

In silico analysis of anther-specific transcription factors belonging to bHLH, MADS-box, and MYB families in *Oryza sativa* and *Arabidopsis*

Naina Miglani[#], Preeti Agarwal[#], Shreya Shrivastava, Radhika, Aparajita Mohanty & Reema Mishra*

Department of Botany, Gargi College, University of Delhi, Siri Fort Road, New Delhi-110 049, Delhi, India

Received 05 October 2025; revised 17 August 2026

Transcription factors are important regulatory proteins that orchestrate plant developmental processes, including the complex regulation of anther development, which is essential for reproductive success in flowering plants. This study analyses three major transcription factor families, bHLH, MYB, and MADS-box, that play crucial roles in anther development in two important model plant species, *Oryza sativa* and *Arabidopsis thaliana*. A total of 48 (20 from rice and 28 from *Arabidopsis*) anther-specific transcription factor coding genes were identified through literature mining. Phylogenetic and protein domain analysis established conserved evolutionary relationships between rice and *Arabidopsis*, suggesting both functional conservation and divergence across monocots and dicots. Chromosomal localization revealed irregular distribution pattern of genes in both plants. Expression analysis showed rice transcription factor genes exhibited maximal expression during pre-meiotic and meiotic stages of anther development, while in *Arabidopsis*, they showed predominant expression in flower pedicels and carpels. Protein-protein interaction analysis revealed that most anther-specific proteins interact with proteins involved in stamen and tapetum development. These findings offer valuable insights into the molecular regulation of anther development and identify candidate transcription factor genes that may be employed for hybrid seed production. Understanding these regulatory mechanisms will contribute to global food security.

Keywords: *Arabidopsis*, Gene expression, Reproductive development, Rice, Transcription factors

Transcription factors (TFs) have gained significant attention in recent years, due to their crucial role in regulating genes essential for various vital plant processes. A key feature of a TF is the presence of a conserved DNA-binding domain (DBD), which varies from a simple basic alpha helix to a complex network of alpha, beta, and loop folds¹. These DNA-binding proteins mediate the various biological processes by recognizing specific *cis*-regulatory DNA sequences located in the promoter regions of respective target genes². They are primarily known for performing two

crucial functions: first, identifying the correct DNA binding site, and second, assembling other regulatory proteins to efficiently execute the transcriptional process³. Several plant TF databases have been developed in recent years to catalogue the TF-encoding genes in plants. The model plant *Arabidopsis* has the largest number of TFs (approximately 2000). The presence of a characteristic DBD serves as the basis of categorizing TFs into their respective families⁴. To date, scientists have characterized numerous TFs involved in a myriad of plant processes, including anther development.

The current study primarily focuses on three crucial classes of TFs, including bHLH (basic helix-loop-helix), MYB (myeloblastosis), and MADS (M-MINICHROMOSOME MAINTENANCE 1- A-AGAMOUS D- DEFICIENS S-SERUM RESPONSE FACTOR) domain-containing proteins, for their involvement in anther development in rice and *Arabidopsis*.

The basic structural composition of a bHLH TF, the second largest class of TFs, includes two highly conserved domains, which comprises of 60 amino-acid long residue region⁵. The first domain, known as the basic domain, and the second domain, the typical

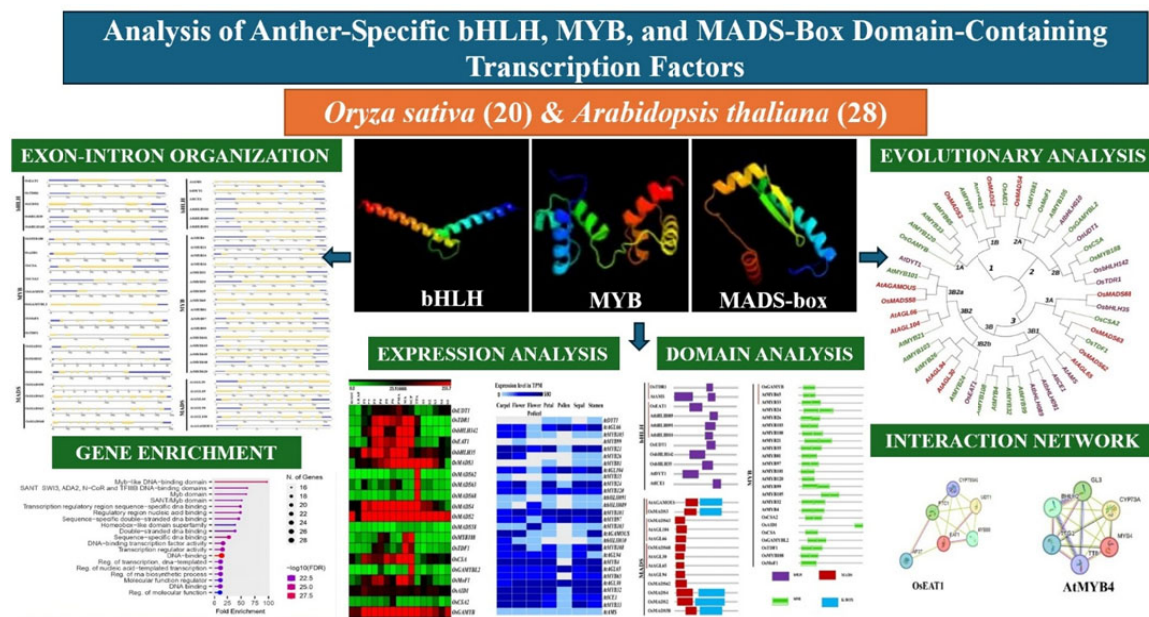
[#]The authors have contributed equally

Correspondence:

E-mail: reema.mishra@gargi.du.ac.in

Suppl. data available on respective page of NOPR

Abbreviations: bHLH, Basic helix-loop-helix; DBD, DNA-binding domain; GEO, Gene Expression Omnibus; GO, Gene Ontology; iTOL, Interactive Tree of Life; MA, Meiotic anther; MADS, M-MINICHROMOSOME MAINTENANCE 1- A- AGAMOUS D- DEFICIENS S-SERUM RESPONSE FACTOR; MEGA, Molecular Evolutionary Genetics Analysis; MeV, Multiple Experiment Viewer; MYB, Myeloblastosis; PCD, Programmed cell death; pI, isoelectric point; PMA, Pre-meiotic anther; RGAP, Rice Genome Annotation Project; SCP, Single cell pollen containing anther; TFs, Transcription factors; TPA, Trinucleate pollen containing anther; URRs, Upstream regulatory regions



Graphical abstract

HLH domain, are located at the N-terminus and C-terminus of the amino acid sequence, respectively, and are followed by two amphipathic α -helices⁶. Interestingly, due to their diverse nature, several past studies have explored their role in anther development along with their role in other crucial plant processes⁷⁻¹⁰. To date, 167 bHLH TFs in *Arabidopsis* and 165 in rice have been identified^{8,11}.

The key feature of a MADS-box TF is the presence of an evolutionarily conserved DBD called MADS domain or M domain. They bind to DNA usually as obligate dimers via recognition of the 'CARG' box motif¹². Studies to date have identified approximately 107 Type I and Type II MADS-box TFs in the model plant *Arabidopsis*, with 61 TFs belonging to Type I and 45 to Type II¹³. Similarly, in rice, 76 MADS-box members (MIKCC, MIKC, $M\alpha$, $M\beta$, and $M\gamma$) have been identified, comprising 32 Type I and 44 Type II TFs¹⁴. In the case of MYB proteins, the presence of a highly conserved DBD termed the MYB domain serves as a key categorizing feature. Being one of the largest classes of TFs, these proteins have been shown to regulate a variety of plant responses, including tolerance against several abiotic stresses, cell cycle control, meristem formation, and hormone signaling¹⁵. As per the studies done so far, an approximate number of 155 MYB genes in rice and 197 in *Arabidopsis* have been identified¹⁶.

Anther development is one of the key events in the life cycle of a sexually reproducing plant and is

crucial for ensuring successful male reproductive development. The generation of hybrid seeds by manipulating anther-specific genes for inducing male sterility provides a valuable alternative to conventional breeding techniques.

Rice is a crucial agricultural crop that represents all monocot species, while *Arabidopsis* serves as a model plant for all dicot crops and is widely regarded as a "lab rat" for conducting a broad range of scientific experiments. Given the immense potential of both of these crops to revolutionize the breeding and agricultural industry, the present study attempts to draw a comparison between the roles of the anther-specific TFs of these two plant species. The present study documents 48 (20 of rice and 28 of *Arabidopsis*) anther-specific TFs belonging to bHLH, MYB, and MADS-box classes. Domain analysis, phylogenetic relationships, gene ontology, and expression profiling provide deeper insights into their role in regulating the anther development process. Additionally, orthologous gene pair analysis further highlights the relationship between anther-specific genes of rice and *Arabidopsis*.

Materials and Methods

Data collection

An extensive literature search using Google Scholar and PubMed database (keywords- anther-specific transcription factors, rice, *Arabidopsis*, bHLH, MADS-box, MYB) resulted in the identification of 48 (20 in rice and 28 in *Arabidopsis*) anther-specific TFs coding genes (Suppl. Tables 1 and 2).

Domain analysis

The domain analysis of anther-specific proteins of rice and *Arabidopsis* was performed using the NCBI CDD database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), InterPro database (<https://www.ebi.ac.uk/interpro/>), SMART, a web tool (<http://smart.embl-heidelberg.de/>) and Pfam (<http://pfam-legacy.xfam.org/search/sequence>) database.

Exon-Intron structural analysis

Gene Structure Display Server 2.0 (<http://gsds.gao-lab.org/>) was employed to analyze exon-intron structural features of anther-specific TFs belonging to the bHLH, MYB, and MADS-box protein families.

Chromosome localization

For chromosome mapping of the anther-specific genes belonging to bHLH, MYB, and MADS-box families, Oryza base map tool (<http://viewer.shigen.info/oryzavw/maptool/MapTool.do>) for rice and TAIR map tool (<https://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp>) for *Arabidopsis* was used.

Phylogenetic analysis

CLUSTAL MUSCLE software was used with the default settings to perform multiple sequence alignment of anther-specific proteins of rice and *Arabidopsis*, followed by the construction of a phylogenetic tree using Molecular Evolutionary Genetics Analysis (MEGA) (<https://www.megasoftware.net>) version 11 software. Further, the Interactive Tree of Life (iTOL) research tool was used for viewing and coloring the phylogenetic tree.

Promoter analysis for the presence of *Cis*-regulatory elements

New PLACE database (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>) was used for determining the *cis*-regulatory motif sites present in upstream regulatory regions (URRs) of anther-specific TF coding genes. URRs were obtained from the 'Sequences' section of TAIR for *Arabidopsis*, while for rice, the URR sequences were downloaded from the 'Downloads' section of the Rice Genome Annotation Project (RGAP).

Gene ontology

To identify enriched GO (Gene Ontology) terms of genes coding for anther-specific TFs of bHLH, MYB, and MADS-box class, ShinyGO 0.77 (<http://bioinformatics.sdstate.edu/go/>) graphical tool was employed. The locus IDs from TAIR and RGAP for *Arabidopsis* and rice were used, respectively.

Expression analysis

Expression analysis of anther-specific TFs in rice was conducted across vegetative tissues (roots and leaves), reproductive tissues including panicle (P1-P6), anther [Pre-meiotic anther (PMA), Meiotic anther (MA), single cell pollen containing anther (SCP), and trinucleate pollen containing anther (TPA)], and different stages of seed (S1-S5) development. Expression Heat Map was built using Multiple Experiment Viewer (MeV) (<https://mybiosoftware.com/mev-4-6-2-multiple-experiment-viewer.html>) version 4.9.0 using raw expression data downloaded from the NCBI Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) in the case of rice (GSE6893 and GSE27726), and for *Arabidopsis*, the Expression Atlas tool (<https://www.ebi.ac.uk/gxa/experiments/E-MTAB-7978/Results>) was used. To further corroborate the gene expression of rice and *Arabidopsis* anther-specific TFs across various tissue types, the RNA-seq database (<https://plantnadb.com/athrdb/>) was examined and expression data plots were constructed.

Physical and Chemical properties, Protein structure and Protein-protein interaction

For understanding the physical and chemical properties of anther-specific proteins of rice and *Arabidopsis*, Prot Param ExPasy tool (<https://web.expasy.org/protparam>) was selected. Through this tool, different parameters, namely length, molecular weight, stability, instability, and pI (isoelectric point) values, were retrieved. In addition to this, subcellular localization of proteins was also documented using ProtComp Version 9.0 (<http://www.softberry.com/>). Furthermore, the structures of proteins were generated using the amino acid sequences of rice and *Arabidopsis* through the Phyre 2 tool (www.sbg.bio.ic.ac.uk/phyre2), with the 'expert mode' option selected for visualizing the results.

STRING (<https://string-db.org/>) version 11.5 was used to perform protein-protein interaction analysis for rice and *Arabidopsis* genes using their protein sequences or locus IDs. This web-based tool was used to visualize different interactions in the form of a network¹⁷. The maximum number of interactors was set to 5 at a medium confidence level of 0.4.

Results

Domain architecture of anther-specific TFs belonging to bHLH, MYB, and MADS-box families of rice and *Arabidopsis*

Protein sequences of 48 (20 of rice and 28 of *Arabidopsis*) anther-specific TFs were analyzed for the presence of conserved domains using the NCBI

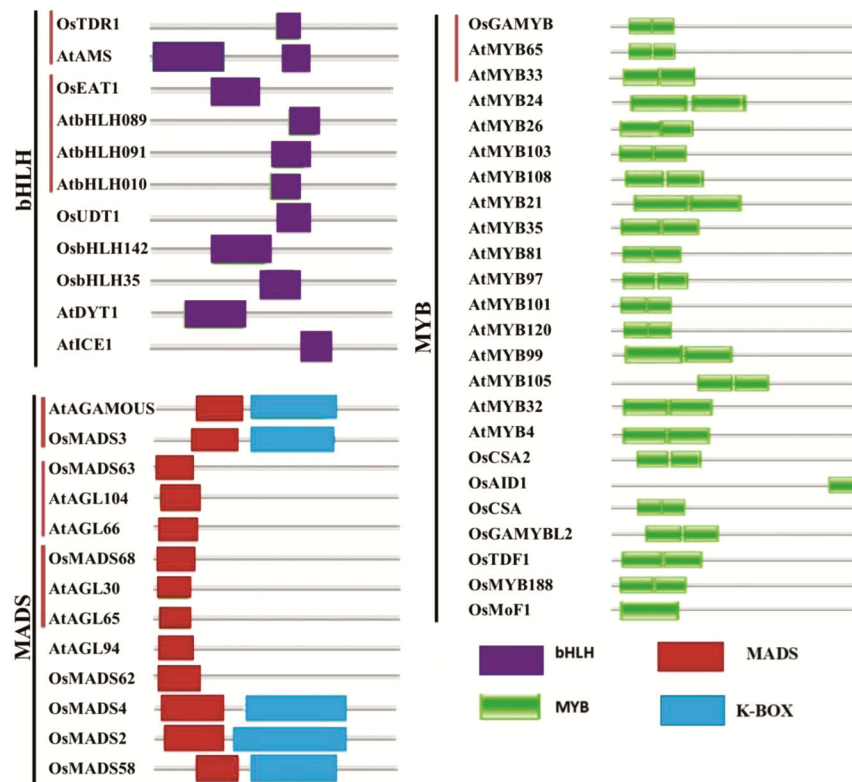


Fig. 1 — Domain analysis of anther-specific TFs in rice and *Arabidopsis*. (The purple box indicates the bHLH domain, the red box signifies the MADS-box domain, the green box indicates the MYB domain, and the blue box represents the K-box domain and 6 orthologous gene pairs are denoted with red vertical lines)

CDD, SMART, InterPro, and Pfam databases. The four types of domains that were identified in different TFs are bHLH, MYB, MADS-box, and K-box. Further, six orthologous gene pairs of rice and *Arabidopsis* were also grouped together (Fig. 1). Two orthologous pairs belonged to bHLH TF class (OsTDR1 and AtAMS; OsEAT1 and AtbHLH089, AtbHLH091, AtbHLH010), three from MADS-box TF family (OsMADS3 and AtAGAMOUS; OsMADS63 and AtAGL104, AtAGL66; OsMADS68 and AtAGL65, AtAGL30), and one from MYB TF group (OsGAMYB and AtMYB33, AtMYB65). Additionally, 23 proteins (6 from rice and 17 of *Arabidopsis*) showed the presence of two domains specific to their respective classes. In rice, this includes six MYB TFs: OsMYB188, OsCSA, OsCSA2, OsGAMYB, OsGAMYBL2, and OsTDF1. In *Arabidopsis*, the proteins include one bHLH TF, AtAMS and 16 MYB proteins (AtMYB4, AtMYB21, AtMYB24, AtMYB26, AtMYB32, AtMYB33, AtMYB35, AtMYB65, AtMYB81, AtMYB97, AtMYB99, AtMYB101, AtMYB103, AtMYB105, AtMYB108, and AtMYB120). Interestingly, members

of the MADS-box group possessed an extra K-box domain both in *Arabidopsis* (AtAGAMOUS) and rice (OsMADS3, OsMADS4, OsMADS2, and OsMADS58).

Different gene attributes such as locus ID, chromosome number, ORF length (bp), protein length (aa), pI, molecular weight in Daltons, number of exons/introns, alternatively spliced variants, putative function, CDS coordinates, and the orthologs in *Arabidopsis*, *Sorghum*, and maize were catalogued using the RGAP database (Suppl. Tables 1-3).

Gene structure: Exon-Intron structure

The exon-intron organization and the presence of untranslated region (UTR) of the anther-specific genes encoding bHLH, MYB, and MADS-box TF proteins were visualized through Gene Display Server 2.0 (Fig. 2). Gene structure analysis revealed that genes belonging to the MADS-box TF family have the maximum number of introns, with the intron number varying from 6 to 10 in both rice and *Arabidopsis*. This is followed by the bHLH TF family with intron numbers ranging from 1 to 6 and 2 to 7 in

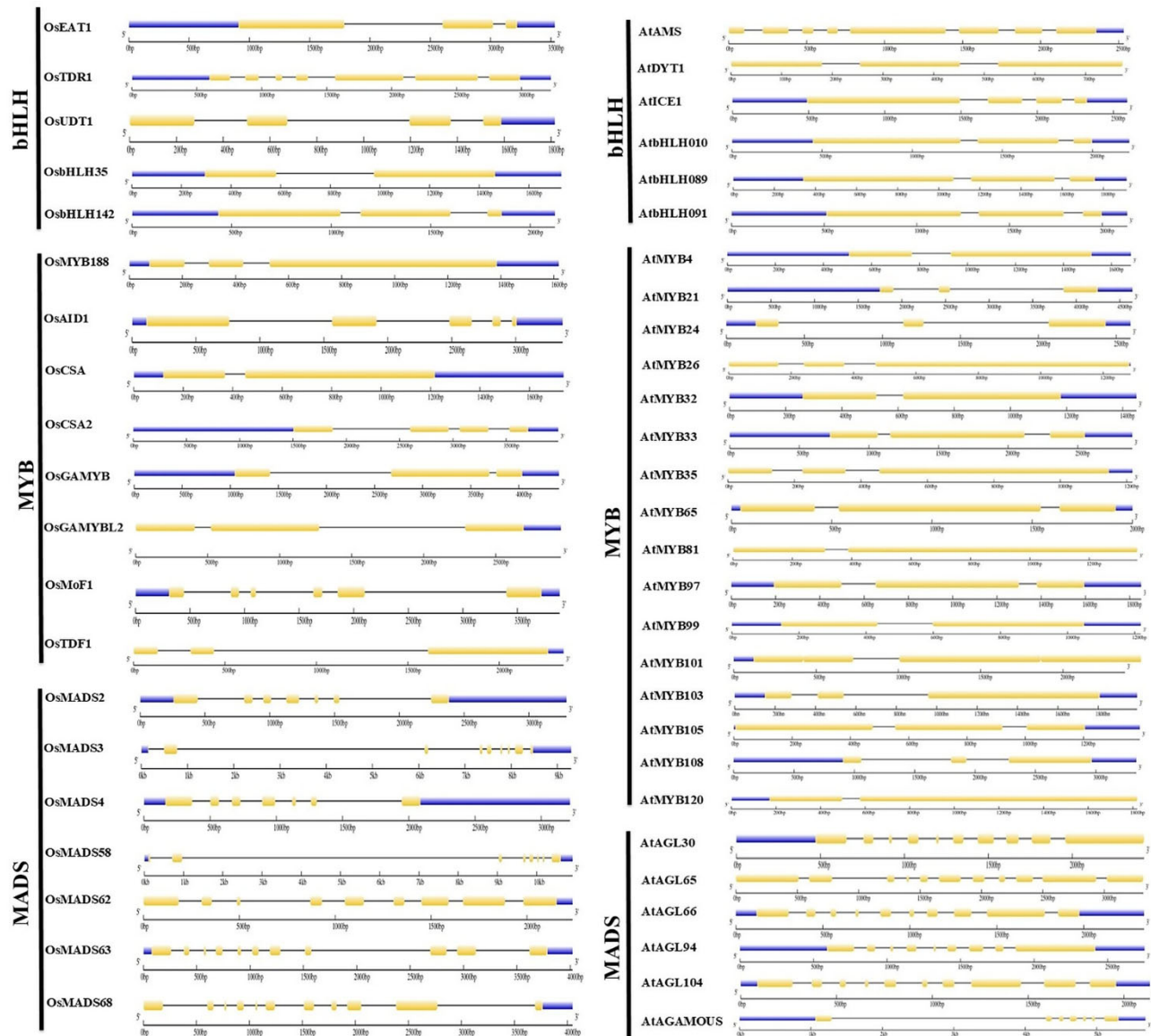


Fig. 2 — Exon-intron organization of rice and *Arabidopsis* anther-specific bHLH, MYB and MADS-box TF coding genes. Exons are represented by yellow rectangles, introns by grey lines, and untranslated regions (UTRs) by blue rectangles

rice and *Arabidopsis*, respectively. Proteins belonging to the MYB TF class exhibited the least number of introns *i.e.* 1 to 5 in rice and 1 to 2 in *Arabidopsis*.

Further, the 5' and 3' UTRs showed significant variations when compared in rice and *Arabidopsis*. In *Arabidopsis*, 1 bHLH member (*AtAMS*) and 2 MYB members (*AtMYB26* and *AtMYB35*) have only 3' UTR. Similarly, in rice, 1 bHLH (*OsUDT1*), 2 MADS-box (*OsMADS62* and *OsMADS68*), and 2 MYB members (*OsTDF1* and *OsGAMYBL2*) exhibited only 3' UTR. In *Arabidopsis*, four bHLH members (*AtbHLH089*, *AtbHLH091*, *AtbHLH010*, and *AtICE1*), four MADS-box members

(*AtAGAMOUS*, *AtAGL66*, *AtAGL94*, and *AtAGL104*), and eleven MYB members (*AtMYB4*, *AtMYB21*, *AtMYB24*, *AtMYB32*, *AtMYB33*, *AtMYB65*, *AtMYB97*, *AtMYB99*, *AtMYB103*, *AtMYB105*, and *AtMYB108*) were found to have both 5' and 3' UTRs. Similarly, in rice, four bHLH (*OsEAT1*, *OsTDR1*, *OsbHLH35*, and *OsbHLH142*), five MADS-box (*OsMADS2*, *OsMADS3*, *OsMADS4*, *OsMDAS58*, and *OsMADS63*), and six MYB members (*OsAID1*, *OsCSA*, *OsCSA2*, *OsGAMYB*, *OsMoF1*, and *OsMYB188*) exhibited both 5' and 3' UTRs. Notably, in *Arabidopsis*, two MYB genes (*AtMYB101* and *AtMYB120*) and one MADS-box gene (*AtAGL30*) showed only 5' UTR, while one

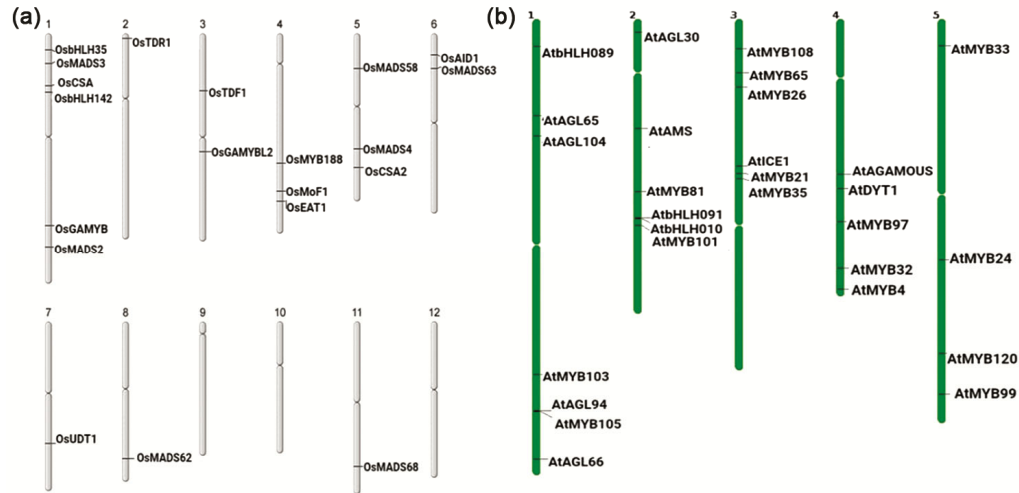


Fig. 3 — The distribution of bHLH, MYB, and MADS-box anther-specific genes across the chromosomes in (3a) rice; and (3b) *Arabidopsis*. The gene names are indicated at their corresponding chromosomal locations

bHLH (*AtDYT1*), one MYB (*AtMYB81*), and one MADS-box member (*AtAGL65*) lacked both 5' and 3' UTRs.

The largest proteins identified were OsGAMYB (554 amino acids) in rice and *AtAGL30* (666 amino acids) in *Arabidopsis*, while the smallest were OsMADS2 (210 amino acids) in rice and *AtDYT1* (207 amino acids) in *Arabidopsis* (Suppl. Tables 1 and 2).

Further, WoLF PSORT software (<https://www.genscript.com/wolf-psort.html>) was employed to study the localization of anther-specific TFs of rice and *Arabidopsis*, and the majority of them were localized in the nucleus (Suppl. Tables 1 and 2).

Chromosomal localization

To gain a better insight into rice genetics, an integrated rice science database called Oryzabase (<https://shigen.nig.ac.jp/rice/oryzabase/about/oryzabase>) was employed to determine the physical location of genes coding for anther-specific TFs in rice. Twenty such genes were mapped onto the 12 chromosomes of the rice using the Oryzabase map tool, with chromosome numbers mentioned at the top of each chromosome (Fig. 3a). The results revealed an uneven distribution pattern of these genes, with the majority of the genes (6) positioned at chromosome 1, with two genes belonging to each of the bHLH, MYB, and MADS-box TF classes, respectively. This was followed by chromosome 4 with two genes belonging to the MYB class and one from bHLH class, while chromosome 5 carried two members belonging to the MADS-box family and one from the MYB class.

Additionally, chromosomes 3 and 6 each had two genes, with chromosome 3 hosting two MYB genes and chromosome 6 comprising one MYB gene and one MADS-box gene. The majority of the rice chromosomes (2, 7, 8, and 11) possessed only a single gene belonging to the bHLH or MADS-box family. On the other hand, chromosomes 9, 10, and 12 did not show any genes coding for anther-specific TFs. Further analysis also showed the accumulation of certain genes on specific locations on specific chromosomes; for instance, on chromosome 1, the majority of the genes were positioned on the upper arm, in contrast to chromosome number 4, where all the genes were located at the lower arm. On chromosome 5, two genes out of three were located on the lower arm, and one was on the upper arm, while chromosome 3 displayed an even distribution pattern with one gene positioned on the upper arm and one on the lower arm. On chromosome 6, both the genes were located on upper arm.

TAIR chromosome map tool was utilized to determine the chromosomal localization of 28 *Arabidopsis* anther-specific TF genes. Chromosome 1 had the maximum number of genes (7), with 1 member of bHLH, 4 of MADS-box, and 2 TFs belonging to the MYB class, respectively. This was followed by chromosome numbers 2 and 3, each harboring 6 genes. Chromosome 2 had 3 members from bHLH, 2 from MYB and 1 from MADS-box, while chromosome 3 showed 1 member of the bHLH class and the other 5 genes from the MYB gene family. Furthermore, chromosomes 4 and 5 exhibited localization of 5 and 4 genes, respectively.

Chromosome 4 reported 1 member from bHLH and MADS-box and the rest 3 members belonging to the MYB family, while chromosome 5 exclusively contained MYB class genes. A closer examination of the figure (Fig. 3b) reveals patterns of gene accumulation, such as all genes on chromosome 3 being located on the upper arm, and those on chromosome 4 being concentrated on the lower arm. In chromosomes 1, 2 and 5, the majority of the genes are positioned on the lower arm.

Evolutionary analysis of bHLH, MADS-box, and MYB TFs in rice and *Arabidopsis*

For examining the evolutionary relationships between rice and *Arabidopsis* anther-specific genes, a neighbor-joining tree was constructed using MEGA 11 software. The tree was built using protein sequences of rice and *Arabidopsis* TFs listed in (Suppl. Tables 1 and 2). A total of 48 anther-specific proteins (20 from rice and 28 from *Arabidopsis*) were distributed among the 3 clades of the tree, which were further divided into 10 subclades. Clade 3 exhibited the highest number of proteins (28), followed by clade 2 with 11 proteins and lastly the smallest clade 1 with 9 proteins. Interestingly, subclade 3B1 consists entirely of members from *Arabidopsis*, while subclade 3A includes only members from rice. However, the other two subclades within clade 3B2 (3B2a and 3B2b) contain members from both rice and *Arabidopsis*. Additionally, subclade 2B contains only rice members, whereas subclade 2A includes proteins from both rice and *Arabidopsis*. Subclades 1A and 1B

showed members from both rice and *Arabidopsis* (Fig. 4).

Determination of *Cis*-regulatory elements (CREs)

The screening of *cis*-regulatory elements (CREs) present in the promoter region, *i.e.*, 1 kb upstream regulatory region (URR) of rice and *Arabidopsis* genes coding for anther-specific TFs, was performed using the New PLACE Database. Among the other elements, the CREs: S000245/POLLENILELAT52 elements possessing motif AGAAA, S000378/GTGANTG10/pollen with motif GTGA, and Q (quantitative)-element with motif AGGTCA, known for conferring anther specificity, were selected. The details of the number of motif sites present in the URR of anther-specific genes are documented in (Suppl. Tables 4 and 5). In both rice and *Arabidopsis*, the URR region of all the genes displayed the presence of CREs POLLENILELAT52 and GTGANTG10. However, the Q (quantitative)-element was found in the URR of 4 genes of rice, *i.e.* *OsEAT1*, *OsMADS4*, *OsMoF1*, and *OsCSA2* and 5 genes of *Arabidopsis* *i.e.* *AtbHLH089*, *AtMYB4*, *AtMYB105*, *AtMYB120*, and *AtAGAMOUS*.

In *Arabidopsis*, URR of *AtMYB101* gene displayed the maximum number of AGAAA motif sites (15), while for rice, *OsbHLH142* URR showed the highest number of AGAAA motif sites (10). Further, for the GTGA motif, the URR of rice *OsTDR1* displayed 17 motif sites, and *AtDYT1* showed 12 motif sites. Additionally, the URR of the *OsTDF1* and *OsMADS4* genes of rice showed the least (2) number of AGAAA and GTGA motifs, respectively, and in *Arabidopsis*, the URR of *AtMYB35* and *AtMYB97* genes displayed the lowest number of AGAAA motifs (3), while *AtMYB65* gene displayed only 1 GTGA motif.

Interestingly, the motif sites of the Q-element were surprisingly fewer in number in both rice and *Arabidopsis* URR. A total of only 4 and 5 anther-specific genes possessed Q-element motif sites in rice and *Arabidopsis*, respectively. Unlike *Arabidopsis*, where URRs of all 5 genes (*AtbHLH089*, *AtMYB105*, *AtMYB120*, *AtMYB4*, and *AtAGAMOUS*) exhibited 1 motif site, the URRs of 3 rice genes (*OsMoF1*, *OsEAT1*, and *OsMADS4*) reported 1 motif site, while the 4th gene (*OsCSA2*) displayed 3 Q-elements.

Gene ontology

ShinyGO 0.77 tool was employed to perform the enrichment analysis of gene sets coding for anther-specific TFs in rice and *Arabidopsis*. Locus IDs of the genes were used to obtain the information regarding

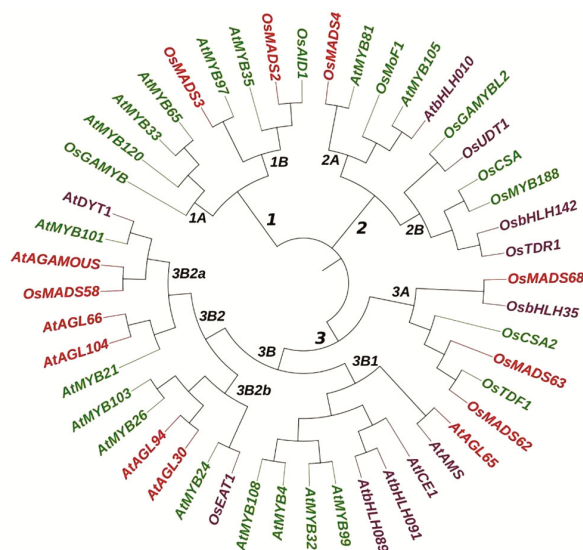


Fig. 4 — A combined phylogenetic tree showing the distribution of 48 anther-specific proteins of rice and *Arabidopsis*. The tree is divided into three main clades, which are further subdivided into 10 subclades (1A, 1B, 2A, 2B, 3A, 3B, 3B1, 3B2, 3B2a, and 3B2b)

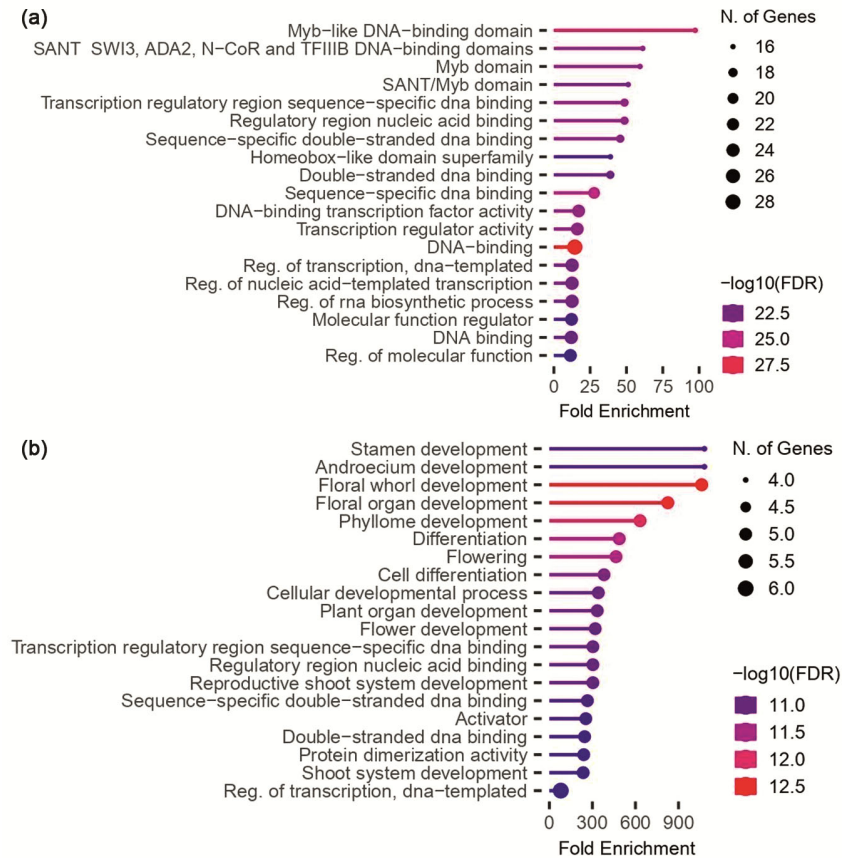


Fig. 5 — Gene enrichment analysis of rice (a) and (b) *Arabidopsis*. The Fold enrichment values are shown at the bottom right corner of both Fig. 5a and 5b and are represented in log values in their decreasing order. The purple color signifies the highest ranked values, while the magenta color indicates the values occupying the medium rank, and the red color represents the lowest ranked values. The respective biological pathways are mentioned on the left side of the figure, and the number of genes is represented by dots at the top right corner

the biological pathways. The top 19 and 20 GO terms of rice and *Arabidopsis* genes, along with their fold enrichment values, ranking from highest to lowest, have been depicted in the figure (Fig. 5). The highest and lowest ranks were concluded based on the FDR values (log score). In rice, the highest rank was observed for the MYB-like DNA-binding domain, and the lowest was for the regulation of molecular function, while in *Arabidopsis*, the highest-ranked value belonged to stamen development, and lowest-ranked was seen for regulation of transcription.

Expression profiling

Expression analysis of 20 rice anther-specific genes (coding for bHLH, MYB, and MADS-box domain containing TFs) in different vegetative (root and leaf), reproductive [panicle (P1-P6), anther (PMA, MA, SCP, and TPA)] and seed (S1-S5) developmental stages was performed. A heat map displaying the gene expression patterns was generated using MeV (Multiple Experiment Viewer) software (Fig. 6). The

transcripts of most of the bHLH domain encoding TF genes were observed in panicles and anthers. Among them *OsUDT1* exhibited the highest expression in PMA stage, while *OsTDRI*, *OsEAT1*, *OsBHLH142* and *OsBHLH35* showed maximum transcript accumulation in different stages of panicle and anther development.

Unlike the bHLH domain-containing TF family, the MADS-box family members displayed differential expression patterns in different tissues and stages analyzed. The transcripts of *OsMADS3* and *OsMADS4* were detected in different stages of panicles, anthers and seeds. Interestingly, *OsMADS62*, *OsMADS63*, and *OsMADS68* showed TPA-specific expression, while *OsMADS58* showed low-level expression across all the tissues analyzed.

The majority of the rice genes coding for MYB domain-containing proteins showed peak expression in reproductive tissues. These include *OsMYB188* and *OsGAMYBL2* accumulating the maximum transcript

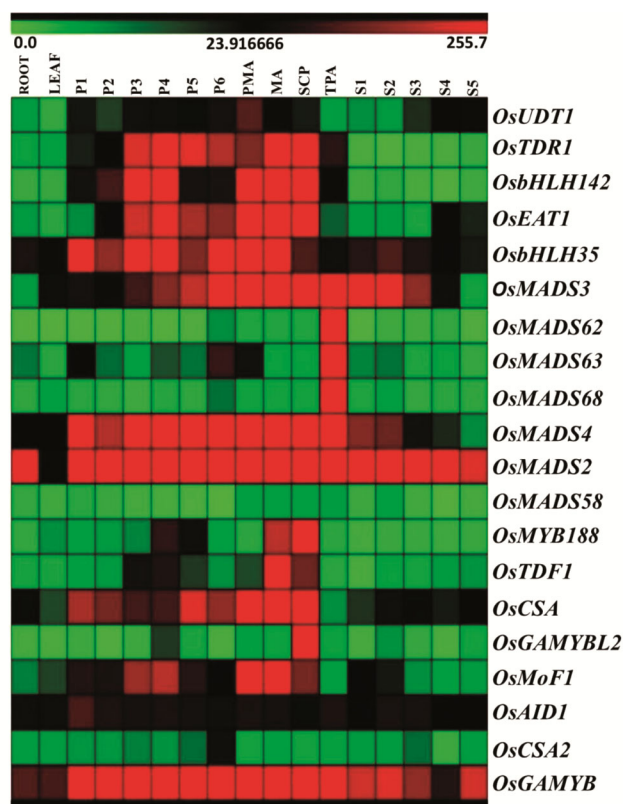


Fig. 6 — Heat map illustrating the expression patterns of bHLH, MYB, and MADS-box genes encoding anther-specific TFs in rice across various vegetative and reproductive tissues at different developmental stages. The color bar at the top of the figure indicates the expression values, with green representing low expression levels, black denoting moderate expression levels, and red illustrating high expression levels. The different developmental stages are mentioned at the top. (PMA- pre-meiotic anther, MA- meiotic anthers, SCP- single-celled pollen, TPA- tri-nucleate pollen containing anther, P1-P6 indicates stages of panicle, and S1 to S5 indicates stages of seed development)

levels in SCP stage, while *OsTDF1* showed peak expression in MA stage. *OsCSA* displayed high expression in MA, along with PMA and SCP stages, and lastly, *OsMoF1* displayed maximum expression during both PMA and MA stages.

It was further observed that *OsMADS2* and *OsGAMYB* exhibited high expression levels in all the tissues, while *OsMADS58* (MADS-box) and *OsCSA2* of the MYB family exhibited low expression levels across all the tissues examined. Interestingly, *OsAID1* (Fig. 6) reported moderate amounts of expression in all the tissues studied.

The expression profile of 28 *Arabidopsis* anther-specific genes across different tissues namely carpel, flower, flower pedicel, petal, pollen, sepal, and stamens was generated using the Expression

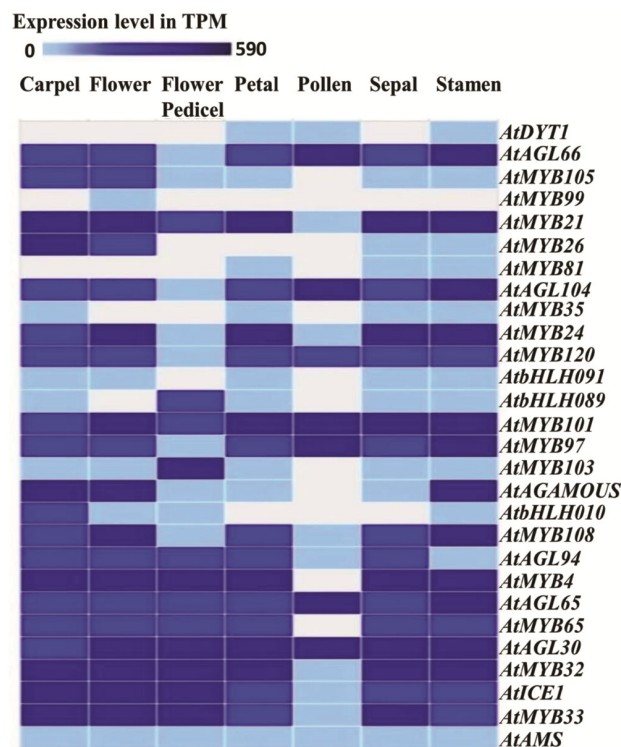


Fig. 7 — Heat map displaying the expression levels of *Arabidopsis* bHLH, MYB, and MADS-box domain encoding TF genes in different reproductive tissues. The color bar at the top of the figure represents expression values. In the figure, the light blue color refers to low levels of expression, while a darker shade of blue signifies medium levels, and the darkest shade of blue depicts the highest expression levels. Different developmental stages in which the expression of the genes is observed are mentioned at the top, while the names of the genes are shown on the right side of the expression map

Atlas tool (<https://www.ebi.ac.uk/gxa/experiments/EMTAB-7978/Results>) (Fig. 7). The bHLH domain-encoding genes exhibited varied expression patterns across the different tissues examined. *AtDYT1* was expressed at low levels in petals, pollen, and stamens, and it did not show expression in any other tissues. In contrast, *AtAMS* displayed moderate expression across all tissues. *AtICE1* displayed higher expression levels in all the tissues except pollen. Additionally, *AtbHLH010* expression was found to be highest in carpels, while *AtbHLH089* was highly expressed in flower pedicels. Conversely, *AtbHLH091* showed low expression across all tissues, with no transcripts detected in either flower pedicels or pollen.

In case of MADS-box group of proteins, *AtAGL66* and *AtAGL104* showed higher expression in all tissues except the flower pedicel, while transcripts of *AtAGL65* and *AtAGL30* were present in all tissues with the highest expression levels in pollen. *AtAGAMOUS* exhibited maximum expression in

carpel, flower, and stamens; interestingly *AtAGL94* showed higher expression in all the tissues studied except pollen and stamens.

Among all the three classes taken in this study, expression profiling of *MYB* genes of *Arabidopsis* displayed the maximum variations. This includes expression of *AtMYB105* and *AtMYB26*, which expressed maximally in the carpel and flowers; *AtMYB103* in the flower pedicel; *AtMYB99* exclusively in flowers (at low levels); *AtMYB81* expressed solely in petals, sepals, and stamens at minimal levels and transcripts of *AtMYB35*, were detected at low levels in petals, sepals, stamens, and carpels. The other members of the *MYB* family, including *AtMYB21*, *AtMYB24*, *AtMYB32*, *AtMYB33*, *AtMYB97*, *AtMYB101*, *AtMYB108*, and *AtMYB120* were expressed significantly in all the tissues under study. Notably, transcripts of two *MYB* TF genes, *AtMYB4* and *AtMYB65* were detected in all the tissues except the pollen.

The anther-specific expression of some of the TF proteins in rice and *Arabidopsis* was further analyzed using the RNA-seq database. Expression levels were examined across various tissues, including root, leaf, seedlings, shoot, stem, meristem, flower, seed, embryo, pollen, endosperm, panicle, female reproductive tissue, male reproductive tissue, and silique. As expected, transcripts for most of the TFs of rice in our study were predominantly found in the male reproductive tissue (green), panicle (blue) and flower (pink) and for *Arabidopsis* the expression was high in pollen (gray) and flower (Fig. 8a and 8b). In rice, *OsTDR1*, *OsBHLH142*, *OsMYB188*, *OsCSA2*, *OsMADS63* and *OsTDF1* gene transcripts were primarily observed in the male reproductive system, while in *Arabidopsis*, the highest expression of *AtAMS*, *AtDYT1*, *AtMYB97*, *AtMYB120*, *AtAGL66*, and *AtAGL104* was observed in the flower and pollen. These findings are consistent with the heat-map expression analysis described earlier and further support the crucial role of these proteins in anther development.

Protein-protein interactions

STRING database (<https://string-db.org/>) was employed for determining the interacting partners of different anther-specific proteins of rice and *Arabidopsis*. The data obtained was divided into three categories, *i.e.* curated, experimentally determined, and data with the highest score (Suppl. Tables 6 and 7). Out of these three, experimentally determined data

was considered to be most reliable and is discussed in detail. The functions of the interacting partners are also documented in (Tables 6 and 7).

In bHLH domain-containing protein family, OsEAT1 was shown to interact with UDT1 and MYB80 (both known to play a role in tapetum development). OsUDT1 showed interaction with EAT1 [involved in the regulation of tapetum programmed cell death (PCD) and degradation during male reproductive development]. OsTDR1 showed interaction with Q0DSC3_ORYSJ, Q2RAZ2_ORYSJ (alpha-L-arabinofuranosidase C-terminus family protein), and A0A0P0W347. OsBHLH35 interacted with CARM1 (probable histone-arginine methyltransferase), OsI_09745 (GH16 domain-containing protein), and uncharacterized proteins like OsI_16462, OsI_01315, and OsI_06757 (Suppl. Table 6 and Fig. 9a).

STRING analysis of *MYB* family members showed the interaction of OsMYB188 with EAT1 and A0A0P0UZW0: Os01g0193900 proteins. OsAID1 with OsI_26549 (uncharacterized protein), OsI_27361 (ATP-dependent DNA helicase), OsI_35576 (3'-5' exonuclease domain-containing protein), OsI_04575 (DNA polymerase), and OsI_31334 (TCP domain-containing protein). OsTDF1 and OsCSA2 proteins interacted with OsI_00746 (SANT domain-containing protein), OsI_12754 (DIRP domain-containing protein), and OsI_12755 (uncharacterized protein). OsGAMYBL2 interacted with SANT domain-containing protein (A0A0E0FHJ3), DIRP domain-containing protein (A0A0E0GQX4), and Cyclin N-terminal domain-containing proteins (A0A0E0GY69 and A0A0E0JAP6). OsMoF1 with kinesin-like protein (KIN14I), serine/threonine-protein phosphatases (BSL1 and Q8RZZ3_ORYSJ), tetramerisation domain-containing protein (Q53NW5_ORYSJ), and A0A0P0V1T6: Os01g0314700 protein. OsGAMYB interacted with A0A0P0UZW0: Os01g0193900 protein (Suppl. Table 6 and Fig. 9a).

Further, in MADS-box family members OsMADS3 showed interaction with MADS6, MADS7, MADS8 and MADS29 proteins; OsMADS62 interacted with protein OsI_36907 (MADS-box domain-containing protein); OsMADS2, OsMADS4, OsMADS63, and OsMADS68 partnered with mitogen-activated protein kinases; OsMADS58 interacted with MADS1 (probable TF involved in the development of floral organs), MADS6 and MADS7 (regulates flowering time) and MADS8 (regulates floral organ identity,

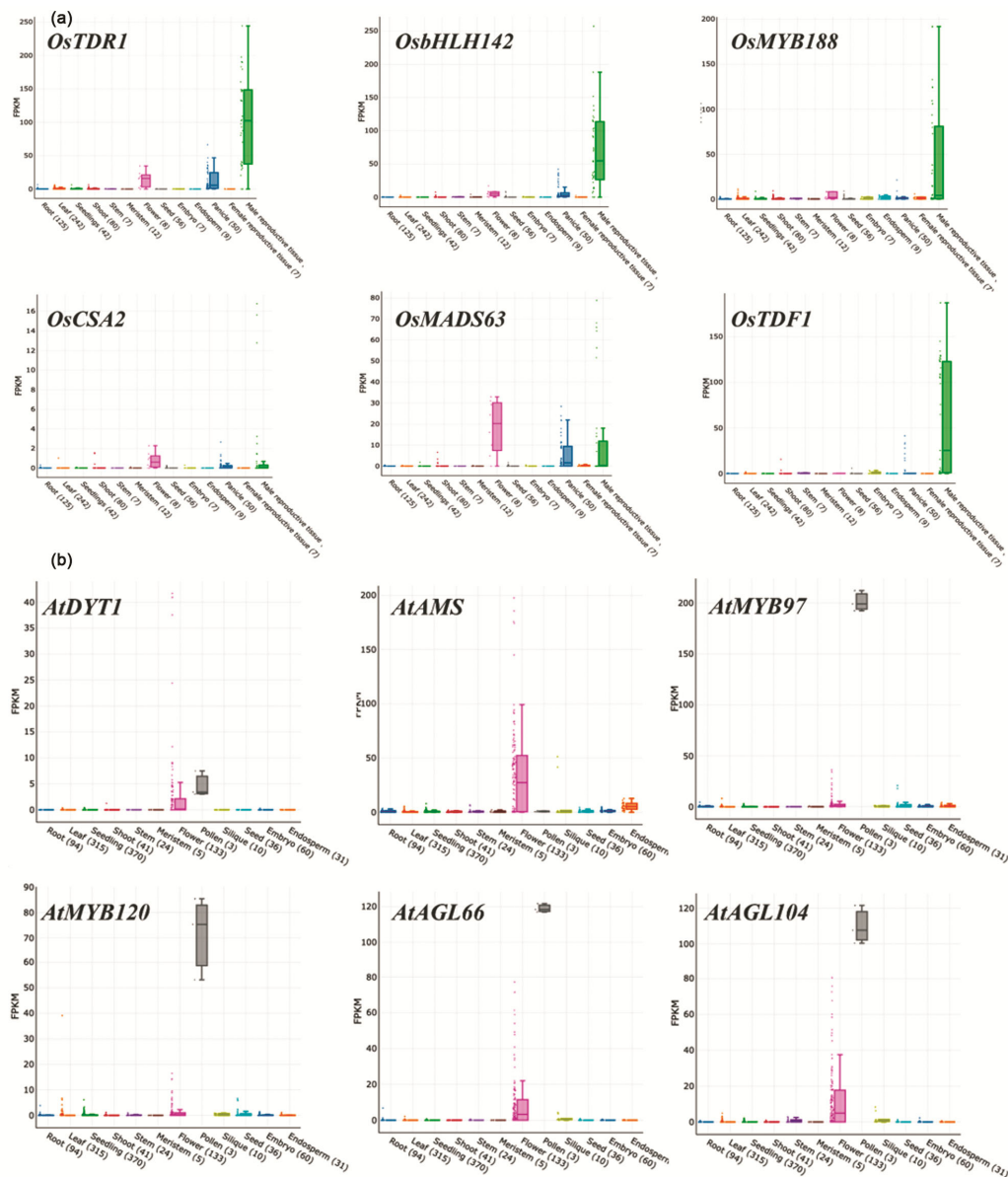


Fig. 8 — Expression profiles of (a) rice and (b) *Arabidopsis* anther-specific TFs using RNA-seq data. The data represents the expression levels of each TF in different tissues using a box and whisker plot. The names of the TFs are written at the top of each plot. The X axis represents the various tissues in which the transcripts have been detected, with the number of transcripts written in the brackets along with tissue name. The Y axis represents the expression level in terms of FPKM (Fragment per kilo base million) which indicates the number of reads mapped to one fragment in other words for paired-end RNA sequences. The figure illustrates transcript density in different tissues, distinguished by different colors

floral meristem determinacy and control of flowering time) proteins¹⁸. Interestingly, OsbHLH142 of the bHLH family and OsCSA from the MYB group did not show any experimentally derived interacting partners in the STRING database (Suppl. Table 6 and Fig. 9a).

In the case of *Arabidopsis*, members of the bHLH family AtbHLH089, AtbHLH091, AtbHLH010, and AtDYL1 (all involved in anther development) interacted with each other. AtbHLH089 and

AtbHLH091 interacted with DYL1. AtbHLH010 interacted with bHLH90 and DYL1. AtICE1 interacted with MUTE, SPCH, HOS1, FAMA, and TIFY6A (their respective roles elaborated in Suppl. Table 7). While TF AtAMS did not show any experimentally determined interacting partner (Suppl. Table 7 and Fig. 9b).

The STRING data analysis results for MYB family members of *Arabidopsis* showed interaction of

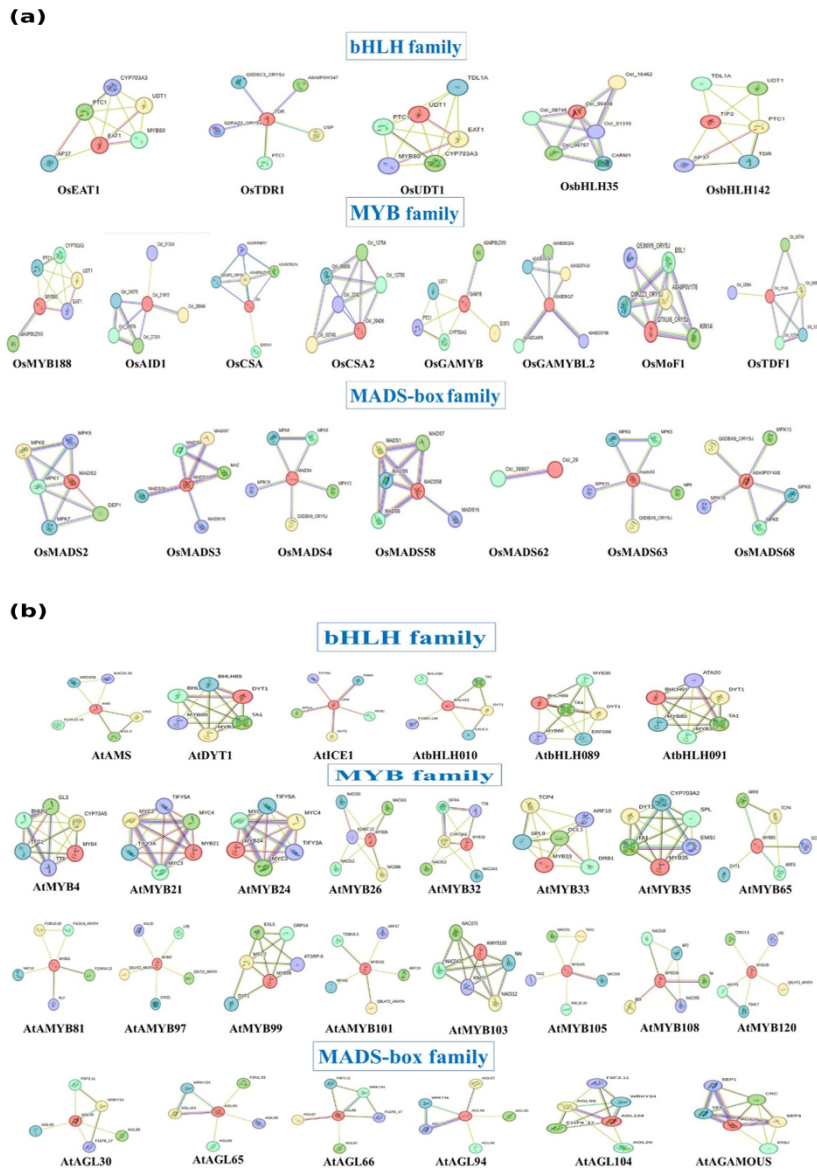


Fig. 9 — Protein-protein interaction network of (a) rice and (b) *Arabidopsis* anther-specific TFs generated using the STRING database. Nodes in the network denote proteins. Predicted interactions are represented by edges. The magenta line represents experimentally determined interaction, and the turquoise blue line represents interaction predicted using a curated database

AtMYB24 with MYC4, MYC3, MYC2, TIFY5A, and TIFY3A; AtMYB21 with MYC2, MYC3, MYC4, TIFY3A, and TIFY5A; AtMYB108 with BOI and NAC003; AtMYB105 with NAC098 (a transcriptional activator); and AtMYB4 with TT8 (another transcriptional activator) (Suppl. Table 7 and Fig. 9b).

Interestingly, a significant number of MYB class members, including AtMYB26, AtMYB32, AtMYB33, AtMYB35, AtMYB65, AtMYB81, AtMYB97, AtMYB99, AtMYB101, AtMYB103, and AtMYB120, did not show any experimentally determined interacting partners (Suppl. Table 7 and Fig. 9b).

In the MADS-box domain-containing protein family, AtAGL65 and AtAGL94 were seen interacting with AGL104 (crucial for regulation of pollen maturation at the late stages of pollen development and pollen tube growth). Further, AtAGL66 showed interaction with AGL82 (required for maintenance of the proper function of the central cell in pollen tube attraction), and AtAGAMOUS displayed interaction with SEP1, SEP2, and SEP3, all of which are required for normal anther development¹⁹. However, *AtAGL30* did not show any experimentally determined interacting partners (Suppl. Table 7 and Fig 9b.).

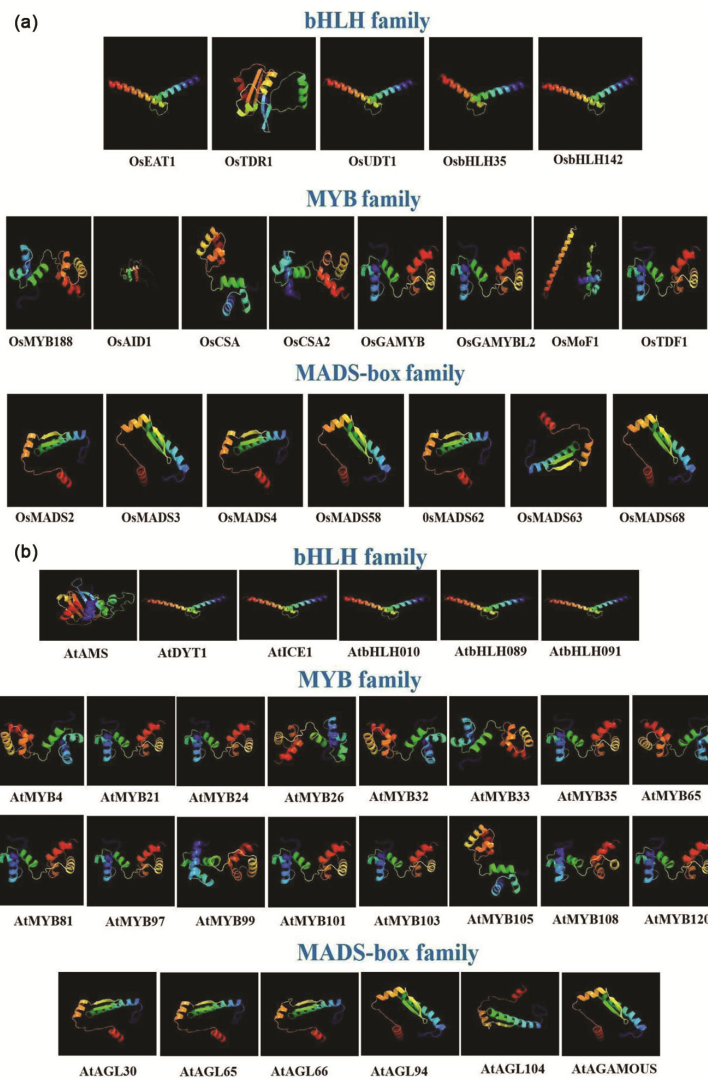


Fig. 10 — Protein structures of anther-specific TFs of (a) rice; and (b) *Arabidopsis* predicted with a 100% confidence level

Physical, chemical, and structural properties of proteins

The physical and chemical details of anther-specific proteins belonging to bHLH, MADS-box and MYB TF families were unfolded using the Prot Param expasy tool and are documented in (Suppl. Tables 8 and 9). Interestingly, proteins from both rice and *Arabidopsis* exhibited amino acid (a.a) lengths ranging from 200 to 700 (Suppl. Tables 1 and 2). In rice, the *OsMADS2* gene was found to encode the smallest protein with 210 a.a, while the *OsGAMYB* protein was the largest with 554 a.a (Suppl. Table 1). In *Arabidopsis*, it ranged between 207 a.a (*AtDYT1*) and 666 a.a (*AtAGL30*) (Suppl. Table 2). Further, all the proteins of rice were found unstable except *OsTDF1*, with a stability index of 38.24 (Suppl. Table 8). Likewise, in *Arabidopsis*, except for one protein, *AtAMS* with a stability index of 38.78, all the

other proteins were found unstable with an instability index of more than 40 (Suppl. Table 9).

The structural details of proteins were analysed using the Phyre 2 tool. In this analysis, it was observed that within the bHLH family, all the proteins were structurally similar except *OsTDR1*. In the MYB family, *OsGAMYB*, *OsGAMYBL2* and *OsTDF1* proteins exhibited similar structures while the others displayed distinct structural features. Similarly, among the MADS-box proteins, *OsMADS2*, *OsMADS4*, and *OsMADS62* shared structural similarity. *OsMADS3*, *OsMADS58*, and *OsMADS68* formed another structurally similar group. However, *OsMADS63* displayed a unique structure (Fig. 10a).

Similarly, in *Arabidopsis*, the MYB proteins exhibited the highest degree of structural variations. *AtMYB4* showed similarity to *AtMYB32*, while, the

TFs AtMYB21, AtMYB24, AtMYB35, AtMYB81, AtMYB97, AtMYB101, AtMYB103, AtMYB108, and AtMYB120 displayed similar structures. In the bHLH group, despite differences in the molecular weights, the structures of AtDYT1, AtICE1, AtbHLH010, AtbHLH089 and AtbHLH091 appeared similar. In the MADS-box family, *AtAGL30* showed similarity to *AtAGL65*, while *AtAGL94* and *AtAGAMOUS* were structurally alike. However, *AtAGL66* and *AtAGL104*, did not show similarity to any other protein (Fig. 10b).

On comparing rice and *Arabidopsis*, in the bHLH family, OsEAT1, OsUDT1, OsbHLH35, OsbHLH142, AtDYT1, AtICE1, AtbHLH010, AtbHLH089 and AtbHLH091 displayed similar structures. Similarly, in the MYB family, OsGAMYB, OsGAMYBL2, OsTDF1, AtMYB21, AtMYB24, AtMYB35, AtMYB81, AtMYB97, AtMYB101, AtMYB103, AtMYB108 and AtMYB120 showed structural similarity. Likewise in the MADS-box family OsMADS2, OsMADS4, OsMADS62, *AtAGL30* and *AtAGL65* showed similar structures, while OsMADS3, OsMADS58, OsMADS68, *AtAGL94* and *AtAGAMOUS* formed another structurally similar group.

Discussion

Anther development, is a highly intricate process involving a regulatory network of various TFs that operates in a spatial and temporal manner. The previous scientific reports have categorized the process of anther development in *Arabidopsis* into 14 and 15 stages^{20,21} and in rice, this process is divided into 9, 8, and 14 stages²²⁻²⁴. The role of TFs is evident from the early stages of anther development, known as the pre-meiotic phase, to the tetrad/meiosis stage, and culminating in the final anther dehiscence stage^{7,25-29}.

TFs serve as master regulators, participating in numerous plant growth and developmental processes, as well as stress responses. By utilizing anther-specific TFs, male sterility can be engineered in crops for hybrid seed production. Many studies have covered the role of TFs in anther development and their potential application in the generation of male sterile lines³⁰⁻³². One such example is AtMYB80, an anther-specific TF that was employed to inhibit and restore fertility in reversible male sterility system³³. This targeted breeding strategy can speed up the development of improved crop varieties suited to ever-changing environmental conditions.

Significant progress has recently been made in identifying important candidate genes associated with

anther development, largely due to the availability of numerous male sterile mutants in the model plant *Arabidopsis*. Additionally, anther development has also been extensively studied in rice (a monocot model system). However, research on the other economically important crops, such as barley, wheat, and maize, remains limited. Inducing male sterility in these crops requires the identification of orthologous genes involved in anther development. Since the regulatory networks governing different phases of anther development are conserved across species, this knowledge can be leveraged to regulate fertility in economically important crops like wheat and maize. In this study, the TFs belonging to three major protein families named bHLH, MYB, and MADS-box in rice and *Arabidopsis* plants were documented. Despite the larger genome size of rice, a greater number of anther-specific TFs have been characterized in *Arabidopsis* (20 in rice and 28 in *Arabidopsis*). This could be due to the fact that more studies have focused on *Arabidopsis* than on rice, opening up potential opportunities for future research in the rice crop. Anther-specific TF genes of rice and *Arabidopsis*, along with their tissue of expression, heterologous expression, overexpression, and mutant phenotypes, have been documented in (Suppl. Tables 10 and 11).

The chromosomal study showed distribution of 48 anther-specific genes across 12 chromosomes in rice and 5 chromosomes in *Arabidopsis*. The maximum number of genes was positioned at chromosome number 1 in both rice and *Arabidopsis*. While all of the 5 chromosomes of *Arabidopsis* exhibited gene positioning, chromosome numbers 9, 10, and 12 of rice lacked the anther-specific TF genes. In both rice and *Arabidopsis*, the maximum numbers of anther-specific TFs (8 in rice and 16 in *Arabidopsis*) belonged to the MYB TF family, making this family the most diverse in both plants. NCBI, SMART, InterPro, and Pfam domain search analysis displayed the presence of bHLH, MYC1_like, MADS, K-box, MEF2-like SRF-type, Myb, and SANT domains, which are found common to both rice and *Arabidopsis*. This indicated the conservation and functional significance of these domains in both species. In *Arabidopsis*, an additional PLN03212 superfamily domain was also reported in the MYB class. The presence of these domains validated the classification of these proteins into their respective families and also provided potential usefulness in determining their functional details.

Moreover, different features of rice and *Arabidopsis* genes, including their ORF and protein length, spliced variants, molecular weights, CDS coordinates, and their location, were also documented in (Suppl. Tables 1 and 2). The protein with the largest size was reported to be OsGAMYB (554 a.a) of the MYB class in rice and *AtAGL30* of the MADS-box family (666 a.a) in *Arabidopsis*. Gene distribution analysis revealed that anther-specific TFs in *Arabidopsis* and rice are unevenly distributed across chromosomes. Motif conservation is indicative of functional conservation, as evidenced by their role in anther development in both species, further supported by expression and promoter analysis.

The analysis of gene structure provides insights into evolutionary processes such as duplication events and divergence³⁴. Comparative gene structure analysis helps in understanding both the evolutionary and functional relationships. In the present study, anther-specific TFs exhibited distinct exon-intron patterns across different genes, despite all being anther-specific. The variation may be attributed to differences in exon and intron lengths. *OsEAT1* of rice and its orthologous counterparts in *Arabidopsis*, *AtbHLH089*, *AtbHLH091*, and *AtbHLH010*, possessed the same numbers of introns (2) and UTRs (presence of both 5' and 3' UTR in all of these proteins). Similarly, another orthologous pair, *OsGAMYB* of rice and *AtMYB65* of *Arabidopsis*, also possessed 2 introns each and displayed both 5'UTR and 3'UTR, suggesting the possibility of functional similarities among them. These similar intron numbers and the presence of both 5' and 3' UTR suggest functional conservation of gene function in both species over the course of evolution. However, the remaining 4 orthologous gene pairs displayed differences in terms of both UTRs and the number of introns, thus implying diversification, functional differences, and loss of introns during the course of evolution.

Subcellular localization studies (an important characteristic for determining the function of proteins)³⁵ predicted that the majority of the anther-specific TF proteins are localized in the nucleus, which supports their role as key regulators of the transcriptional process. Phylogenetic analysis grouped 48 anther-specific genes into 3 clades and their respective subclades, with clade 3 being the largest with 28 proteins. Notably, subclades 1A, 1B, 2A, 3B2a, and 3B2b contained members from both rice and *Arabidopsis*, indicating a significant degree of relatedness between both these plants and implying

conservation of genes during the evolutionary process.

The URR of a gene contains various CREs that are specific to different proteins involved in the initiation and regulation of transcription³⁶⁻³⁸. *Cis*-regulatory elements play an important role in operating and assisting the process of transcription as they possess TF binding sites. Upstream regulatory regions of all the anther-specific TF genes were analyzed for the presence of three CREs, known for their role in the regulation of anther-specific gene expression³⁹. These CREs include S000245/POLLEN1LELAT52, possessing motif AGAAA; S000378/GTGANTG10/pollen with motif GTGA; and Q (quantitative)-element with motif AGGTCA. It has been found that all of these elements are involved in processes related to pollen development. For instance, Q-element increases the pollen-specific expression of the *Zm13* promoter⁴⁰. Another report showed that the GTGANTG10 motif in the promoter of tobacco *g10* plays a significant role in regulating the pollen-specific expression of the gene⁴¹. POLLEN1LELAT52 was observed to drive pollen-specific induction of tomato *LAT52* gene⁴². The maximum number of motif sites was identified in the URR of *OsTDR1* gene, a bHLH protein family member in rice, while in *Arabidopsis*, the *AtMYB101* gene URR (a MYB family protein) exhibited the highest motif sites. This highlights the involvement of these genes in anther development in both rice and *Arabidopsis* as well as their functional dominance, since the abundance of CREs can be directly correlated with the intensity of a gene's expression level⁴³.

Gene ontology enrichment analysis of 48 identified TF genes demonstrated that the two biological processes pertaining to the highest ranks in rice were stamen development and androecium development. The appearance of these terms at the top positions shows that genes taken into consideration for this particular study are upregulated and are crucial for anther development, thus validating our search for anther-specific genes. Interestingly, the top ranked biological pathway in *Arabidopsis* was observed to be MYB-like DNA-binding domain region, which aligns with our study, as the MYB TF is the most diverse among all the anther-specific genes identified.

Expression heat maps were generated using raw expression data downloaded from NCBI GEO datasets to examine the functional relevance of these 48 TFs in rice and *Arabidopsis*. As expected, the highest expression levels were observed in

reproductive tissues. In rice, the expression was predominant across different stages of anther development examined, while in *Arabidopsis*, TFs were majorly expressed in other reproductive tissues like flower pedicels and carpels. This data was validated using RNA-seq data which also showed highest transcript levels in flower and male reproductive parts of rice and *Arabidopsis*. These findings not only confirm their functional significance in anther development process but also help to identify the key tissues that can be targeted for generating hybrid crop varieties.

One potential way of understanding the functions of proteins is to gain an insight into their interacting partners, for which the STRING database was employed, and the interacting partners of both rice and *Arabidopsis* anther-specific proteins were unveiled using experimentally determined data. Rice anther-specific TFs were found to interact with proteins that have a role in anther specification, anther development, tapetum development and its PCD, flowering, floral organ formation, and floral organ identity determination, which suggested that different pathways with different proteins are required for anther development. For example, OsEAT1 was found to interact with OsUDT1 and OsMYB80 (both are known to play important roles in tapetum development). OsMADS58 interacted with MADS1, MADS6, MADS7, and MADS8 (crucial for stamen development) proteins¹⁸. It is also known that anther development is sensitive to abiotic stress⁴⁴. Interestingly, some of the TF proteins (MADS-box family) interacted with enzymes such as mitogen-activated kinases, which are involved in conferring stress tolerance. *Arabidopsis* TFs mainly interacted with proteins having potential roles in pollen tube attraction, stamen development, tapetum development, pollen differentiation, post-meiotic pollen development and maturation, and male sterility that supported the roles of these proteins in anther development.

Understanding the molecular processes involved in anther development is an effective approach in regulating male reproduction in plants and has proven beneficial for generating hybrid varieties. Hybrid rice varieties generally produce higher yields than conventionally produced rice varieties. The present study facilitates the selection and use of anther-specific TFs from different families (bHLH, MADS-box, and MYB) as candidate genes for various breeding and transgenic-based approaches to develop improved varieties of rice and other crop plants.

Conclusion

This study aimed to analyse TFs involved in anther development and to gain deeper insights into their regulatory mechanisms through a comparative approach. A total of 48 anther-specific proteins belonging to three major classes of TFs, bHLH, MYB, and MADS-box were identified in rice and *Arabidopsis*. Among these six orthologous TF pairs were identified between these two crop plants suggesting a conserved molecular framework for anther development across monocots and dicots. Gene ontology and expression profiling studies further revealed that these proteins are primarily expressed in the reproductive tissues mainly in stages critical for male fertility. Additionally, protein-protein interaction and upstream regulatory region analysis showed that most anther-specific proteins in rice and *Arabidopsis* interact with proteins involved in stamen and tapetum development and form a tight regulatory network. These findings provide important insights into the regulatory network of TFs governing anther development. Thus, unravelling the TFs and their roles in anther development will help further in understanding the mechanisms controlling male sterility, which can be effectively employed in hybrid seed production. Moreover, the identification and genetic manipulation of candidate anther-specific genes could contribute to the development of improved rice varieties, providing an efficient and sustainable solution to the global food crisis.

Acknowledgement

The authors acknowledge the support provided by Gargi College, University of Delhi, Delhi, which was instrumental in facilitating the successful completion of this research work.

Conflict of interest

All authors declare no conflict of interest.

References

1. Blanc-Mathieu R, Dumas R, Turchi L, Lucas J & Parcy F, Plant-TFClass: a structural classification for plant transcription factors. *Trends Plant Sci*, 29 (2024) 40.
2. Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P & Solano R, DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci U S A*, 116 (2014) 2367.
3. Strader L, Weijers D & Wagner D, Plant transcription factors – being in the right place with the right company. *Curr Opin Plant Biol*, 65 (2022) 102136.
4. Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman, Pilgrim M, Broun P, Zhang JZ, Ghandehari D,

- Sherman BK & Yu G, *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, 290 (2000) 2105.
- 5 Jones S, An overview of the basic helix-loop-helix proteins. *Genome Biol*, 5 (2004) 226.
 - 6 Nair SK & Burley SK, Recognizing DNA in the library. *Nature*, 404 (2000) 715.
 - 7 Jung KH, Han MJ, Lee YS, Kim YW, Hwang I, Kim MJ, Kim YK, Nahm BH & An G, Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. *Plant Cell*, 17 (2005) 2705.
 - 8 Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang Z & Zhang D, Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and *Arabidopsis*. *Plant Physiol*, 141 (2006) 1167.
 - 9 Niu N, Liang W, Yang X, Jin W, Wilson ZA, Hu J & Zhang D, EAT1 promotes tapetal cell death by regulating aspartic proteases during male reproductive development in rice. *Nat Commun*, 4 (2013) 1445.
 - 10 Fu Z, Yu J, Cheng X, Zong X, Xu J, Chen M, Li Z, Zhang D & Lian W, The rice basic helix-loop-helix transcription factor TDR INTERACTING PROTEIN2 is a central switch in early anther development. *Plant Cell*, 26 (2014) 1512.
 - 11 Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, & Robertson DL, Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, poplar, rice, moss, and algae. *Plant Physiol*, 153 (2010) 1398.
 - 12 Tilly JJ, Allen DW & Jack T, The CArG boxes in the promoter of the *Arabidopsis* floral organ identity gene APETALA3 mediate diverse regulatory effects. *Development*, 12 (1998) 1647.
 - 13 Parenicová L, De Folter S, Kieffer M, Hornzer DS, Favalli C, Busscher J, Holly HE, Ingram RM, Kater MM, Davies B, Angenent GC & Colombo L, Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell*, 15 (2003) 1538.
 - 14 Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK & Kapoor S, MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genom*, 8 (2007) 1.
 - 15 Zhang T, Zhao Y, Wang Y, Liu Z & Gao C, Comprehensive analysis of MYB gene family and their expressions under abiotic stresses and hormone treatments in *Tamarix hispida*. *Front Plant Sci*, 9 (2018) 1303.
 - 16 Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V & Bansal KC, Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC Genom*, 13 (2012) 1.
 - 17 Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ & Mering CV, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47 (2019) D607.
 - 18 Li H, Liang W, Hu Y, Zhu L, Yin C, Xu J, Dreni L, Kater MM & Zhang D, Rice MADS6 interacts with the floral homeotic genes SUPERWOMAN1, MADS3, MADS58, MADS13, and DROOPING LEAF in specifying floral organ identities and meristem fate. *Plant Cell*, 23 (2011) 2536.
 - 19 Castillejo C, Romera-Branchat M & Pelaz S, A new role of the *Arabidopsis* SEPALLATA3 gene revealed by its constitutive expression. *Plant J*, 43 (2005) 586.
 - 20 Sanders PM, Bui AQ, Weterings K, McIntire KN, Hsu YC, Lee PY, Truong MT, Beals TP & Goldberg RB, Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sex Plant Reprod*, 11 (1999) 297.
 - 21 Gómez JF, Talle B & Wilson ZA, Anther and pollen development: a conserved developmental pathway. *J Integr Plant Biol*, 57 (2015) 876.
 - 22 Raghavan V, Anther and pollen development in rice (*Oryza sativa*). *Am J Bot*, 75 (1988) 183.
 - 23 Zhang D & Wilson ZA, Stamen specification and anther development in rice. *Chin Sci Bull*, 54 (2009) 2342.
 - 24 Zhang D, Luo X & Zhu L, Cytological analysis and genetic control of rice anther development. *J Genet Genom*, 38 (2011) 379.
 - 25 Zhu QH, Ramm K, Shivakkumar R, Dennis ES & Upadhyaya NM, The ANOTHER INDEHISCENCE1 gene encoding a single MYB domain protein is involved in anther development in rice. *Plant Physiol*, 135 (2004) 1514.
 - 26 Hu L, Liang W, Yin C, Cui X, Zong J, Wang X, Hu Z & Zhang D, Rice MADS3 regulates ROS homeostasis during late anther development. *Plant Cell*, 23 (2011) 515.
 - 27 Ortolan F, Trenz TS, Delaix CL, Lazzarotto F & Margis-Pinheiro M, bHLH-regulated routes in anther development in rice and *Arabidopsis*. *Genet Mol Biol*, 46 (2023) e20230171.
 - 28 Qian Z, Shi D, Zhang H, Li Z, Huang L, Yan X & Lin S, Transcription factors and their regulatory roles in the male gametophyte development of flowering plants. *Int J Mol Sci*, 25 (2024) 566.
 - 29 Miglani N, Agarwal P, Ranjan R, Mohanty A & Mishra R, Transcription Factors: Major Regulators of Anther Development in Rice. *J Plant Growth Regul*, 44 (2025) 4227.
 - 30 Zhang H, Xu C, He Y, Zong J, Yang X, Si H, Sun Z, Hu J, Liang W & Zhang D, Mutation in CSA creates a new photoperiod-sensitive genic male sterile line applicable for hybrid rice seed production. *Proc Natl Acad Sci U S A*, 110 (2012) 76.
 - 31 Ko SS, Li MJ, Lin YJ, Hsing HX, Yang TT, Chen TK, Jhong CM & Ku MS, Tightly controlled expression of bHLH142 is essential for timely tapetal programmed cell death and pollen development in rice. *Front Plant Sci*, 8 (2017) 1258.
 - 32 Xiang XJ, Sun LP, Yu P, Yang ZF, Zhang PP, Zhang YX, Wu WX, Chen DB, Zhan XD, Khan RM, Abbas A, Cheng SH & Cao LY, The MYB transcription factor Baymax1 plays a critical role in rice male fertility. *Theor Appl Genet*, 134 (2021) 453.
 - 33 Li SF, Iacuone S & Parish RW, Suppression and restoration of male fertility using a transcription factor. *Plant Biotechnol J*, 5 (2007) 297.
 - 34 Yang WY, Novembre J, Eskin E & Halperin E, A model-based approach for analysis of spatial structure in genetic data. *Nat Genet*, 44 (2012) 725.
 - 35 Sharma R, Kapoor M, Tyagi AK & Kapoor S, Comparative transcript profiling of TCP family genes provide insight into gene functions and diversification in rice and *Arabidopsis*. *J Plant Mol Biol Biotechnol*, 1 (2010) 24.
 - 36 Ong CT & Corces VG, Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*, 12 (2011) 283.

- 37 Ho CL & Geisler M, Genome-wide computational identification of biologically significant cis-regulatory elements and associated transcription factors from rice. *Plants*, 8 (2019) 441.
- 38 Mulat MW & Sinha VB, Distribution and abundance of CREs in the promoters depicts crosstalk by WRKYs in Tef [*Eragrostis tef* (Zucc.) Trotter]. *Gene Rep*, 23 (2021) 101043.
- 39 Hernandez-Garcia CM & Finer JJ, Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci*, 217 (2014) 109.
- 40 Hamilton DA, Schwarz YH & Mascarenhas JP, A monocot pollen-specific promoter contains separable pollen-specific and quantitative elements. *Plant Mol Biol*, 38 (1998) 663.
- 41 Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, Lonsdale DM & Twell D, Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. *Plant Mol Biol*, 45 (2001) 577.
- 42 Filichkin SA, Leonard JM, Monteros A, Liu PP & Nonogaki H, A novel endo- β -mannanase gene in tomato LeMAN5 is associated with anther and pollen development. *Plant Physiol*, 134 (2004) 1080.
- 43 Worsley-Hunt R, Bernard V & Wasserman WW, Identification of cis-regulatory sequence variations in individual genome sequences. *Genome Med*, 3 (2011) 65.
- 44 Zhang Z, Hu M, Xu W, Wang Y, Huang K, Zhang C & Wen J, Understanding the molecular mechanism of anther development under abiotic stresses. *Plant Mol Biol*, 105 (2021) 1.