing: it contains approximately 32% protein, 2.7% fat, 362.3 calories per 100 g, 1.1% crude fiber, and 2.2% ash. These properties make it a valuable food and feed resource.

The MADS-box gene family has received considerable attention because of its key functions in plant development, responses to stress, and environmental adaptability. The functional diversity of MADS-box genes appears beneficial for stress tolerance, flowering regulation, and environmental adaptability. Characterizing these genes in the ancient crop grass pea (*Lathyrus sativus*), recognized as a promising crop for drought-prone regions, presents an opportunity to leverage this diversity. Although grass pea is known as a resilient and nutritious crop, limited research has specifically focused on the function of MADS-box genes within it, necessitating further exploration.

Plants often rely on transcriptional regulation for precise biological processes such as flower development, fruit ripening, and responses to abiotic stress. This regulation is mediated by several transcription factor gene families, including MADS-box genes. Their role in stress tolerance regulation is conserved across a wide range of plants. For instance, OsMADS27 in rice improves tolerance to salt stress by regulating the expression of stress-responsive genes [5]. Likewise, SiMADS51 plays an important role in drought tolerance in foxtail millet, suggesting both conservation and functional diversification of MADS-box genes involved in stress tolerance mechanisms among different crops [6]. Furthermore, studies in sheepgrass revealed the specific participation of eleven differentially expressed MADS-box genes in response to multiple abiotic stresses, indicating potential for similar functional roles in grass pea [7]. Thus, further characterization and validation of MADS-box genes in grass pea may provide useful tools for breeding programs aimed at improving the plant's adaptation to environmental stress.

MADS-box genes also have essential functions in flowering regulation, with intricate gene regulatory networks affecting reproductive development. For example, PsEND1, a pea MADS-box gene, regulates pollen development and plays a role in both development and reproduction [8]. Additionally, studies on pineapple flowering networks have identified major MADS-box interactions that mediate environmental inputs and flowering outputs [9]. Elucidating such networks in grass pea could reveal factors affecting flowering time, which would be critical for matching flowering to varying climatic conditions for reproductive success. Consequently, regulating MADS-box gene expression could dramatically improve the environmental adaptability of grass pea. Linkages between developmental regulation and environmental adaptability have been reported in both the model plant Arabidopsis and in wheat, where the genetic diversity of MADS-box genes was directly associated with rapid adaptation to environmental chemistry [10]. This demonstrates the potential for linking developmental processes and stress adaptation, exemplified by how mutations in MADS-box genes regulating flowering time strongly enhanced cold stress tolerance in Arabidopsis [11]. Identifying similar intersections in grass pea could be exploited for enhancing cultivars to thrive in more diverse and challenging environments under climate change.

Experimental validations, such as overexpression and knockout studies, are crucial for confirming the functional role of grass pea MADS-box genes. Overexpression studies of specific MADS-box genes in other species, for example, rice and Arabidopsis, have demonstrated that these genes could be useful for the genetic enhancement of drought tolerance [6]. For instance, in grass pea, the overexpression of a zinc-finger gene (GpZF) enhanced drought tolerance, indicating that analogous strategies could be applied with MADS-box genes [12]. Conversely, gene function becomes clearer in knockout studies, such as those involving the Arabidopsis AGL16 gene, which exhibited altered stress tolerance and perturbed developmental timing [6]. Applying these methods to grass pea can provide tremendous insight into the gene-specific contributions to stress response and developmental regulation. This gene family is divided into two main evolutionary groups: Type I genes, which encode SRF-like (Serum Response Factor-like) domain proteins, and Type II genes, which encode MEF2-like (Myocyte Enhancer Factor 2-like) proteins [13]. Further classification divides the type I genes into M-type and N-type. A phylogenetic analysis based on the MADS-box domains led to the classification of type I genes into Mα, Mβ, Mγ, and Mδ groups [14, 15]. Type II genes contain four characteristic domains: the MADS (M) domain, the Intervening (I) domain, the Keratin-like (K) domain, and the C-terminal (C) domain. Additionally, they can be categorized into the MIKC^c-type (classic) or the MIKC*-type based on structural differences, particularly in the I and K domains [16, 17].

A closer look at type II genes reveals the importance of the M domain, which spans approximately 60 amino

acids and is responsible for sequence-specific DNA binding. The intermediate domain (I), consisting of about 35 amino acids, contributes to the dimerization specificity of the DNA-binding complex. Furthermore, the K domain, a semi-conserved region spanning about 70 amino acids, facilitates protein-protein interactions. Although the C-terminal region (C) is less conserved in its sequence, it frequently contributes to transcriptional activation and the formation of higher-order protein complexes [18, 19]. This comprehensive understanding emphasizes the complex role of the MADS-box gene family in diverse organisms and complicated developmental processes. The MADS-box gene family comprises transcription factors that are critical for regulating flowering time, floral organ and meristem identity, fruit ripening, embryonic development, and root and leaf growth [20–22].

The broader impact of this research is that through detailed characterization and functional analysis of MADS-box genes in grass pea, breeders will gain information useful for designing cultivars with enhanced stress tolerance, appropriate flowering fitness, and better adaptability to changing environments. These advancements are warranted for promoting grass pea as an important crop in a changing agricultural future, particularly as climate-related stresses intensify globally.

Results

It is worth noting that the following gene names follow the format 'LSMADS', where 'LS' refers to *Lathyrus sativus* and 'MADS' to the MADS-box gene family. The suffix '_D' or '_R' indicates the genomic or transcriptomic origin, respectively.

Identification of MADS-box genes in grass pea

To identify LSMADS genes in grass pea, our analysis began with a search for known MADS-box protein sequences in NCBI, followed by alignment against grass pea genomic and transcriptomic data. We used the TBLASTN and HMMER algorithms, leading to the discovery of 53 potential LSMADS sequences belonging to different subtypes (MIKCc, MIKC*, Mα, Mβ, and Mγ) that potentially contain MADS-box domains. MIKC-type genes (including MIKCc and MIKC*) are typically involved in regulating floral organ identity, flowering time, and reproductive development. In contrast, type I genes (Mα, Mβ, and Mγ) are generally associated with roles in embryogenesis, endosperm development, and sometimes responses to environmental stress. Subsequently, these candidates were subjected to gene prediction using the AUGUSTUS tool. Nine sequences were excluded as they were predicted to be non-functional or incomplete (Additional file 1). The predicted encoded amino acid sequences exhibited varying lengths, ranging from 82 (LSMADS_D37) to 344 (LSMADS_D1) amino acids, with an average length of 209.13 amino acids (Additional file 2).

Phylogenetic analysis of the MADS-box gene family

The MADS-box gene family has been identified and characterized in the genomes of numerous plant species [23]. To investigate the evolutionary relationships of LSMADS genes, we constructed a phylogenetic tree using the maximum likelihood method (Fig. 1). This analysis included a total of 338 MADS-box proteins: 141 from apple (*Malus domestica*), 100 from Arabidopsis (*Arabidopsis thaliana*), 51 from rice (*Oryza sativa*), and the 46 identified from grass pea. This comparative approach provides valuable insights into evolutionary history and functional conservation by classifying genes based on established relationships [24]. The 46 LSMADS genes identified in this study were categorized into two main groups: Type I (31 genes) and Type II (15 genes) (Additional file 5). Further subclassification of the type I LSMADS genes resulted in three subgroups: Mα (21 genes), Mγ (7 genes), and Mβ (3 genes). The Mα subgroup was the most abundant, followed by Mγ and then Mβ. Type II LSMADS genes were categorized into two subgroups: MIKCc (14 genes) and MIKC* (1 gene). The MIKCc subgroup represents the most extensive evolutionary branch in this analysis, containing 184 genes from Arabidopsis, apple, and rice, in addition to the 14 LSMADS genes (Fig. 1). In contrast, the MIKC* subgroup was the smallest, consisting of 12 Arabidopsis and apple genes and only one LSMADS gene. This comprehensive classification provides a basis for understanding the diversity and evolutionary history of LSMADS genes.

Gene structure and conserved motif identification

Diversity in exon-intron structure and distribution within gene families provides crucial clues for understanding the evolutionary relationships and origins of the different gene families [25]. To explore the gene structure of the MADS-box genes in grass pea, we compared the predicted cDNA sequences with the corresponding genomic DNA sequences for the 46 LSMADS genes. Within the type II LSMADS genes, six genes (*LSMADS_D23, LSMADS_D30, LSMADS_R2, LSMADS_R3, LSMADS_R8, LSMADS_R9*) contain a single intron, while *LSMADS_D32* has the highest number with three introns (Fig. 2). All type II LSMADS genes possess exons; specifically, *LSMADS_D32*, *LSMADS_D37*, and *LSMADS_R5* contain 4, 3, and 2 exons, respectively, while the rest have a varying number of exons, including intronless genes (Fig. 2, Additional files 6 and 7). In contrast, most type I LSMADS genes (26 of 31) were predicted to be intronless. Exceptions include *LSMADS_D12*, *LSMADS_D15*, *LSMADS_D16*,

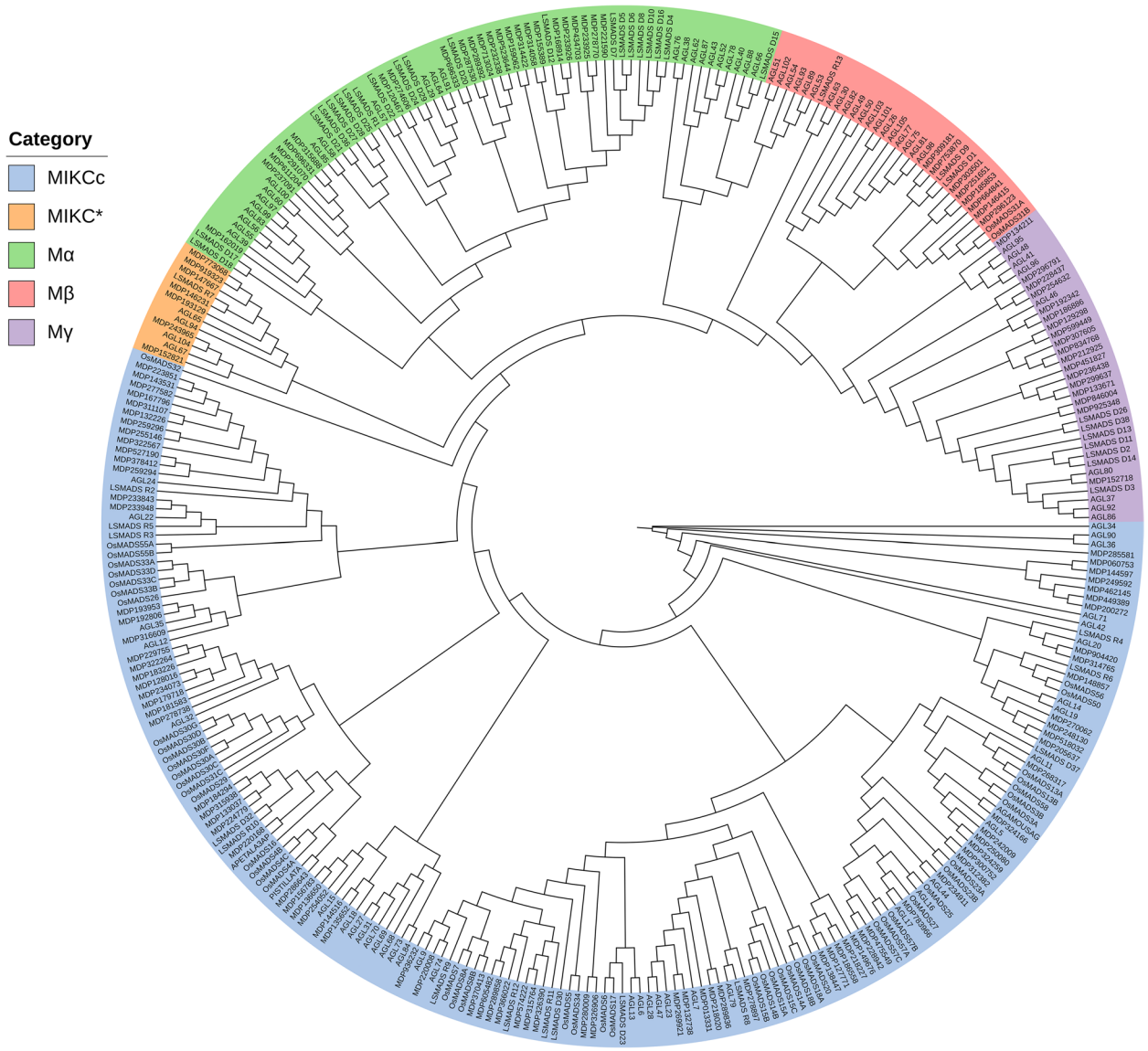

Fig. 1 Maximum likelihood phylogenetic tree of MADS-box genes constructed using 338 MADS-box proteins sourced from apple (141), Arabidopsis (100), rice (51), and grass pea (46). The genes are categorized into Type I (Mα, Mβ, Mγ) and Type II (MIKCc, MIKC*) subfamilies. Each subgroup is color-coded according to the lighter Tableau 10 scheme used in the figure: MIKCc (Light Blue), MIKC* (Light Orange), Mα (Light Green), Mβ (Light Red/Pink), and Mγ (Light Purple/Lavender). This color scheme highlights the evolutionary relationships and structural classification of the LSMADS genes

LSMADS_D20, and *LSMADS_D38*, each possessing one intron. Furthermore, all type I genes contain exons, with most having a single exon, except for *LSMADS_D15* and *LSMADS_D38*, which have two exons each (Additional files 6 and 7). To investigate conserved regions within the predicted proteins, we used the online MEME suite and NCBI-CDD. This analysis revealed ten conserved motifs, designated motifs 1–10, ranging in length from 24 to 51 amino acids (Fig. 2). Motif 1, corresponding to the MADS-box domain, stands out as fundamental, being found in 37 LSMADS proteins. Motifs 2 and 3 appear to be specific to certain Mα proteins. This comprehensive study of gene structure and motifs enhances our understanding of the structural diversity and potential functional elements of LSMADS genes in grass pea.

PPI network construction and gene ontology enrichment analysis

Protein-protein interactions (PPIs) involving LSMADS proteins were investigated to uncover potential

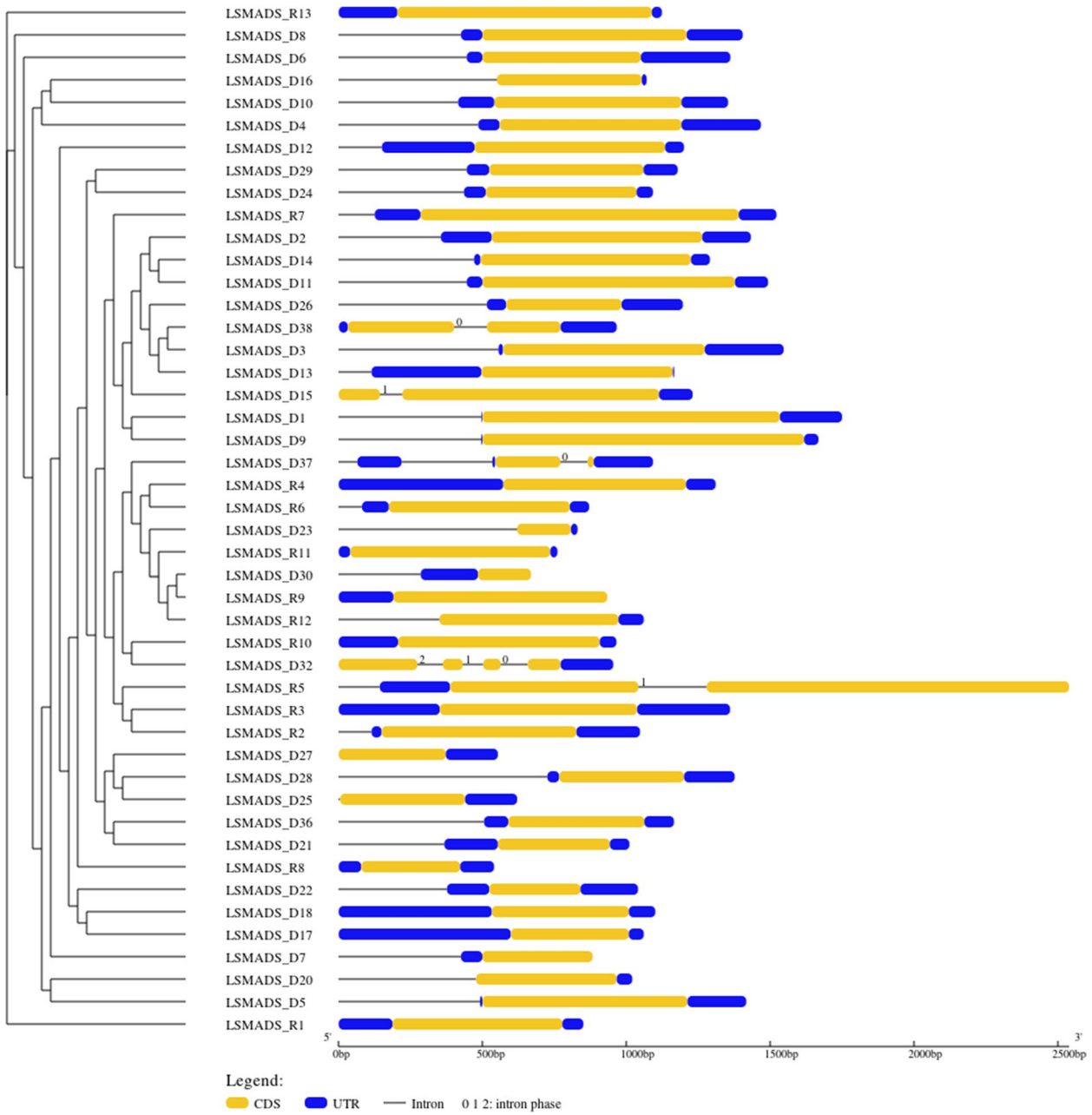

Fig. 2 Gene structure and conserved motif analysis of LSMADS genes. (Top) Phylogenetic tree based on the LSMADS protein sequences. (Middle) Conserved motifs identified by MEME analysis. Different colored boxes represent different motifs (legend below). (Bottom) Exon-intron structure. Green boxes represent exons, black lines represent introns, and grey boxes represent UTRs (if predicted, but usually not present in this representation). The scale bar indicates protein length or gene length as appropriate contextually. Intron phases (0, 1, 2) indicating the position within the codons are marked

functions, signaling pathways, and metabolic processes. The STRING database was used to predict interactions between the 46 LSMADS proteins, using ortholog information primarily from *Arabidopsis thaliana* (Fig. 3). The analysis revealed interactions for a subset of genes, with limited predicted interactions for others, likely due to the nascent stage of grass pea functional genomics and reliance on cross-species predictions. The majority of the interacting proteins belonged to the type II MADS-box gene family (MIKCc). At the center of the interaction network appeared *LSMADS_R12*, predicted to interact with six other MIKCc proteins (*LSMADS_R10*,

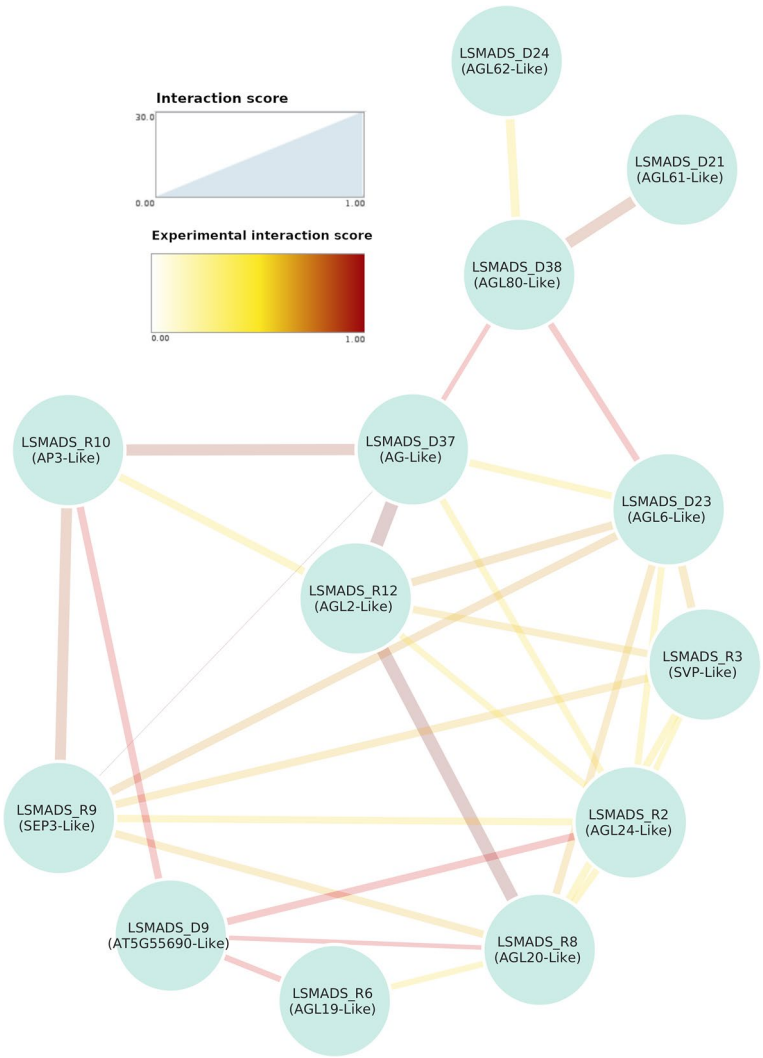

Fig. 3 Predicted Protein-Protein Interaction (PPI) network of LSMADS proteins. Nodes represent LSMADS proteins, and edges depict predicted interactions among them. Gene names in parentheses may indicate the primary ortholog (*A. thaliana*) used for prediction if directly imported from STRING defaults. Interaction edge colors and widths often correspond to the confidence scores of the interactions (e.g., experimental evidence, co-expression, orthology) obtained from the STRING database. Lines connecting nodes represent predicted interactions

LSMADS_D37, *LSMADS_D23*, *LSMADS_R3*, *LSMADS_R2*, and *LSMADS_R8*). Another highly connected protein was *LSMADS_R9*, predicted to interact with the same set (*LSMADS_R10*, *LSMADS_D37*, *LSMADS_D23*, *LSMADS_R3*, *LSMADS_R2*, and *LSMADS_R8*). Only three type I proteins (*LSMADS_D24*, *LSMADS_D21*, and *LSMADS_D38*) were involved in predicted interactions within this network. *LSMADS_D24* and *LSMADS_D21* did not show predicted interactions with type II proteins but were linked to *LSMADS_D38*. *LSMADS_D38* itself was linked to the type II proteins *LSMADS_D23* and *LSMADS_D37*. Gene Ontology (GO) enrichment analysis (Additional file 8) indicated that these genes are involved in various biological processes, molecular functions, and cellular components related to plant development and response. Enriched GO terms included DNA binding, transcription regulation, cell differentiation, ovule development, reproductive structure development, stamen development, meristem maintenance, and positive regulation of flower development, as well as associations with intracellular membrane-bound organelles (Fig. 4).

Analysis of putative promoter regions in LSMADS genes

Cis-regulatory elements in promoter regions act as binding sites for transcription factors, playing a key role in controlling the timing, location, and level of gene expression [6]. In the current study, analysis of the 1.5 kb

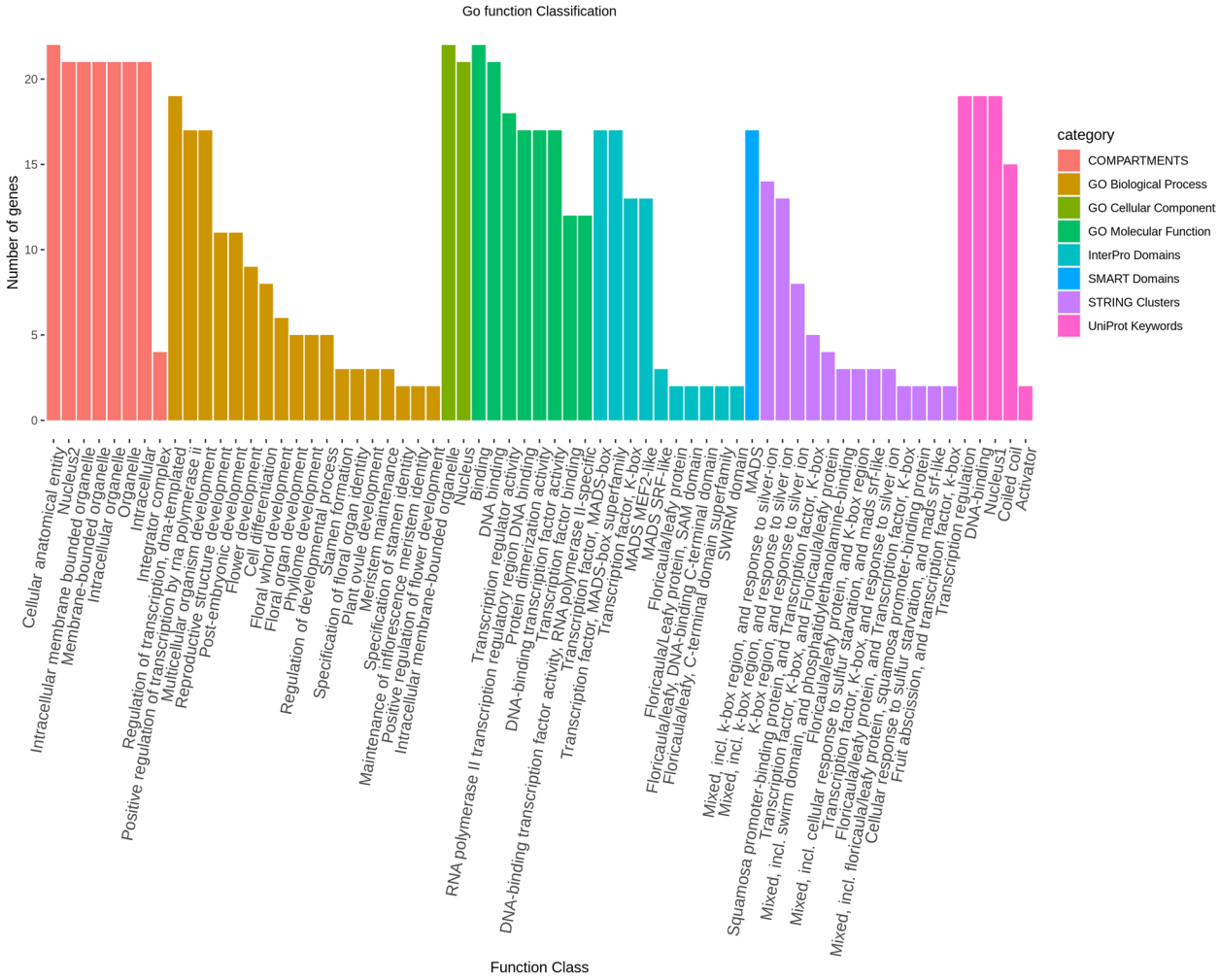

Fig. 4 Gene Ontology (GO) Enrichment Analysis of the 46 LSMADS genes. The plot shows enriched GO terms categorized into Biological Process, Cellular Component, and Molecular Function. Other categories like InterPro Domains, SMART Domains, STRING Clusters, and UniProt Keywords may also be included depending on the specific analysis tool used (original text mentioned 8 categories, this caption focuses on GO). Bar length or point size typically represents the significance (e.g., *p*-value) or the number of genes associated with the term. Color might represent the category or significance level

upstream sequences of LSMADS genes identified numerous potential *cis*-elements (up to 76 unique types across all genes), with individual genes like *LSMADS_D11* having a high number of diverse elements (Additional file 9). These predicted *cis*-elements are associated with various functions in plants and can be broadly categorized based on their known responsiveness (Fig. 5). One category includes elements related to hormone responses (e.g., abscisic acid - ABRE; gibberellin - P-box, TATC-box; salicylic acid - TCA element; methyl jasmonate - TGACG motif, CGTCA motif) and light responses (e.g., G-box, Box 4, AE-box, GT1-motif), as well as elements involved in circadian control and anaerobic induction (ARE). A second category includes elements associated with tissue-specific expression, such as endosperm expression (O2 site, GCN4 motif) and meristem expression (CAT-box), along with elements potentially involved in seed-specific regulation (RY-element) and auxin response (TGA-element). A third category comprises elements linked to stress responses, including drought (MBS - MYB binding site), low temperature (LTR), and defense/general stress responses (TC-rich repeats, W-box). A fourth category encompasses basic promoter elements like the CAAT-box and the TATA-box, which are involved in transcription initiation.

Salicylic acid is known to promote systemic acquired resistance (SAR) to pathogens and to regulate plant growth, development, and signaling pathways [26, 27].

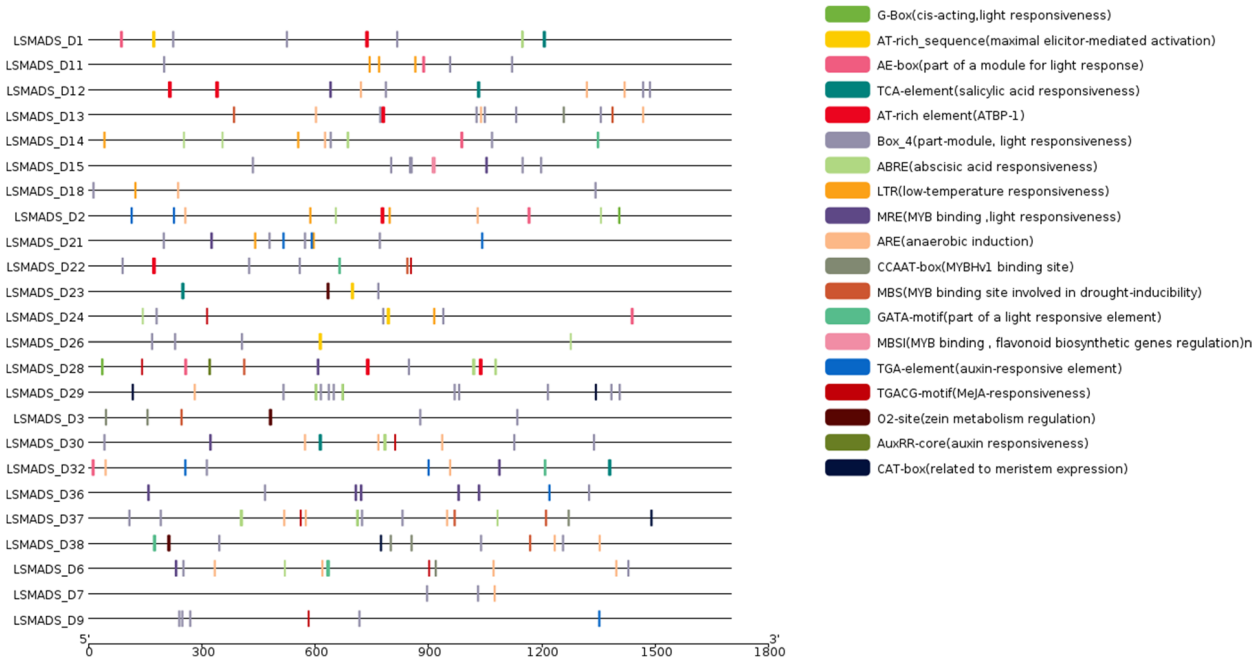

Fig. 5 *Cis*-acting regulatory elements predicted within the 1.5 kb upstream promoter regions of LSMADS genes. Gene names are listed on the left. The colored boxes represent different types of predicted *cis*-elements, potentially grouped by function (e.g., light response, hormone response, stress response). A legend (often provided separately or implicitly through color consistency) indicates the specific element type represented by each color. The scale bar below indicates the length of the promoter sequence analyzed (1500 bp)

Jasmonates are hormones involved in defense against biotic and abiotic stressors and the production of secondary metabolites [28]. The W-box plays a role in stress responses and senescence [29]. The G-box is involved in light-responsive gene expression and can interact with other elements like the H-box (involved in phenylpropanoid biosynthesis) or CACCTG motifs (involved in seed storage protein gene regulation) to modulate expression [30–32]. This analysis highlights the potential for complex regulation of LSMADS genes in response to developmental cues and environmental stimuli.

Expression pattern analyses of LSMADS genes using RNA-seq

We investigated the expression patterns of LSMADS genes under different conditions using publicly available RNA-seq data from different tissues (shoots, seeds, leaves) and germplasm accessions. RNA-seq is a powerful technique for assessing gene activity and discovering new novel transcripts [24]. A key finding relates to the response of LSMADS genes to salt stress (data derived from processed results, potentially from study SRP327502 or similar, although not explicitly stated as salt stress in the method description of SRA datasets - assuming this connection based on subsequent qPCR). Differentially expressed genes were observed in response to different NaCl concentrations (inferred context). Most tested genes showed upregulation at 50 mM NaCl. For instance, *LSMADS_D11* showed approximately a 4-fold increase upon initial salt stress. More responsive genes, like *LSMADS_D5* and *LSMADS_R7* (note: R7 is MIKC*, D5 is MIKCc), exhibited 32- and 38-fold increases, respectively, suggesting potential roles in early stress signaling. Notably, *LSMADS_R13* (Type I Mγ) showed a dramatic increase (up to 605-fold), suggesting a significant regulatory role during salt stress adaptation. However, it is possible that certain spatial and temporal expression patterns, particularly for type II genes involved in development, are not be fully captured in the available RNA-seq datasets [33].

In experiment SRP045652 (samples treated with fungal spores), most LSMADS genes showed low or no expression, with the exceptions of MIKCc genes *LSMADS_R2*, *R3*, *R4*, *R5*, *R6*, *R7*, *R8*, *R11*, and *R12* (Fig. 6). This suggests specific roles for these genes under biotic stress or that the conditions did not significantly induce most LSMADS genes. In contrast, experiment SRP092875 (germplasm tissue) showed significant transcript levels for many genes. Here, several type II MIKCc genes exhibited higher expression levels compared to many type I genes (Fig. 6). The genes *LSMADS_D5*, *D16*, *D25*, *D26*, *D27*, and *D36* showed little to no expression, while *LSMADS_D9* (MIKCc) and *LSMADS_R3* (MIKCc) showed exceptionally high expression levels.

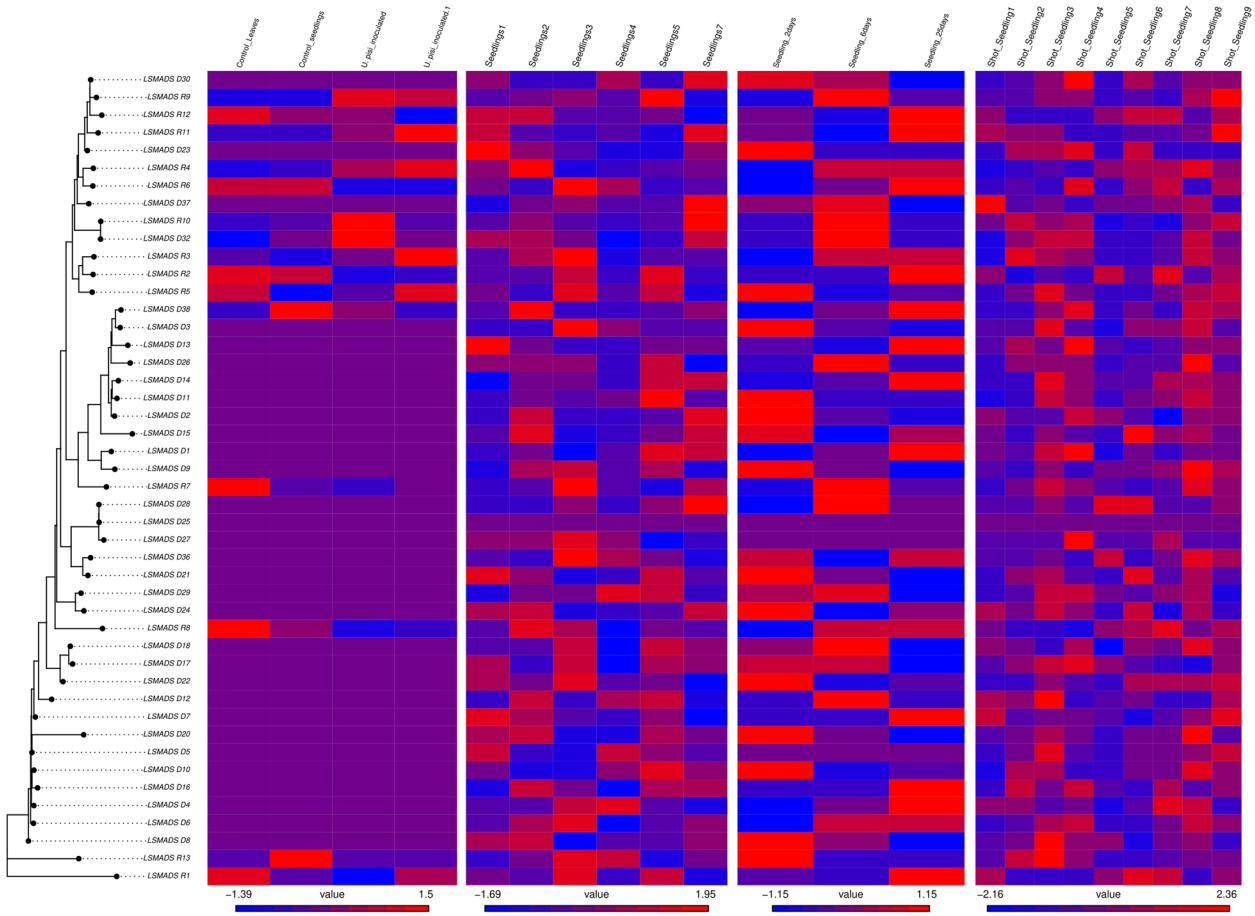

Fig. 6 Expression levels of LSMADS genes across different tissues and experiments based on RNA-seq data. The heatmap displays transcript abundance (e.g., log2(TPM+1) or similar normalized values) for each gene (rows) across different samples/conditions (columns) from four SRA experiments (SRP045652, SRP092875, SRP145030, SRP327502). The color scale indicates relative expression levels, typically with red indicating higher expression and blue/purple indicating lower expression, compared to the mean or a reference sample

In experiment SRP145030 (different shoot tissues over time), gene expression varied depending on the developmental stage. Tissue isolated from young shoots 25 days after sowing generally showed higher expression for several LSMADS genes compared to tissue isolated after 2 and 6 days. In experiment SRP327502 (seedling shoots), *LSMADS_R3*, *R4*, and *R5* showed high expression levels, while *LSMADS_D16*, *D17*, *D21*, and *R6* showed very low or no expression (Fig. 6). In summary, based on these datasets, several type II MIKCc genes (*LSMADS_R2*, *R3*, *R4*, *R5*, *R7*, *R8*, *R9*, *R12*) often showed detectable to high expression levels across different experiments. Conversely, some type I genes like *LSMADS_D16*, *D25*, and *D30* consistently showed low or no expression in the sampled tissues and conditions. These results highlight the diverse expression patterns and potential functional roles of LSMADS genes.

While the RNA-seq analysis revealed valuable insights into tissue-specific and potentially stress-induced expression patterns of LSMADS genes, comprehensive RNA-seq analyses, including detailed time-course experiments under various abiotic stresses like drought or heat, are beyond the scope of this study but are warranted in future investigations.

qRT-PCR validation of LSMADS gene expression under salt stress

Salt stress poses a significant threat to crop productivity, especially in the arid and semi-arid regions where grass pea is commonly cultivated. The detrimental effects of salinity on plant growth and yield underscore the urgency of developing salt-tolerant varieties. To investigate the molecular response of grass pea to salinity, we exposed plants to different concentrations of NaCl and measured the expression of selected LSMADS genes using quantitative real-time PCR (qRT-PCR). Our aim was to identify key genes involved in the salt stress response and gain

insights into the molecular mechanisms potentially contributing to salt tolerance in grass pea.

We selected eight LSMADS genes representing different subgroups (Type I: *D1*-Mα, *D11*-Mα, *D13*-Mγ, *R13*-Mγ, *D29*-Mα; Type II: *R5*-MIKCc, *R7*-MIKC*, *R9*-MIKCc) for qRT-PCR analysis under increasing salt concentrations (0, 50, 100, and 200 mM NaCl) (Fig. 7). At 50 mM NaCl, most selected genes showed an upregulation compared to the control (0 mM), notably *LSMADS_R5*, *LSMADS_R7*, *LSMADS_R9*, *LSMADS_D13*, *LSMADS_R13*, and *LSMADS_D29*. The expression of several genes (*LSMADS_R5*, *LSMADS_R9*, *LSMADS_D11*) remained relatively stable or slightly decreased between 50 mM and 100 mM NaCl. However, *LSMADS_R7* expression peaked at 100 mM NaCl before declining significantly at 200 mM, suggesting an optimal response range or potential negative feedback at higher stress levels. Interestingly, the expression of *LSMADS_D1*, *LSMADS_D13*, *LSMADS_R13*, and *LSMADS_D29* continued to increase or remained highly elevated at 100 mM and 200 mM NaCl. Specifically, *LSMADS_D1*, *LSMADS_D13*, and *LSMADS_D29* showed substantial upregulation (approximately 10-, 57-, and 27-fold, respectively) at 200 mM NaCl compared to the control, suggesting roles in sustained or late responses to severe salt stress. *LSMADS_R13* also maintained very high expression levels at 200 mM NaCl. These results confirm the responsiveness of these specific LSMADS genes to salt stress and suggest their involvement in grass pea's adaptation mechanisms to salinity.

Discussion

Comparative analysis of LSMADS gene characteristics in grass pea (*Lathyrus sativus*) with related legumes highlights their potential evolutionary significance and functional conservation. The MADS-box gene family has been extensively studied in legumes such as pea (*Pisum sativum*), *Medicago truncatula*, and soybean (*Glycine max*). Research often reveals strong purifying selection acting on key developmental genes across legume species, suggesting functional conservation under evolutionary pressure [34]. Given that essential functions of

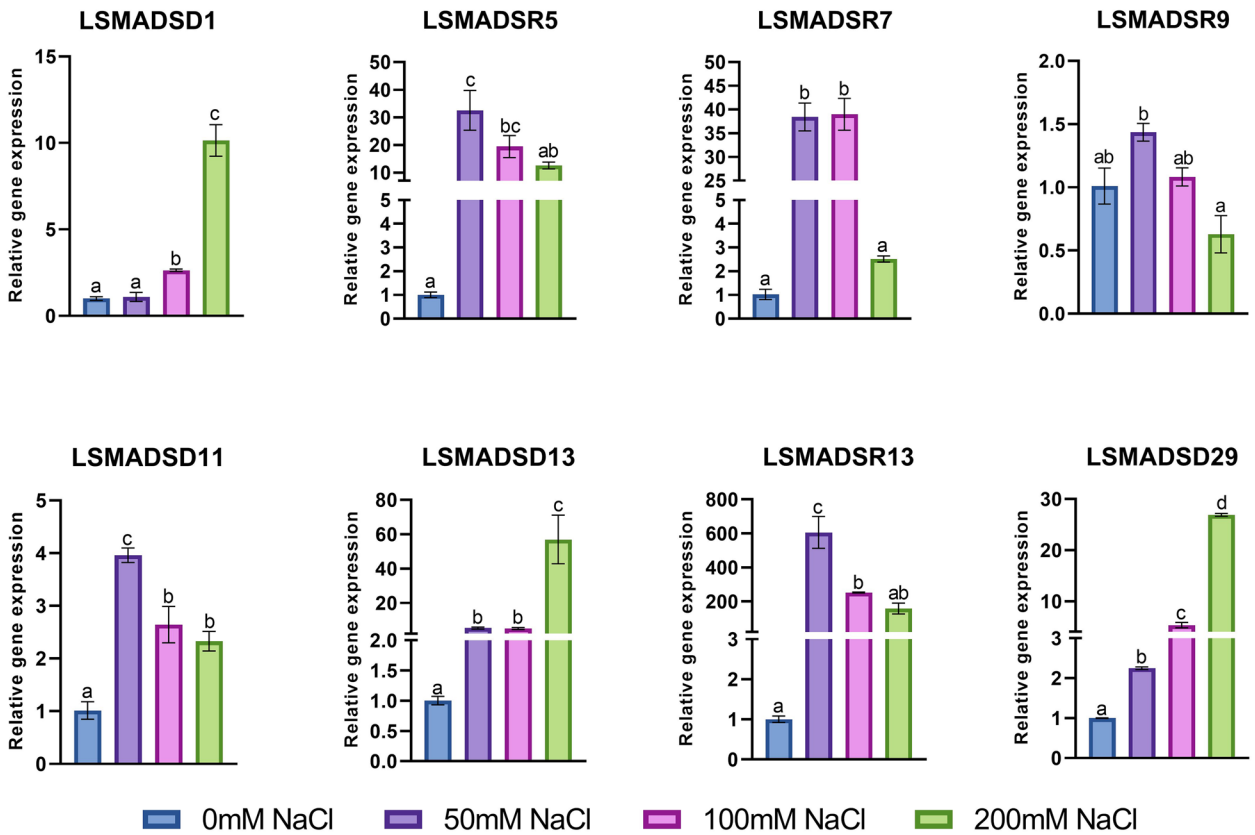

Fig. 7 Relative expression patterns of eight selected LSMADS genes (*LSMADS_D1*, *LSMADS_R5*, *LSMADS_R7*, *LSMADS_R9*, *LSMADS_D11*, *LSMADS_D13*, *LSMADS_R13*, and *LSMADS_D29*) under varying concentrations of salt stress (0, 50, 100, and 200 mM NaCl) determined by qRT-PCR. Expression levels are shown relative to the control (0 mM NaCl) after normalization using reference genes (ABCT and Elf1b). Different letters above the bars indicate statistically significant differences ($P < 0.05$) between treatments for each gene, as determined by Duncan's multiple range test. Error bars represent the standard deviation of three biological replicates

MADS-box genes in flowering regulation and environmental adaptation are likely maintained by such pressures, the grass pea MADS-box genes examined here may be under similar evolutionary constraints. Supporting this idea, phylogenetic relationships and conserved intron-exon structures or motif arrangements in homologous MADS-box proteins across legumes often indicate evolutionary conservation of function [35]. However, variations in gene number and intron patterns, as observed between grass pea and other species, also point towards lineage-specific evolution and potential functional diversification [36]. Studying these structural variations may provide insights into the evolutionary mechanisms shaping the MADS-box gene family in grass pea.

The study of synteny provides further evidence of evolutionary relationships and gene conservation among legumes. For instance, high syntenic conservation of orthologous genes in legume genomes underscores the utility of comparative frameworks for reconstructing evolutionary history and inferring functional similarity [35]. Comparative phylogenetic analysis can also unveil conserved regulatory functions, particularly in processes such as flowering time control. Key flowering time genes, such as the FTa1 orthologs in *Medicago truncatula* and pea (*Pisum sativum*), are thought to play conserved roles, suggesting that similar ancestral regulatory pathways likely exist in grass pea [37, 38]. These evolutionary perspectives underscore the value of placing LSMADS gene data in the broader legume evolutionary context to identify potentially conserved adaptive traits (e.g., stress tolerance, flowering regulation). While comparative genomics provides valuable hypotheses, functional validation through mutation or overexpression studies is necessary to confirm the specific roles of LSMADS genes. Combining evolutionary, comparative genomic, and functional approaches will strengthen grass pea breeding programs aiming to enhance climate resilience.

The MADS-box gene family plays a central role in regulating flower development and fruit ripening in various plant species, although some members have evolved novel functions over time [39]. These genes encode important transcription factors involved in diverse biological processes [6]. Despite their characterization in species such as *A. thaliana*, poplar (*Populus trichocarpa*), rice, grape, apple, maize, cucumber, soybean, tomato, *Brachypodium distachyon*, and *Prunus mume*, MADS-box genes in grass pea (*L. sativus*) remained largely uncharacterized [40]. Our study addressed this gap by conducting a comprehensive analysis of grass pea genome and transcriptome data to identify, characterize, annotate, and validate MADS-box genes in this species.

We identified and analyzed 46 putative functional MADS-box genes in grass pea, designated LSMADS_D (genomic origin) or LSMADS_R (transcriptomic origin). The number of LSMADS genes (46) is comparable to that in pineapple (48), *Erigeron breviscapus* (44), and cucumber (43). However, it is significantly lower than in Arabidopsis (107), *Populus trichocarpa* (105), *Saccharum spontaneum* (182), *Salix suchowensis* (64), or coconut palm (69) [6, 40]. These differences might reflect lineage-specific gene loss or duplication events during evolution [41]. The number of MADS-box genes varies greatly between plant species, influenced by factors such as whole-genome duplications and subsequent gene retention or loss. Our study provides the first systematic inventory and characterization of MADS-box genes in grass pea.

Understanding the phylogenetic relationships of LSMADS genes is crucial for inferring their potential functions and evolutionary history. Our phylogenetic analysis, including MADS-box proteins from apple, Arabidopsis, rice, and grass pea, classified the 46 LSMADS genes into 31 type I and 15 type II genes (Fig. 1). The predominance of type I MADS-box genes in grass pea is similar to the pattern observed in *A. thaliana*, rice, cucumber, and *Populus trichocarpa* [14, 17, 25, 42]. All identified type II genes in grass pea belong to the MIKC type, with 14 classified as MIKCc and one (*LSMADS_R7*) as MIKC*. Grass pea possesses fewer type II genes compared to maize (43), Arabidopsis (55), poplar (64), soybean (88), *Brassica oleracea* (42), or apple (91). Due to the limited number of MIKCc genes identified and the lack of extensive functional data in grass pea, further subdivision of the MIKCc group into established clades (e.g., AP1/FUL, SEP, AGL6, etc.) requires more detailed comparative analysis, although the phylogenetic tree suggests distinct branches (Fig. 1). The number of MIKCc genes varies considerably among species [43–46]. The 31 type I genes were subdivided into Mα (21), Mβ (3), and Mγ (7) subgroups. The number of type I genes in grass pea differs from poplar (41), *P. mume* (48), apple (56), Arabidopsis (61), and soybean (75), but is comparable to rice and maize (about 32) [40, 47]. This detailed phylogenetic analysis lays the groundwork for future studies on the functional diversity and evolutionary significance of the MADS-box gene family in this important legume.

Gene structure analysis revealed that many LSMADS genes, particularly type I genes (26 of 31), are intronless. Type II genes exhibit more variability: seven are intronless, six have a single intron, and *LSMADS_D32* contains three introns (Fig. 2, Additional file 7). In general, MIKC-type genes tend to have multiple introns, whereas type I (Mα, Mβ, Mγ) genes often have few or no introns [48]. Our results are generally consistent with this pattern, although the number of introns can vary significantly

even within subfamilies in different species [40, 47]. Studies suggest that intron number can sometimes correlate with expression levels or breadth, although this is not a universal rule [49]. Most LSMADS genes consist of a single exon, with the exceptions noted in the Results section (*LSMADS_D15*, *LSMADS_D38*, *LSMADS_D32*, *LSMADS_D37*, *LSMADS_R5*). MIKC-type genes typically possess more exons than type I genes. For example, MIKC genes in Arabidopsis usually contain five to eight exons [40, 48, 50]. Using MEME, we have identified ten conserved motifs within the LSMADS protein sequences.

Members belonging to the same phylogenetic subfamily often have similar motif compositions (Fig. 2). The MIKCc subgroup proteins typically function as transcription factors containing conserved MADS and K-box domains, crucial for functions such as floral organ identity and development [51]. The K domain (involved in protein dimerization) has been identified (e.g., via Motif 6) in most MIKCc proteins, though absent or truncated in a few (*LSMADS_D30*, *LSMADS_D23*, *LSMADS_D37*). The MADS domain (Motif 1, responsible for DNA binding) was identified in nearly all LSMADS proteins (except *LSMADS_R13*, which might be truncated or represent a different class). The conserved MADS domain in LSMADS consists of approximately 51 amino acids. We have also confirmed the presence of the SRF-TF (Type I) and MEF2-like (Type II) signatures within the MADS domains [52]. This detailed annotation provides valuable insight into the conserved structural features and potential functional domains within the grass pea MADS-box gene family.

The Study of protein interaction networks can provide insights into functional relationships and cellular pathways [53]. A predicted PPI network for LSMADS proteins was constructed using STRING (Fig. 3). The analysis suggested significant interconnections between a subset of LSMADS proteins, involving mainly type II MIKCc members. The limited number of predicted interactions for many genes likely reflects the reliance on orthology-based predictions and the current lack of experimental PPI data for grass pea. MIKCc proteins showed strong predicted associations with each other among themselves, largely separate from the few interactions involving type I proteins (*LSMADS_D24*, *LSMADS_D21*, *LSMADS_D38*). This pattern of intra-type II interactions is consistent with the known formation of multimeric complexes by MIKCc proteins during developmental processes, similar to networks observed in other species such as *Punica granatum* [50]. Furthermore, GO enrichment analysis highlighted the potential involvement of LSMADS genes in critical stages of the grass pea life cycle (Fig. 4).

Enriched GO terms indicated a role in transcriptional regulation, DNA binding, and various developmental processes, including floral organ development, ovule development, stamen development, and meristem maintenance. Several specific LSMADS genes have been associated with several key functions. For example, *LSMADS_R4* (MIKCc) has been linked to transcriptional control, DNA binding, and floral organ development. *LSMADS_R8* (MIKCc) has been associated with flower development and meristem identity maintenance. *LSMADS_R2* (MIKCc) has been associated with transcription factor activity and postembryonic development. *LSMADS_R13* (Mγ) has been associated with transcriptional control, DNA binding, and possibly protein dimerization. *LSMADS_D9* (MIKCc) was associated with transcriptional control and floral growth regulation. *LSMADS_R7* (MIKC*) was associated with stamen, floral organ, and phyllome development. Taken together, the PPI and GO analyses suggest significant roles for LSMADS genes, particularly MIKCc members, in regulating plant development, especially floral organs.

Cis-regulatory elements within promoters are crucial for controlling gene expression patterns [54]. Our analysis identified 76 distinct types of potential *cis*-elements in the promoter regions of the 46 LSMADS genes. Notably, CAAT-box and TATA-box elements, common in eukaryotic promoters, were prevalent, similar to findings in other plants like beautyberry [55]. Among the functionally annotated elements, the ABRE (involved in ABA response) was found in numerous LSMADS promoters, suggesting potential roles in abiotic stress responses like drought and high salinity [12]. Other key elements included the G-box (light and hormone response) [32, 56], W-box (pathogen/stress response) [57], TCA element (salicylic acid response) [27, 58], ARE (anaerobic response) [59], and GATA motif (light/nitrate response) (Fig. 5). Many of the *cis*-elements identified in grass pea MADS-box gene promoters are similar to those found regulating homologous genes in maize, lettuce, and beautyberry [55, 60, 61]. The diverse array of predicted *cis*-elements underscores the potential for complex regulation of LSMADS genes by various developmental and environmental signals throughout the plant life cycle.

Given the central role of the MADS-box gene family in flower development, examining their expression under stress conditions is important [62]. We selected eight LSMADS genes from different subfamilies and assessed their response to salt stress using qRT-PCR in leaf tissue (Fig. 7). Several type I genes showed significant upregulation. Specifically, *LSMADS_D1* (Mα), *LSMADS_D13* (Mγ), *LSMADS_R13* (Mγ), and *LSMADS_D29* (Mα) exhibited increased expression in response to salt stress, particularly at higher concentrations (100–200 mM

NaCl). These findings align with studies on type I MADS-box genes in other species. For example, rice type I genes *OsMADS64* and *OsMADS77* showed significant transcript accumulation under salt stress [42]. In contrast, the type II genes *LSMADS_R5* (MIKCc) and *LSMADS_R7* (MIKC*) showed significant upregulation at lower salt concentrations (50–100 mM NaCl) but reduced or stable expression at 200 mM NaCl (Fig. 7).

These varied responses are consistent with findings in *Brassica rapa* and sheepgrass, where MIKC genes displayed diverse expression patterns under salt stress [7, 63]. The qRT-PCR validation confirms the salt-responsiveness of these selected LSMADS genes and provides insights into their potential involvement in grass pea's mechanisms for coping with salinity, highlighting distinct response dynamics between different gene types and individual genes.

It is important to note that while we have computationally identified potential *cis*-regulatory elements and provided indirect experimental support through expression analysis (RNA-seq and qRT-PCR under salt stress), direct experimental validation, such as promoter-reporter assays or chromatin immunoprecipitation (ChIP) experiments, would be required to confirm the functional roles of these promoter elements in stress-responsive gene regulation. Future validation experiments would significantly strengthen our current predictions.

Conclusion

The MADS-box gene family plays a pivotal role in plant growth and development, particularly in specifying floral organ identity. In this study, we conducted the first comprehensive genome-wide survey of MADS-box genes in grass pea using genomic and transcriptomic data, leading to the identification of 46 LSMADS genes. These genes were divided into 31 type I (Mα, Mβ, Mγ) and 15 type II (MIKCc, MIKC*) members based on phylogenetic analysis with related species. We characterized their gene structures, conserved motifs, predicted protein interactions, and putative promoter elements. Expression analysis using RNA-seq and qRT-PCR revealed diverse expression patterns across tissues and identified several genes responsive to salt stress. Based on their phylogenetic position, predicted interactions, and strong stress-induced expression, genes such as *LSMADS_R13*, *LSMADS_D13*, *LSMADS_D1*, *LSMADS_D29*, and potentially key interacting MIKCc members like *LSMADS_R9* and *LSMADS_R12* emerge as promising candidates for further functional characterization and potential targets for breeding programs aimed at enhancing stress tolerance and optimizing development in grass pea.

Although the differential expression of MADS-box genes between salt-treated and control samples provides valuable preliminary guidance, this study has limitations. A lack of comprehensive RNA-seq time-course profiling under various abiotic stresses (e.g., drought, heat, cold) restricts a full understanding of their dynamic regulation. Furthermore, detailed RNA-seq experiments covering multiple time points and stress treatments are recommended to gain deeper mechanistic insights into the regulatory networks involving these genes in grass pea under stress. Moreover, while our study provides initial insights, functional validation through techniques such as gene editing (e.g., CRISPR/Cas9) or overexpression/knockdown experiments is necessary to confirm the precise role of these LSMADS genes in development and stress response.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11661-3.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.

Acknowledgements

The authors acknowledge the African Supercomputing Center at Mohamed VI Polytechnic University for the supercomputing resources (https://ascc.um6p.ma/) made available for conducting the research reported in this paper. The authors would like to express their gratitude to all donors and organizations who support the CGIAR system and helped fund the CGIAR Research Program from which this work benefited.

Authors' contributions

Authors' contributions Mabd, AHU, AMA, AEN, KHM, and MSB collected samples and performed greenhouse experiments. KHR, MSI, AEA, ALH, and ZKE contributed to conception and design of the study. MAB and AHU performed lab-validation. AMA, AEN, KHM, Mabd, AHU, KHR, MSI, AEA, and MSB performed the analysis of genomic and transcriptomic data. AMA, AEN, KHM, Mabd, AHU, KHR, MSI, AEA, and MSB organized the data and performed the statistical analysis. AMA, AEN, KHM, Mabd, AHU, KHR, MSI, and ALH drafting original version of manuscript. AMA, AEN, Mabd reviewing and editing manuscript. AEA, ALH, and ZKE supervised, managed and funded the research. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the Templeton World Charity Foundation, Inc. (TWCF) project TWCF0400 'Safeguarding crop diversity for food security: Pre-breeding complemented with Innovative Finance', which is managed by the Global Crop Diversity Trust (https://www.cwrdiversity.org/).

Data availability
All data generated or analysed during this study are included in this published article and its supplementary information files. Publicly available datasets used were sourced from NCBI SRA under accession numbers SRP045652, SRP092875, SRP145030, and SRP327502, and the grass pea genome assembly CABITX010000000.

Declarations

Ethics approval and consent to participate
Not applicable (no human subjects). All authors consented to their participation in this research.

Consent for publication
Not applicable (no human subjects). All authors consent to the publication of this manuscript.

Competing interests
The authors declare no competing interests.

Author details
[1]Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), Giza P.O. 12619, Egypt. [2]International Center for Agricultural Research in the Dry Areas, (ICARDA), Giza P.O. 11742, Egypt. [3]National Biotechnology Network of Expertise, ASRT, Cairo, Egypt. [4]Bioinformatics Laboratory, College of Computing, Mohammed VI Polytechnic University, Ben Guerir, Morocco. [5]Biodiversity and Crop Improvement Program, International Center for Agricultural Research in the Dry Areas, Rabat 10100, Morocco.

Received: 4 October 2023 Accepted: 1 May 2025
Published online: 21 May 2025

References

1. Kumar S, Bejiga G, Ahmed S, Nakkoul H, Sarker A. Genetic improvement of grass pea for low neurotoxin (β-ODAP) content. Food Chem Toxicol. 2011;49(3):589–600.
2. Dixit GP, Parihar AK, Bohra A, Singh NP. Achievements and prospects of grass pea (Lathyrus sativus L.) improvement for sustainable food production. Crop J. 2016;4(5):407–416.
3. Grela ER, Rybiński W, Matras J, Sobolewska S. Variability of phenotypic and morphological characteristics of some Lathyrus sativus L. and Lathyrus cicera L. accessions and nutritional traits of their seeds. Genet Resour Crop Evol. 2012;59(8):1687–1703.
4. Aci MM, Lupini A, Badagliacca G, Mauceri A, Lo Presti E, Preiti G. Genetic diversity among Lathyrus ssp. based on agronomic traits and molecular markers. Agronomy. 2020;10(8):1182.
5. Alfatih A, Zhang J, Song Y, Jan S, Zhang Z, Xia J, et al. Nitrate-responsive OsMADS27 promotes salt tolerance in rice. Plant Commun. 2023;4(2):100458. https://doi.org/10.1016/j.xplc.2022.100458.
6. 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. 2021;12(11):1799.
7. Jia J, Zhao P, Cheng L, Yuan G, Yang W, Liu S, et al. MADS-box family genes in sheepgrass and their involvement in abiotic stress responses. BMC Plant Biol. 2018;18:1–11.
8. Hamza R, Roque E, Gómez-Mena C, Madueño F, Beltrán J, Cañas L. PsEND1 is a key player in pea pollen development through the modulation of redox homeostasis. Front Plant Sci. 2021;12. https://doi.org/10.3389/fpls.2021.765277.
9. Zhang X, Fatima M, Zhou P, Ming R. Analysis of MADS-box genes revealed modified flowering gene network and diurnal expression in pineapple. 2019. Preprint. https://doi.org/10.21203/rs.2.10423/v1.
10. Raza Q, Riaz A, Atif R, Hussain B, Rana I, Ali Z, et al. Genome-wide diversity of MADS-box genes in bread wheat is associated with its rapid global adaptability. Front Genet. 2022;12. https://doi.org/10.3389/fgene.2021.818880.
11. Xu N, Chu Y, Chen H, Li X, Wu Q, Liang J, et al. Rice transcription factor OsMADS25 modulates root growth and confers salinity tolerance via the ABA–mediated regulatory pathway and ROS scavenging. PLoS Genet. 2018;14(10):e1007662. https://doi.org/10.1371/journal.pgen.1007662.
12. Parsa M, Kashanchi M, Zeinali A, Pourfakhraei E. Overexpression of zinc finger (GPZF) promotes drought tolerance in grass pea (Lathyrus sativus). 2021. Preprint. https://doi.org/10.1101/2021.10.19.464954.
13. Liu J, Zhang J, Zhang J, Miao H, Wang J, Gao P, et al. Genome-wide analysis of banana MADS-box family closely related to fruit development and ripening. Sci Rep. 2017;7(1):1–13.
14. 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(7):1538–51.
15. De Bodt S, Raes J, Florquin K, Rombauts S, Rouzé P, Theißen G, et al. Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J Mol Evol. 2003;56:573–86.
16. Vickers NJ. Animal communication: when i'm calling you, will you answer too? Curr Biol. 2017;27(14):R713–5.
17. 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.
18. Huo S, Li Y, Li R, Chen R, Xing H, Wang J, et al. Genome-wide analysis of the MADS-box gene family in Rhododendron hainanense Merr. and expression analysis under heat and waterlogging stresses. Ind Crop Prod. 2021;172:114007.
19. Wei B, Zhang RZ, Guo JJ, Liu DM, Li AL, Fan RC, et al. Genome-wide analysis of the MADS-box gene family in Brachypodium distachyon. PLoS ONE. 2014;9(1):e84781.
20. Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, et al. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. Plant J. 2000;24(4):457–66.
21. Riechmann JL, Meyerowitz EM. MADS domain proteins in plant development. Biol Chem. 1997;378(10):1079–102.
22. Moore S, Vrebalov J, Payton P, Giovannoni J. Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. J Exp Bot. 2002;53(377):2023–30.
23. Lin Z, Cao D, Damaris RN, Yang P. Genome-wide identification of MADS-box gene family in sacred lotus (Nelumbo nucifera) identifies a SEPALLATA homolog gene involved in floral development. BMC Plant Biol. 2020;20(1):1–15.
24. Alsamman AM, Mousa KH, Nassar AE, Faheem MM, Radwan KH, Adly MH, et al. Identification, characterization, and validation of NBS-encoding genes in grass pea. Front Genet. 2023;14:1187597.
25. Hu L, Liu S. Genome-wide analysis of the MADS-box gene family in cucumber. Genome. 2012;55(3):245–56.
26. Wang D, Pajerowska-Mukhtar K, Culler AH, Dong X. Salicylic acid inhibits pathogen growth in plants through repression of the auxin signaling pathway. Curr Biol. 2007;17(20):1784–90.
27. Raskin I. Role of salicylic acid in plants. Annu Rev Plant Biol. 1992;43(1):439–63.
28. Zhou M, Memelink J. Jasmonate-responsive transcription factors regulating plant secondary metabolism. Biotechnol Adv. 2016;34(4):441–9.
29. Liu L, Xu W, Hu X, Liu H, Lin Y. W-box and G-box elements play important roles in early senescence of rice flag leaf. Sci Rep. 2016;6(1):20881.
30. Kobayashi K, Obayashi T, Masuda T. Role of the G-box element in regulation of chlorophyll biosynthesis in Arabidopsis roots. Plant Signal Behav. 2012;7(8):922–6.
31. Loake GJ, Faktor O, Lamb CJ, Dixon RA. Combination of H-box [CCT ACC (N) 7CT] and G-box (CACGTG) cis elements is necessary for feedforward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate p-coumaric acid. Proc Natl Acad Sci. 1992;89(19):9230–4.
32. Kawagoe Y, Campell BR, Murai N. Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein β-phaseolin} gene. Plant J. 1994;5(6):885–90.
33. Duan W, Song X, Liu T, Huang Z, Ren J, Hou X, et al. Genome-wide analysis of the MADS-box gene family in Brassica rapa (Chinese cabbage). Mol Gen Genomics. 2015;290:239–55.
34. Kim D, Parupalli S, Azam S, Lee S, Varshney R. Comparative sequence analysis of nitrogen fixation-related genes in six legumes. Front Plant Sci. 2013;4. https://doi.org/10.3389/fpls.2013.00300.
35. Restrepo-Montoya D, McClean P, Osorno J. Orthology and synteny analysis of receptor-like kinases "RLK" and receptor-like proteins "RLP"

in legumes. BMC Genomics. 2021;22(1). https://doi.org/10.1186/s12864-021-07384-w.
36. Garg V, Agarwal G, Pazhamala L, Nayak S, Kudapa H, Khan A, et al. Genome-wide identification, characterization, and expression analysis of small RNA biogenesis purveyors reveal their role in regulation of biotic stress responses in three legume crops. Front Plant Sci. 2017;8. https://doi.org/10.3389/fpls.2017.00488.
37. Laurie R, Diwadkar P, Jaudal M, Zhang L, Hecht V, Wen J, et al. The Medicago flowering locus T homolog, MtFTa1, is a key regulator of flowering time. Plant Physiol. 2011;156(4):2207–24. https://doi.org/10.1104/pp.111.180182.
38. Campbell CG. Grass pea, Lathyrus sativus L, vol 18. Bioversity International; 1997.
39. Smaczniak C, Immink RG, Angenent GC, Kaufmann K. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development. 2012;139(17):3081–98.
40. Wei X, Wang L, Yu J, Zhang Y, Li D, Zhang X. Genome-wide identification and analysis of the MADS-box gene family in sesame. Gene. 2015;569(1):66–76.
41. Airoldi CA, Davies B. Gene duplication and the evolution of plant MADS-box transcription factors. J Genet Genomics. 2012;39(4):157–65.
42. 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(1):1–21.
43. Xu Z, Zhang Q, Sun L, Du D, Cheng T, Pan H, et al. Genome-wide identification, characterisation and expression analysis of the MADS-box gene family in Prunus mume. Mol Gen Genomics. 2014;289:903–20.
44. Zhou H, Li G, Zhao X, Kang Z, Guo A. Genome-Wide sequence identification of MADS-box transcription factor gene family from Fragaria vesca and cloning of FaMADS1 gene from F. ananassa Fruit. Plant Phys J. 2014;50:1630–1638.
45. Wells CE, Vendramin E, Jimenez Tarodo S, Verde I, Bielenberg DG. A genome-wide analysis of MADS-box genes in peach [Prunus persica (L.) Batsch]. BMC Plant Biol. 2015;15:1–15.
46. Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ. Genome-wide analysis of MIKCC-type MADS box genes in grapevine. Plant Physiol. 2009;149(1):354–69.
47. Sheng XG, Zhao ZQ, Wang JS, Yu HF, Shen YS, Zeng XY, et al. Genome wide analysis of MADS-box gene family in Brassica oleracea reveals conservation and variation in flower development. BMC Plant Biol. 2019;19(1):1–15.
48. 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(2):277–90.
49. Chung BY, Simons C, Firth AE, Brown CM, Hellens RP. Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. BMC Genomics. 2006;7:1–13.
50. Zhao Y, Zhao H, Wang Y, Zhang X, Zhao X, Yuan Z. Genome-Wide Identification and Expression Analysis of MIKC-Type MADS-Box Gene Family in Punica granatum L. Agronomy. 2020;10(8):1197.
51. Mondragón-Palomino M, Theißen G. Conserved differential expression of paralogous DEFICIENS-and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the 'orchid code'. Plant J. 2011;66(6):1008–19.
52. Alvarez-Buylla ER, Pelaz S, Liljegren 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. 2000;97(10):5328–33.
53. Hakes L, Pinney JW, Robertson DL, Lovell SC. Protein-protein interaction networks and biology–what's the connection? Nat Biotechnol. 2008;26(1):69–72.
54. Yocca AE, Edger PP. Current status and future perspectives on the evolution of cis-regulatory elements in plants. Curr Opin Plant Biol. 2022;65:102139.
55. Alhindi T, Al-Abdallat AM. Genome-wide identification and analysis of the MADS-box gene family in American beautyberry (Callicarpa americana). Plants. 2021;10(9):1805.
56. Yu Z, Zhang G, da Silva JAT, Zhao C, Duan J. The methyl jasmonate-responsive transcription factor DobHLH4 promotes DoTPS10, which is involved in linalool biosynthesis in Dendrobium officinale during floral development. Plant Sci. 2021;309:110952.
57. Dhatterwal P, Basu S, Mehrotra S, Mehrotra R. Genome wide analysis of W-box element in Arabidopsis thaliana reveals TGAC motif with genes down regulated by heat and salinity. Sci Rep. 2019;9(1):1681.
58. Li R, Zhu F, Duan D. Function analysis and stress-mediated cis-element identification in the promoter region of VqMYB15. Plant Signal Behav. 2020;15(7):1773664.
59. Dolferus R, Klok EJ, Ismond K, Delessert C, Wilson S, Good A, et al. Molecular basis of the anaerobic response in plants. IUBMB Life. 2001;51(2):79–82.
60. 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. 2021;12(12):1956.
61. Ning K, Han Y, Chen Z, Luo C, Wang S, Zhang W, et al. Genome-wide analysis of MADS-box family genes during flower development in lettuce. Plant Cell Environ. 2019;42(6):1868–81.
62. Kang YI, Choi YJ, Lee YR, Seo KH, Suh JN, Lee HR. Cut flower characteristics and growth traits under salt stress in lily cultivars. Plants. 2021;10(7):1435.
63. Saha G, Park JI, Jung HJ, Ahmed NU, Kayum MA, Chung MY, et al. Genome-wide identification and characterization of MADS-box family genes related to organ development and stress resistance in Brassica rapa. BMC Genomics. 2015;16(1):1–21.

Publisher's Note