Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide landscape of m6A regulatory genes in Marchantia polymorpha: conservation, diversification, and expression dynamics

Joong-Tak Yoon
Duke Kunshan University
Shaban Muhammad
Duke Kunshan University
Yicheng Liu
Duke Kunshan University
Eun Yu Kim
eunyu.kim@dukekunshan.edu.cn

Duke Kunshan University

Research Article

Keywords: Marchantia polymorpha, N6-methyladenosine, m6A, RNA methylation, Genome-wide identification, Comparative genomics, Gene expression, Abiotic stress, cis-acting elements

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11429316/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
No competing interests reported.

Tables are available in the Supplementary Files section.

Genome-wide landscape of m^{6}A regulatory genes in *Marchantia polymorpha*: conservation, diversification, and expression dynamics

Joong-Tak Yoon[1,†,*], Shaban Muhammad[1,†], Yicheng Liu[1], Eun Yu Kim[1,2,3,*]

[1] Division of Natural and Applied Sciences, Duke Kunshan University, Kunshan, Jiangsu 215316, China

[2] Environmental Research Center, Duke Kunshan University, Kunshan, Jiangsu 215300, China.

[3] Center for Genome Engineering, Institute for Basic Science, Daejeon 34126, Republic of Korea

†These authors contributed equally to this work

[*]Correspondence:

Joong-Tak Yoon (joongtak.yoon@dukekunshan.edu.cn)

Eun Yu Kim (eunyu.kim@dukekunshan.edu.cn)

Abstract

Background

N^6-Methyladenosine (m^6A) regulates RNA fate in flowering plants, but the repertoire and expression patterns of m^6A-related genes remain incompletely characterized in liverworts. Here, we characterized these gene families in the liverwort *Marchantia polymorpha* within a comparative framework spanning diverse plant lineages.

Results

A comparative survey of 122 plant genomes using sequence similarity and conserved-domain criteria identified 18 candidate m^6A-related loci in *M. polymorpha*, comprising seven writer-related genes, five AlkB homolog genes, and six reader-related genes. Across the surveyed genomes, canonical writer components generally had low copy numbers, whereas AlkB homolog and reader-related families showed greater copy-number variation. Phylogenetic and domain analyses supported the family assignments of the *M. polymorpha* candidates. Eight of the 18 candidate loci showed detectable collinearity with *Marchantia quadrata*. Analysis of putative promoter regions identified motifs annotated as hormone-, light-, and stress-responsive, with hormone-responsive motifs forming the largest category among functionally classified sites. Public transcriptome analysis revealed variation in candidate gene expression across developmental stages and tissues. Quantitative reverse-transcription PCR showed significantly higher transcript abundance for four of the six reader-related genes in thalli than in both gemmae and gemma cups. Cold, heat, osmotic, salt, and dark treatments elicited gene-specific expression responses. In the qRT-PCR assays, darkness increased transcript abundance of several candidates, including *MpMTA*, *MpALKBH1A*, and *MpECT2*, relative to unstressed controls. A predicted protein-association network suggested links between subsets of the candidates and RNA-

processing proteins.

Conclusions

These findings define a candidate m^6A-related gene repertoire in *M. polymorpha* and identify tissue- and stress-responsive genes for targeted functional studies of RNA methylation in this liverwort.

Keywords

Marchantia polymorpha, N^6-methyladenosine, m^6A, RNA methylation, Genome-wide identification, Comparative genomics, Gene expression, Abiotic stress, *cis*-acting elements

Background

N^6-Methyladenosine (m^6A) is a widespread internal RNA modification that adds a post-transcriptional layer to gene regulation [1, 2]. On plant transcripts, m^6A has been associated with RNA processing, alternative polyadenylation, transcript stability, and translation [3-6]. Its effects depend on the modified RNA and the proteins that install, remove, or recognize the methyl group [7-9]. These proteins are commonly classified as writers, erasers, and readers, respectively (Figure 1A) [7-9]. Studies in flowering plants have linked this regulatory framework to embryogenesis, organ development, and responses to environmental signals, but the functions of individual family members can differ across biological contexts [1, 8, 10, 11].

In *Arabidopsis thaliana*, METHYLTRANSFERASE A (MTA) and METHYLTRANSFERASE B (MTB) are central components of the mRNA m^6A writer complex, together with FKBP12 INTERACTING PROTEIN 37 KD (FIP37), VIRILIZER (VIR), and HAKAI [1, 8]. Depletion of several of these components lowers mRNA methylation and can disrupt embryogenesis, shoot meristem activity, and root development [1, 8, 10]. The effects of the writer machinery also extend beyond protein-coding mRNAs [12]. MTA-dependent methylation of primary microRNAs (pri-miRNAs) promotes their efficient processing by the nuclear microRNA machinery, which includes DICER-LIKE 1 (DCL1) and HYPONASTIC LEAVES 1 (HYL1) (Figure 1B) [12]. FIONA1 (FIO1), a methyltransferase related to METHYLTRANSFERASE-LIKE 16 (METTL16), represents a distinct route to RNA methylation: its modification of U6 small nuclear RNA (snRNA) affects splice-site selection and pre-mRNA splicing [4]. Thus, the presence of a methyltransferase-related domain does not necessarily imply membership in the MTA–MTB writer complex or the same RNA substrate specificity [4, 8].

Removal of m^6A adds another potential level of regulation [7, 13]. The *A. thaliana* AlkB homolog

(ALKBH) protein ALKBH10B demethylates transcripts encoding flowering regulators, including FLOWERING LOCUS T (FT), SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3 (SPL3), and SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 9 (SPL9), and thereby affects their stability and the timing of floral transition [7]. A different family member, ALKBH9B, influences the m^6A content of alfalfa mosaic virus RNA and viral infection [13]. These examples show that plant ALKBH proteins can act in different cellular and biological contexts (Figure 1A) [7, 13]. However, RNA demethylase activity and substrate choice have been demonstrated for only a subset of plant ALKBH proteins; they cannot be inferred for every family member from sequence similarity alone [7, 13, 14].

The consequences of m^6A are also determined by reader proteins [5, 9]. In *A. thaliana*, the long isoform of CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR 30 (CPSF30-L) contains an m^6A-recognizing YT521-B homology (YTH) domain and contributes to alternative polyadenylation of target transcripts in the nucleus, with reported effects on nitrate signaling, floral transition, and abscisic acid (ABA) responses [3, 9]. In the cytoplasm, EVOLUTIONARILY CONSERVED C-TERMINAL REGION (ECT) proteins constitute another group of YTH-domain readers [5, 15]. *A. thaliana* ECT2, ECT3, and ECT4 support organogenesis and can promote the stability of bound transcripts; they have also been implicated in translation of immune-related mRNAs under specific conditions [5, 6, 15, 16]. ECT8 provides a distinct stress-associated example: it sequesters m^6A-modified PYRABACTIN RESISTANCE 1-LIKE 7 (PYL7) mRNA in stress granules and limits its translation during ABA responses [17], whereas its association with processing bodies promotes target mRNA decay during salt stress [11]. Accordingly, reader-dependent outcomes range from RNA persistence to selective translation or turnover and should be evaluated in the relevant cellular context (Figure 1B) [5, 6, 9, 11, 17].

Genome-wide and transcriptome-scale studies in *A. thaliana*, *Oryza sativa*, *Litchi chinensis*, *Pinus massoniana*, poplar, and *Brassica napus* have catalogued m^{6}A-related gene families and described variation in copy number, promoter motifs, and developmental or stress-associated expression [18-23]. A broader comparative analysis across 154 plant species examined the evolutionary relationships, family expansions, and conserved genomic contexts of m^{6}A-related proteins [14]. The liverwort *Marchantia polymorpha* belongs to an early-diverging lineage of land plants, providing an evolutionary comparison with flowering plants and a model for examining gene expression across vegetative, asexual, and sexual developmental contexts [24]. Homologs of the core writer-complex components MTA, MTB, FIP37, VIR, and HAKAI have already been identified in *M. polymorpha* [25]. In addition, a *M. polymorpha* YTH-domain family (YTHDF) protein partially complemented leaf-development defects in *A. thaliana ect2 ect3 ect4* mutants in a heterologous assay [26]. These findings establish a basis for investigating m^{6}A-related proteins in *M. polymorpha*, but do not resolve how the broader candidate repertoire is expressed across developmental contexts and abiotic stress conditions. Integrating comparative sequence analyses with expression profiling can help identify candidates for functional investigation.

Here, we surveyed candidate m^{6}A-related proteins across 122 plant genomes and identified 18 candidate loci in *M. polymorpha*: seven writer-related genes, five ALKBH-family genes, and six reader-related genes (Figure 1C). We analyzed phylogenetic relationships, conserved domains, gene structures, predicted protein structures, chromosomal collinearity with *Marchantia quadrata*, putative promoter motifs, expression across developmental contexts and abiotic stresses, and predicted protein associations. This integrated analysis defines a candidate repertoire and identifies specific genes for future tests of RNA methylation and function in a liverwort.

Methods

Plant materials and growth conditions

The male *Marchantia polymorpha* subsp. *ruderalis* accession Takaragaike-1 (Tak-1), previously identified by Prof. Masaki Shimamura [27], was propagated on half-strength Gamborg's B5 medium with vitamins (Duchefa Biochemie, G0210) supplemented with 1% (w/v) sucrose (Adamas life, 66841U) and 1.4% (w/v) phyto agar (Duchefa Biochemie, P1003), adjusted to pH 5.5. Plants were grown at 22 °C under continuous LED illumination at a photosynthetic photon flux density of 100 μmol m^{-2} s^{-1}. The Tak-1 stock used in this study was originally obtained from the laboratory of Prof. Masaki Shimamura and subsequently maintained at Yonsei University, Republic of Korea. No wild plants were collected, and no voucher specimen was deposited as part of this study.

Abiotic stress treatments

Abiotic stress treatments were performed as previously described, with modifications [28]. For salt and osmotic stress treatments, gemmae were grown for 15 days on sucrose-free half-strength Gamborg's B5 medium with vitamins and 1.4% (w/v) phyto agar, supplemented with 40 mM NaCl (Macklin, S805286) or 100 mM mannitol (Greagent, G77529B), respectively. Both media were adjusted to pH 5.5. For cold, heat, and dark stress treatments, gemmae were initially grown on the same sucrose-free basal medium without NaCl or mannitol. On day 14, plates were transferred to 3 °C or 33 °C for 24 h for cold and heat treatments, respectively. For dark treatment, plates were wrapped in aluminum foil on day 12 and maintained in darkness for 3 days. Corresponding control plants were grown on the same sucrose-free basal medium without stress treatment at 22 °C under continuous LED light until day 15. On day 15, control and stress-treated plants were harvested and

used for further analysis.

Genome-wide identification of m^{6}A writers, erasers, and readers in 122 plant genomes

Predicted protein sequences for 119 plant species were obtained from Ensembl Plants release 62 [29]. *A. thaliana* Araport11 protein sequences were obtained from TAIR [30]. The full protein dataset (Araport11_pep_20250411.gz) was included among the target proteomes, and reference query sequences were selected from the representative-gene-model dataset (Araport11_pep_20250411_representative_gene_model.gz). The *M. quadrata* var. MQ v1.1 primary-transcript protein sequences were obtained from JGI Phytozome [31, 32]. The *M. polymorpha* MpTak_v7.1 assembly, annotation, and predicted protein sequences were obtained from MarpolBase, bringing the dataset to 122 plant genomes [33]. Previously reported *A. thaliana* m^{6}A-related proteins were used as reference queries [18, 34]. The query set comprised writer-related proteins, ALKBH-family proteins, and reader-related proteins. Candidates assigned to ALKBH8 or ALKBH8A were excluded after domain-compatible query assignment because the corresponding *A. thaliana* reference proteins have characterized roles in tRNA wobble-uridine modification: AtALKBH8 catalyzes hydroxylation of mcm^{5}U to (S)-mchm^{5}U, whereas AtALKBH8A (AtTRM9) catalyzes methylation of cm^{5}U to mcm^{5}U [35]. The names and locus identifiers of the reference queries are provided in Supplementary Table S1 (Additional file 1).

Predicted protein sequences from the 122 species were combined into a single FASTA file, with sequence identifiers linked to their species of origin and original protein identifiers. Only protein sequences beginning with methionine were retained. A local protein database was constructed from the combined sequences using makeblastdb (NCBI BLAST+ v2.17.0+). BLASTp v2.17.0+ was used to search this database with the *A. thaliana* reference sequences at an E-value cutoff of 1 ×

185 10^{-5} [36]. Hits with less than 30% amino acid identity across the aligned region were excluded.
186 Proteins passing these criteria were extracted for conserved-domain analysis.
187 Candidate proteins were searched against the NCBI Conserved Domain Database (CDD v3.21)
188 using RPS-BLAST v2.17.0+ at an E-value cutoff of 1×10^{-2} [36, 37]. The diagnostic domain
189 models examined were MT-A70 (CDD: 368270, pfam05063), Wtap (CDD: 465345, pfam17098),
190 VIR_N (CDD: 464938, pfam15912), Methyltransf_10, MTase-10 (CDD: 399160, pfam05971),
191 RING-HC_HAKAI-like (CDD: 438171, cd16508), 2OG-FeII_Oxy_2, ALKBH (CDD: 433285,
192 pfam13532), YTH (CDD: 410979, cd21134, CDD: 461195, pfam04146), Cleavage and
193 polyadenylation specificity factor, CPSF (CDD: 227416, COG5084), and RRM_ist3_like, RRM
194 (CDD: 409845, cd12411). Candidates were retained only when they shared at least one selected
195 CDD model with a reference query and met the corresponding BLAST criteria. For assignments
196 involving the RRM_ist3_like model, an additional BLAST E-value threshold of $\leq 1 \times 10^{-40}$ was
197 applied before query assignment. The remaining proteins were classified as writer-, eraser-, or
198 reader-related candidates based on their similarity to the reference queries and their conserved-
199 domain composition.
200 For cross-species copy-number analysis, the longest protein isoform among those passing both the
201 BLASTp and RPS-BLAST criteria was retained for each locus, and each locus was counted once
202 within its assigned family. For cross-species phylogenetic analysis, the same representative-
203 isoform approach was used for the other species, whereas proteins annotated with the .1 isoform
204 suffix in MarpolBase were used for *M. polymorpha*. All annotated protein isoforms of the
205 identified *M. polymorpha* loci were evaluated for isoform-level characterization. The short
206 MpECT12 isoform Mp3g00420.2 was excluded from downstream analyses because it lacked the
207 YTH-domain region, leaving 23 selected protein isoforms. For the subsequent isoform-level

analyses, the 23 selected transcript/protein isoforms listed in Table 1 were used.

The theoretical molecular weights and isoelectric points of the identified *M. polymorpha* proteins were calculated using ExPASy Compute pI/Mw [38-40]. Subcellular localization was predicted using CELLO2GO, Plant-mPLoc v2.0, and LOCALIZER v1.0.4 [41-43]. The identified *M. polymorpha* loci, selected transcript identifiers, physicochemical properties, and predicted localizations are listed in Table 1.

Taxonomic distribution of m^{6}A-related gene copy numbers

Species names were mapped to NCBI taxonomy identifiers using ETE3 NCBITaxa, with manual correction of synonyms where required. NCBI lineage information was used to assign taxonomic ranks and organize the 122 species in a circular taxonomic display. For each species, copy numbers were calculated for five writer groups (MT-A70, Wtap, VIR_N, MTase-10, and RING-HC_HAKAI-like), one eraser group (ALKBH), and four reader groups (YTH, YTH+CPSF, CPSF, and RRM). Counts were derived from the locus-level representative protein sets prepared for the cross-species phylogenetic analyses. Within each group, one representative sequence was retained per species-specific gene identifier, prioritizing the longest isoform among proteins that passed both the BLASTp and conserved-domain criteria. Each represented locus contributed one copy to its assigned group. The taxonomic grouping and ten group-specific copy-number heatmap tracks were visualized in iTOL. The taxonomic display does not represent a species phylogeny inferred from the m^{6}A-related protein sequences.

Multiple sequence alignment and phylogenetic analysis

For the cross-species analyses, records sharing the same species-specific gene identifier within a

conserved-domain dataset were reduced to one representative sequence. For species other than *M. polymorpha*, the longest sequence was prioritized, with BLAST E-values, amino acid identities, BLAST bit scores, and protein identifiers, in that order, used to resolve ties. For *M. polymorpha*, the .1 isoforms annotated in MarpolBase were used. Full-length proteins were analyzed in five writer groups (MT-A70, Wtap, VIR_N, MTase-10, RING-HC_HAKAI-like), one eraser group (ALKBH), and four reader groups (YTH, YTH+CPSF, CPSF, and RRM). Each group was aligned independently with MAFFT v7.526 using the --auto option [44]. No trimming was applied to the resulting alignments.

A maximum-likelihood tree was inferred independently for each alignment with IQ-TREE v3.1.4 [45]. Model selection was performed using ModelFinder with the -m MFP option, and branch support was assessed with 3,000 ultrafast bootstrap replicates and bootstrap-tree refinement with the -bb 3000 -bnni options [46, 47]. To display the independently inferred writer trees in one panel, their consensus trees were grafted onto a predefined synthetic backbone using a custom Python v3.12.14 script based on ETE3 v3.1.3 [48]. The four reader trees were assembled in the same way. The ALKBH eraser tree was displayed directly. Relationships connecting different writer or reader groups were specified by the synthetic backbones and were not inferred or assigned bootstrap support by IQ-TREE. The trees were visualized and annotated with iTOL v7 [49]. Group annotations were based on conserved-domain composition and similarity to *A. thaliana* reference proteins.

For analyses restricted to *M. polymorpha*, the 23 selected protein isoforms were aligned separately for the writer, eraser, and reader groups using MUSCLE [50] implemented in MEGA12. Alignment parameters were set as follows: gap-opening penalty, −2.90; gap-extension penalty, 0.00; hydrophobicity multiplier, 1.20; maximum iterations, 16; clustering method, UPGMA for

both the initial two iterations and subsequent iterations; and minimum diagonal length (λ), 24. Separate maximum-likelihood trees were then generated using MEGA12 [51]. Model selection based on the Bayesian information criterion identified rtREV+G+F, rtREV+I+F, and WAG+G+I+F for the writer, eraser, and reader datasets, respectively. Node support was assessed using 2,000 bootstrap replicates.

Gene structure and conserved domain architecture analysis

Gene structures and protein domain architectures of the identified *M. polymorpha* m^6A-related genes were visualized using custom Python scripts with pandas v2.2.3 and matplotlib v3.8.4. For gene-structure visualization, exon and CDS annotations were extracted from the MpTak_v7.1 GFF3 file using transcript parent IDs. Gene structures and domain architectures were visualized for 23 selected transcript and protein isoforms from 18 loci. The short MpECT12 isoform Mp3g00420.2, which lacks the YTH-domain region present in Mp3g00420.1, was excluded. The genomic start and end positions and strand of each selected mRNA feature were extracted from the GFF3 file to generate a transcript information table, and genomic coordinates were converted into transcript-oriented relative positions. Introns were defined as the intervals between adjacent exons, and UTRs were inferred as exon regions that did not overlap CDSs. Introns were defined as the intervals between adjacent exons, and UTRs were inferred as exon regions that did not overlap CDSs. In the final gene-structure plot, UTRs were shown as blue boxes, CDSs as yellow boxes, and introns as black lines. For domain visualization, the 23 selected protein sequences were extracted from the MpTak_v7.1 protein FASTA file and searched against a local Conserved Domain Database using RPS-BLAST with an E-value cutoff of 1×10^{-3}. Hits to ten specified CDD identifiers were retained in a script-generated tab-delimited file. Protein lengths were calculated

from the extracted protein FASTA file, and the RPS-BLAST query start and end positions were plotted relative to the full-length protein sequences. Overlapping hits from the two selected YTH domain models were collapsed when their overlap covered at least 80% of the shorter hit, prioritizing CDD identifier 410979. Domain names were simplified for plotting, while the original domain descriptions were retained in the tab-delimited RPS-BLAST results. Protein backbones were represented as grey horizontal bars, and annotated domains were displayed as colored rectangles according to domain type.

Chromosomal location and collinearity analysis

The chromosomal locations of the 18 candidate m^{6}A-related genes in *M. polymorpha* were determined from the MpTak_v7.1 GFF3 annotation. The corresponding genome assembly, v1.1 gene annotation, and primary-transcript protein dataset of *Marchantia quadrata* var. MQ were obtained from JGI Phytozome [31]. For *M. polymorpha*, primary protein sequences were mapped to their parent gene loci, with one sequence retained per locus. For *M. quadrata*, the primary-transcript-only protein dataset was used. Transcript-level identifiers of the candidate *M. polymorpha* m^{6}A-related genes were collapsed to their corresponding gene loci, and species-specific prefixes (MP_ and MQ_) were added to gene and chromosome identifiers to prevent identifier conflicts between the two genomes. Gene positions corresponding to the representative protein sequences from both genomes were extracted from the respective genome annotations and formatted as positional input for MCScanX v1.0.0.

Bidirectional protein similarity searches between the genome-wide representative protein sets of the two species were performed using DIAMOND v2.2.8 in BLASTP sensitive mode, with an E-value threshold of 1×10^{-3} and a maximum of 20 target sequences per query [52]. The results from

both search directions were combined, and interspecific collinear blocks were identified using MCScanX with the parameters -b 2 -s 3 -e 1e-5 -m 30 -a [53]. These settings restricted the analysis to interspecific blocks, required a minimum of three anchor gene pairs per block, applied a block E-value threshold of 1×10^{-5}, and set the maximum gap parameter to 30. The -a option limited output to the pairwise collinearity file. Duplicate *M. polymorpha–M. quadrata* collinear gene pairs were collapsed before downstream analysis.

Collinear pairs containing one of the 18 candidate *M. polymorpha* m^6A-related genes were grouped according to whether the candidate belonged to the writer-related, ALKBH-family, or reader-related group. Pairs whose *M. polymorpha* member was not among the 18 candidates were classified as background relationships. The visualization included *M. polymorpha* chromosomes 1–8, U, and V and *M. quadrata* chromosomes 1–9. Collinear relationships were visualized using the Multiple Synteny Plot module of TBtools-II v2.498 [54]. Candidate genes and their detected *M. quadrata* collinear counterparts were highlighted in red, green, and purple for the writer-related, ALKBH-family, and reader-related groups, respectively. Links involving these highlighted loci were displayed in the corresponding colors, including links connecting a highlighted *M. quadrata* locus to a non-candidate *M. polymorpha* gene. All other links were displayed in light grey. Only pairs containing one of the 18 candidate *M. polymorpha* genes were counted as candidate-associated collinear pairs. This analysis was used to assess conservation of chromosomal context and did not independently establish strict one-to-one orthology or the direction of gene duplication or loss.

Identification of *cis*-regulatory elements in the putative promoter regions

For *cis*-regulatory element analysis, 2-kb genomic regions immediately upstream of the annotated

5′ end of each selected transcript model of the identified *M. polymorpha* m^6A-related genes were extracted using the MpTak_v7.1 genome assembly and corresponding annotation. Upstream sequences were extracted according to transcript orientation, with sequences from negative-strand transcripts reverse-complemented. For loci represented by multiple transcript models in the curated dataset, a separate promoter entry was retained for each transcript model, resulting in 23 transcript-level promoter entries corresponding to 18 candidate m^6A-related loci. Entries from the same locus could represent identical or overlapping upstream regions.

Putative *cis*-regulatory elements within the 2-kb promoter sequences were identified using PlantCARE (accessed 27 September, 2026) [55, 56]. Site-level predictions were compiled from the PlantCARE output and processed using custom Python scripts. PlantCARE hits were retained for functional classification when their annotations could be assigned to one of the functional subcategories listed in Table 2. Hits outside these subcategories were excluded from the functional summaries. For Figure 5A, named motif hits were plotted regardless of whether they could be assigned to these subcategories; unnamed motif hits were omitted.

The retained motif hits were grouped into five major categories based on their PlantCARE functional annotations: light-responsive, hormone-responsive, environmental stress-related, development-related, and other elements (Table 2). MYBHv1-binding sites were assigned to the other-elements category. These annotations do not establish regulatory functions in *M. polymorpha*.

To characterize motif distributions, the positions and identities of individual predicted elements were plotted along each 2-kb upstream sequence, from its distal 5′ end toward the annotated transcript 5′ end. A promoter-entry-by-functional-subcategory count matrix was generated by counting retained motif hits assigned to each functional subcategory within each transcript-level promoter entry and was visualized as a heatmap (Figure 5B). Counts within each entry were

additionally aggregated into the five major functional categories and displayed as stacked horizontal bar plots (Figure 5C). Totals reported in Table 2 were calculated across the 23 transcript-level promoter entries. All data processing and visualization were performed using custom Python scripts with pandas, NumPy v1.26.4, matplotlib, and seaborn v0.13.2.

Expression analysis across tissues and developmental stages

Expression data in transcripts per million (TPM) for the 18 candidate *M. polymorpha* m^6A-related genes, comprising seven writer-related genes, five ALKBH-family genes, and six reader-related genes, were obtained from the MarpolBase Expression database (MBEX; accessed 9 September 2026) [57]. The 18 candidate loci were matched to MBEX entries by their MarpolBase gene IDs, and gene-level TPM values were retrieved for expression analysis. The dataset included 16 tissue and developmental-stage conditions, covering a spore germination time course (0–96 h), vegetative and meristematic tissues, male and female reproductive tissues or cells, and early sporophyte tissue. Each tissue or developmental-stage condition was represented by two to four biological replicates. Replicate-level TPM values for the 18 candidate genes across the 16 tissue and developmental-stage conditions, together with their gene names and MarpolBase gene IDs, are provided in Supplementary Table S2 (Additional file 2).

Sample labels were standardized before analysis. Expression values from individual replicates were $\log_2(\text{TPM} + 1)$-transformed and subsequently averaged within each tissue or developmental stage. No gene-wise Z-score transformation was applied. The heatmap display scale was set from −1 to 8, with values outside this range displayed at the corresponding limit. Separate heatmaps were generated for writers, erasers, and readers using the pheatmap and grid packages in R. Genes were hierarchically clustered within each functional class, whereas tissues and developmental

stages were arranged in a predefined biological order without column clustering.

RNA-seq data processing and differential expression analysis under abiotic stress

Raw RNA-seq data were obtained from ArrayExpress under accession number E-MTAB-11141 [28]. The analyzed dataset comprised seven single-stress conditions (cold, heat, osmotic stress, salt stress, nitrogen deficiency, darkness, and high light), 18 pairwise stress combinations, and two unstressed control groups from independent experimental batches, designated Control D2 and Control H2. Sample accession numbers, treatment details, and batch assignments were obtained from the associated SDRF metadata; each condition comprised three biological replicates. Zeitgeber time-course samples were excluded. Transcript abundance was quantified using Salmon v1.12.0 with mapping validation (--validateMappings) against the *M. polymorpha* Tak-1 v5.1r2 cDNA reference transcriptome obtained from MarpolBase [58]. For transcript-to-gene mapping, only transcript identifiers ending in .1 were retained from the Tak-1 v5.1 cDNA annotation. Salmon abundance estimates for retained .1 transcripts were imported into R using tximport v1.30.0, with countsFromAbundance = "lengthScaledTPM", and assigned to their corresponding gene identifiers by removing the terminal .1 suffix [59]. These estimates represent primary-transcript-based abundance rather than the summed abundance of all isoforms of each locus. Gene identifiers in the v5.1 reference were matched to the candidate loci identified using the v7.1 annotation by exact gene identifier matching.

Genes with fewer than 10 estimated counts summed across all analyzed samples were excluded before model fitting; all 18 candidate m^6A-related genes passed this prefilter. A DESeq2 v1.42.0 model was fitted to all retained genes across the 27 sample groups using the design formula ~ StressSimple. Each stress condition was compared separately with Control D2 and Control H2

using Wald tests. Within each contrast, *P* values were adjusted using the Benjamini–Hochberg method before extracting the results for the 18 candidate genes [60]. Under each stress condition, a candidate gene was classified as consistently differentially expressed when both control comparisons yielded adjusted *P* values < 0.05 and $\log_2$ fold changes with the same sign. For each gene–stress combination meeting these criteria, the arithmetic mean of the $\log_2$ fold changes from the two control comparisons was used for heatmap visualization. Combinations that did not meet the criteria were displayed as zero, indicating failure to meet the selection criteria rather than an estimated absence of expression change. Separate heatmap panels were generated for single-stress and pairwise-stress conditions using pheatmap v1.0.12 and grid in R v4.3.3. Both panels used the same asymmetric color scale ranging from −10 to +3, with white assigned to zero, red to positive values, and blue to negative values. Within each panel, genes were hierarchically clustered separately within the writer-related, ALKBH-family, and reader-related groups using Euclidean distance and complete linkage. Clustering was performed on the displayed matrix after entries that did not meet the differential-expression criteria were set to zero and values were limited to the display range. Stress conditions were displayed in a predefined order without column clustering.

qRT-PCR analysis

Quantitative reverse-transcription PCR (qRT-PCR) was performed to analyze transcript levels of the 18 candidate *M. polymorpha* m^6A-related genes. Total RNA was extracted from thalli, gemmae, and gemma cups for tissue comparisons and from control and stress-treated plants for stress-response analyses using the Complex Plant RNA Extraction Kit (Adamas Life, G8028). cDNA was synthesized from 1 μg of total RNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, K1622). qRT-PCR was conducted using ChamQ Universal SYBR

qPCR Master Mix (Vazyme, Q711-03) on a StepOnePlus Real-Time PCR system (Applied Biosystems). *MpSAND* was used as an internal reference based on its previously reported suitability for developmental and abiotic-stress comparisons in *M. polymorpha* [61]. Relative transcript levels were calculated using the $2^{-\Delta\Delta Ct}$ method, with *MpSAND* as the internal reference gene and thallus samples or the corresponding unstressed controls as calibrators for tissue comparisons or stress-response analyses, respectively [62]. For each gene, relative expression values were expressed on a scale with the arithmetic mean of the thallus or corresponding unstressed control group set to 1. Three independent biological replicates were analyzed for each condition, with three technical replicates performed for each biological replicate. Within each qRT-PCR run, Cq values from technical triplicates were averaged for each target and reference gene before calculating relative expression. One relative expression value per biological replicate was used for plotting and statistical analysis. Individual-well Cq values and associated quality metrics for the tissue comparisons (Figure 7), abiotic-stress response analyses (Figure 9), and stress-marker validation (Supplementary Figure 4 in Additional file 3) are provided in Supplementary Table S3 (Additional file 4).

For stress-response analyses, plants were subjected to 40 mM NaCl or 100 mM mannitol treatment, cold treatment at 3 °C for 24 h, heat treatment at 33 °C for 24 h, or darkness for 3 days before sample collection. Primer sequences used for qRT-PCR are provided in Supplementary Table S4 (Additional file 5). Previously published primer sets for *MpSAND* and selected stress-responsive marker genes, including *MpFAD3*, *MpHSP17.8A1*, *MpLEA-like4*, *MpCIPK-B*, *MpCBL-A*, *MpABI3A*, *MpABI5B*, and *MpSIG1*, were adopted from previous studies [61, 63-67], with the corresponding source references provided in Supplementary Table S4 (Additional file 5). The *MpSIG2* primer pair was designed in this study. *MpSIG1* and *MpSIG2* were selected as darkness-

responsive markers based on the previously reported decrease in *MpSIG1* and increase in *MpSIG2* transcript accumulation after 24 h of darkness following growth under continuous light [68].

Protein structure prediction and domain visualization

Three-dimensional structural models of 21 of the 23 candidate m^6A-related protein isoforms were obtained from the AlphaFold Protein Structure Database [69]. Precomputed models from the Tak-1 proteome were used for all proteins except MpVIR.1 (Mp1g10290.1) and MpVIR.2 (Mp1g10290.2), for which models were unavailable. The structures of MpVIR.1 and MpVIR.2 were therefore predicted *de novo* using ColabFold, with multiple sequence alignments generated using MMseqs2 and five recycles used for model refinement. All other prediction parameters were maintained at their default settings [70, 71].

Predicted structures were visualized in PyMOL v2.5.0 using cartoon representation. For visualization only, residues with predicted local distance difference test (pLDDT) scores below protein-specific thresholds ranging from 20 to 70 were hidden. The pLDDT scores were obtained from the B-factor column of each PDB file [69]. Conserved domains were identified using CD-Search against the NCBI Conserved Domain Database with RPS-BLAST, and the resulting domain boundaries were mapped onto the corresponding structural models [36, 37]. MT-A70, Wtap, VIR_N, RING-HC_HAKAI-like, MTase-10, ALKBH, YTH, CPSF, and RRM domains were distinguished using a fixed color scheme, whereas regions without annotated domains were displayed in light grey.

All structures were rendered against a white background using consistent PyMOL settings. Individually rendered structures were assembled into a composite panel to compare the domain-associated structural features of writer, eraser, and reader proteins (Figure 10).

461

462 **Predicted protein-interaction network analysis**

463 Predicted functional associations involving 18 representative candidate *M. polymorpha* m^6A-
464 related proteins, with one protein selected per locus, were retrieved from STRING v12.0 using the
465 uploaded *M. polymorpha* proteome dataset (organism upload ID: STRG0A11VKF) [72]. For loci
466 with multiple annotated isoforms, the .1 isoform was used. Candidate proteins from the writer-
467 related, ALKBH-family, and reader-related groups were submitted as query nodes. The minimum
468 required interaction score was set to 0.4, and the maximum numbers of additional interactors in
469 the first and second shells were set to 20 and 5, respectively. All other STRING parameters were
470 maintained at their default settings.

471 The resulting network was visualized using Cytoscape v3.10.4, with unconnected query proteins
472 retained [73]. Gene identifiers and annotations of additional associated proteins were compiled
473 from MarpolBase. These annotations were used to describe the nodes and do not constitute
474 experimental validation of their functions. Network edges were interpreted as predicted functional
475 protein associations supported by the evidence sources integrated in STRING and not necessarily
476 as direct physical protein–protein interactions [72].

477

478 **Statistical analysis**

479 Statistical analyses were performed using GraphPad Prism 10.5.0. For qRT-PCR analyses, each
480 biological replicate was treated as one independent observation (n = 3 per condition). Statistical
481 tests were performed separately for each gene using the relative expression values. Data are
482 presented as means ± SD, with individual biological replicates shown as points. For qRT-PCR
483 comparisons among tissues and among control and abiotic-stress conditions, data were analyzed

by ordinary one-way analysis of variance (ANOVA), assuming equal variances, followed by Tukey's multiple-comparisons test for all pairwise comparisons. For two-group comparisons between the control and an individual stress treatment in the stress-marker validation experiment, a two-tailed unpaired Welch's t-test was used. For Tukey's test, the family-wise significance level was set at $\alpha = 0.05$, and multiplicity-adjusted *P* values were reported. Significance was denoted as ns ($P \geq 0.05$), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$), and **** ($P < 0.0001$).

Use of large language models

ChatGPT (OpenAI; GPT-6 Astra) was used to assist with the development and revision of custom analysis scripts and with English-language editing. The authors reviewed and tested the scripts, checked the resulting outputs, and critically revised the AI-assisted text. The authors take full responsibility for the analyses, interpretation, and final content of the manuscript.

Results

Genome-wide analysis reveals conserved plant m^{6}A-related gene families and identifies 18 genes in *Marchantia polymorpha*

To define the evolutionary landscape of plant m^{6}A regulation, we first performed a genome-wide identification of putative writer, eraser, and reader proteins across 122 plant genomes. Candidate genes were identified using *Arabidopsis thaliana* m^{6}A-related proteins as reference queries and retained when their encoded proteins contained conserved domains consistent with their family assignments. The annotated domains included MT-A70, Wtap, VIR_N, Methyltransf_10 (MTase-10), RING-HC_HAKAI-like, 2OG-FeII_Oxy_2 (ALKBH), YT521-B homology (YTH), cleavage and polyadenylation specificity factor (CPSF), and RRM_ist3_like (RRM) domains. This survey showed that the three functional classes of m^{6}A-related proteins were broadly represented across the sampled plant lineages, although individual families differed substantially in copy number among species (Figure 2). Canonical writer components generally occurred at relatively low copy numbers, whereas ALKBH-family erasers and reader proteins showed broader copy-number variation. These patterns are consistent with broad conservation of plant m^{6}A regulatory components and lineage-dependent variation in gene family size [14]. Together, the copy-number survey provides a broad evolutionary context for defining the m^{6}A-related gene repertoire in *M. polymorpha*.

Maximum-likelihood phylogenetic analyses were performed separately for each conserved-domain group (Supplementary Figures 1–3). The writer analyses included the MT-A70 group,

containing MTA-, MTB-, and MTC-related clades, and separate FIP37-, VIR-, FIO1-, and HAKAI-related groups. Eraser proteins clustered within ALKBH-related clades. The reader dataset included proteins annotated in four domain categories: YTH, YTH+CPSF, CPSF, and RRM. *M. polymorpha* candidates were placed within the corresponding plant clades, supporting their family assignments.

Using this identification and domain-validation strategy, we identified 18 candidate m^{6}A regulatory genes in *M. polymorpha*, including seven writers, five erasers, and six readers (Table 1). The writer genes comprised *MpMTA*, *MpMTB*, *MpMTC*, *MpFIP37*, *MpVIR*, *MpFIO1*, and *MpHAKAI*. The eraser genes comprised *MpALKBH1A*, *MpALKBH1C*, *MpALKBH6*, *MpALKBH8B*, and *MpALKBH9C*. The reader genes comprised *MpECT2*, *MpECT12*, *MpCPSF30-L1*, *MpCPSF30-L3*, *MpCPSF30-L4*, and *MpCPSF30-L5*. In the selected dataset, five loci—*MpMTA*, *MpFIP37*, *MpVIR*, *MpFIO1*, and *MpALKBH1C*—were each represented by two transcript models, giving a total of 23 transcript models across the 18 loci. Nuclear localization was predicted for many of the identified proteins. Cytoplasmic or organellar localization was also predicted for several proteins, including *MpFIO1*, *MpALKBH6*, *MpALKBH8B*, *MpALKBH9C*, *MpCPSF30-L4*, and *MpCPSF30-L5*, by one or more tools (Table 1). These predictions suggest potential differences in subcellular distribution among the proteins. This set of 18 *M. polymorpha* m^{6}A-related genes provided the basis for subsequent analyses of gene structure, domain architecture, chromosomal distribution and collinearity, putative promoter elements, expression patterns, predicted protein structures, and predicted protein associations.

Gene structure and domain analyses support the annotation of *M. polymorpha* candidate m^{6}A regulators

We next examined the exon–intron structures and conserved domain composition of the identified *M. polymorpha* candidate m^{6}A regulators. Exon–intron structures and protein domain architectures are shown in Figure 3A and B, respectively. Exon–intron organization varied among the three functional classes, although several candidates within related phylogenetic groups showed comparable structural features (Figure 3A). The writer genes displayed substantial structural diversity. Among the writer candidates, MpVIR encoded one of the largest predicted proteins and showed an extended gene model with numerous exons, whereas MpMTC and MpHAKAI displayed more compact gene structures. MpFIP37 and MpFIO1 were each represented by two annotated transcript models, with similar overall exon–intron organization between the two models at each locus. Among the reader proteins, MpCPSF30-L3 had the greatest predicted length (Figure 3B). Overall, these genes showed diverse exon–intron structures, with similarities among some phylogenetically related members.

Conserved-domain analysis revealed the domain architectures of the identified proteins (Figure 3B). Among the writer proteins, MpMTA, MpMTB, and MpMTC contained MT-A70 domains, whereas MpFIP37, MpVIR, MpFIO1, and MpHAKAI contained Wtap, VIR_N, MTase-10, and RING-HC_HAKAI-like domains, respectively. All eraser proteins contained ALKBH-related domains. Proteins assigned to the reader group contained YTH, CPSF, and/or RRM domains. Together, the domain annotations and phylogenetic analyses supported the family assignments of the 18 identified m^{6}A-related genes. We next examined the chromosomal distribution of these genes and their collinear relationships with loci in *Marchantia quadrata*.

Chromosomal distribution and collinearity analysis reveal conserved genomic context for eight *M. polymorpha* m^{6}A-related genes

To place the *M. polymorpha* candidates in a genomic and genus-level comparative context, we examined their chromosomal distribution and interspecific collinearity with the closely related species *M. quadrata*. The 18 candidate m^{6}A-related genes, comprising seven writers, five erasers, and six readers, were distributed across multiple *M. polymorpha* chromosomes rather than being concentrated within a single genomic region (Figure 4). Specifically, the candidates were located on chromosomes 1, 3, 4, 6, 7, and 8, whereas no candidate m^{6}A-related genes were identified on chromosomes 2, 5, U, or V. Chromosome 1 contained the largest number of candidates, with eight loci comprising five writers, two erasers, and one reader. Chromosome 3 contained three candidates. Chromosome 4 contained one candidate, whereas chromosomes 6, 7, and 8 each contained two.

Genome-wide comparison between *M. polymorpha* and *M. quadrata* identified 5,560 nonredundant interspecific collinear gene pairs. Among the 18 *M. polymorpha* m^{6}A-related loci, eight were associated with detectable collinear relationships in *M. quadrata*, comprising three writer, three eraser, and two reader genes (Figure 4). Thus, detectable collinearity was retained for 3 of 7 writer loci, 3 of 5 eraser loci, and 2 of 6 reader loci under the criteria used in this analysis.

At the individual-gene level, the writer genes Mp1g04450, Mp3g09940, and Mp7g09040 formed collinear pairs with *M. quadrata* loci Maqua.1G029900, Maqua.7G060800, and Maqua.8G103000, respectively. Among the eraser candidates, Mp1g08100, Mp1g21480, and Mp3g00190 formed collinear pairs with Maqua.2G120300, Maqua.2G183700, and Maqua.4G032100, respectively. Two reader genes, Mp3g00420 and Mp8g06500, formed collinear pairs with Maqua.4G034600 and Maqua.6G082600, respectively. These relationships involved eight distinct loci in each species and were distributed across several chromosomes rather than being restricted to a single conserved chromosomal segment.

By contrast, no collinear counterpart was detected under the applied parameters for the remaining four writer genes (Mp1g03170, Mp1g08870, Mp1g10290, and Mp1g24270), two eraser genes (Mp6g08620 and Mp7g10560), or four reader genes (Mp1g00140, Mp4g02260, Mp6g15200, and Mp8g18820). The absence of a detected collinear counterpart under the applied criteria does not establish the absence of a homologous gene in *M. quadrata*. Together, these results identified eight *M. polymorpha* m^6A-related loci with conserved genomic context in *M. quadrata*, complementing the sequence-based phylogenetic analyses.

Putative promoters of *M. polymorpha* m^6A-related genes contain abundant light-, hormone-, development-, and stress-associated *cis*-elements

To explore potential transcriptional regulation of *M. polymorpha* m^6A-related genes during development and environmental responses, we analyzed predicted cis-regulatory elements in 2-kb genomic regions immediately upstream of the annotated 5′ ends of 23 transcript models representing 18 loci. These regions were treated as putative promoters, and the positions of predicted *cis*-elements were mapped for each transcript model (Figure 5A).

In total, 636 predicted motif occurrences were retained for functional classification across the 23 putative promoter sequences, including 282 in writer-gene promoters, 153 in eraser-gene promoters, and 201 in reader-gene promoters (Table 2). These elements were grouped into five functional categories based on PlantCARE annotations: light-responsive, hormone-responsive, environmental stress-related, development-related, and other elements (Figure 5B). This classification provided an overview of the potential regulatory inputs associated with the identified m^6A-related genes.

Hormone-responsive elements represented the most abundant major category, with 262 predicted

motif occurrences, including 90 in writer-gene promoters, 74 in eraser-gene promoters, and 98 in reader-gene promoters. Among individual functional subcategories, the light-responsive element subcategory was the most abundant, with 229 occurrences, including 121 in writer-gene promoters, 50 in eraser-gene promoters, and 58 in reader-gene promoters. Hormone-associated elements were also highly represented across the putative promoter regions. Methyl jasmonate (MeJA)-responsive and abscisic acid (ABA)-responsive elements accounted for 138 and 78 predicted motif occurrences, respectively. Auxin-responsive elements were also present, with 22 predicted sites across the analyzed promoter regions. Stress-associated motifs were detected across the putative promoters, including MYB binding sites associated with drought inducibility, low-temperature-responsive elements, defense/stress-responsive elements, anoxic inducibility elements, and anaerobic-responsive elements. Development-related and growth-associated motifs, including motifs annotated as zein metabolism-related motifs, meristem expression elements, endosperm expression elements, circadian elements, and cell cycle regulation elements, were also detected. Because several motif names are derived from annotations in seed plants, these elements should be interpreted as sequence motifs associated with regulatory categories rather than as direct evidence for the same biological processes in *M. polymorpha*.

The distribution of predicted *cis*-elements differed among individual m^6A-related transcript-level promoter entries (Figures 5B and 5C). MpECT12 had the highest number of functionally classified occurrences, including 32 hormone-responsive annotations. Together, these motif annotations generate hypotheses about light-, hormone-, stress-, and development-associated regulations that require experimental testing. These predictions provide a basis for comparing promoter architecture with tissue- and stress-dependent expression patterns in the following sections.

M. polymorpha m^{6}A-related genes show tissue- and developmental-stage-dependent expression

We next examined the expression patterns of the 18 *M. polymorpha* m^{6}A-related genes across 16 tissue and developmental-stage conditions using data from the MarpolBase Expression database (MBEX) [57]. Heatmap analysis showed that all 18 genes had nonzero TPM values in at least one sampled condition, but their expression levels varied across samples (Figure 6). Several genes showed relatively broad expression across samples from the spore germination time course, thallus, reproductive tissues or cells, and early sporophytes. For example, *MpECT2* showed relatively high expression across the sampled conditions, whereas *MpCPSF30-L3* showed lower expression in the 9-day gametophytic apical cell sample and higher expression in the antheridium. Others showed more restricted or condition-enriched expression patterns. Hierarchical clustering within the writer, eraser, and reader groups organized genes according to similarities in their expression profiles across the sampled tissues and developmental stages. These heatmap-based expression profiles provided context for the subsequent qRT-PCR analysis of tissue-associated expression. To further examine tissue-associated expression, we measured transcript levels in thallus, gemmae, and gemma cups using qRT-PCR (Figure 7). Most writer genes showed significant differences in expression among tissues. *MpMTC*, *MpFIO1*, and *MpHAKAI* transcript levels were significantly higher in thallus than in gemmae or gemma cups. *MpMTA* and *MpVIR* were significantly more highly expressed in thallus than in gemma cups, whereas *MpFIP37* was significantly more highly expressed in thallus than in gemmae. No significant differences among tissues were detected for *MpMTB*. Among erasers, *MpALKBH1A* expression was significantly higher in gemmae and gemma cups than in thallus, whereas *MpALKBH1C*, *MpALKBH6*, and *MpALKBH9C* were more highly expressed in thallus than in gemmae or gemma cups. *MpALKBH8B* expression was

significantly higher in thallus than in gemmae, whereas the difference between thallus and gemma cups was not significant. Reader genes also showed tissue-dependent expression: *MpECT12*, *MpCPSF30-L1*, *MpCPSF30-L3*, and *MpCPSF30-L4* were significantly more highly expressed in thallus than in gemmae or gemma cups. *MpECT2* expression was significantly higher in thallus than in gemmae, whereas *MpCPSF30-L5* expression was significantly higher in thallus than in gemma cups. None of the six reader-related genes showed a significant difference between gemmae and gemma cups. Thus, the qRT-PCR results complemented the MBEX expression profiles by revealing gene-specific expression differences among the three tested tissues, with many writer, eraser, and reader genes showing higher expression in the thallus. Together, these expression patterns suggest potential roles for m^6A-related genes in distinct vegetative and reproductive contexts in *M. polymorpha*.

Abiotic stress alters the expression of *M. polymorpha* m^6A-related genes

We examined stress-associated expression changes in the 18 *M. polymorpha* m^6A-related genes by reanalyzing published RNA-seq data [28]. Under the seven single-stress conditions, writer, eraser, and reader genes displayed stress-dependent expression changes relative to unstressed controls (Figure 8A). The response was not uniform across the writer, eraser, and reader gene groups. Under darkness, the dual-control analysis identified increased primary-transcript-based abundance of *MpMTA* (mean $\log_2$FC = 1.54) and *MpVIR* (0.87), whereas *MpFIP37* decreased (−5.23). Among reader-related candidates, *MpECT2* increased under cold (2.10), whereas *MpCPSF30-L3* decreased under darkness (−1.20). Eraser genes also showed divergent responses, suggesting that stress may differentially affect the expression of genes associated with m^6A methylation and demethylation pathways. Reader genes also showed gene-specific increases and decreases across

691 the tested stresses. Heatmaps for the 18 pairwise stress combinations showed expression patterns
692 that differed from those observed under the corresponding individual stresses (Figure 8A, B).
693 These data suggest that *M. polymorpha* m^6A-related genes are associated with stress-responsive
694 transcriptional programs. These RNA-seq analyses provided a broader context for the subsequent
695 qRT-PCR analysis under selected abiotic stress conditions.
696 Expression changes in stress-responsive marker genes confirmed responses to the selected
697 treatments (Supplementary Figure 4). qRT-PCR analysis showed that many *M. polymorpha* m^6A-
698 related genes exhibited altered transcript levels relative to unstressed controls following exposure
699 to cold, heat, osmotic stress, salt stress, or darkness (Figure 9). Darkness significantly increased
700 the transcript levels of multiple genes, including *MpMTA*, *MpMTC*, *MpVIR*, *MpFIO1*,
701 *MpALKBH1A*, *MpECT2*, *MpCPSF30-L1*, and *MpCPSF30-L3*, while significantly reducing those
702 of *MpFIP37*, *MpHAKAI*, and *MpALKBH8B*. Heat significantly increased *MpALKBH9C*
703 expression and reduced *MpVIR* and *MpHAKAI* expression. Osmotic stress significantly reduced
704 the expression of *MpMTC*, *MpVIR*, and *MpCPSF30-L1*. Salt stress significantly increased
705 *MpALKBH1A* and *MpALKBH6* expression and reduced *MpVIR* expression. Several genes showed
706 significantly reduced expression under cold stress, including *MpMTA*, *MpMTC*, *MpFIP37*, *MpVIR*,
707 *MpFIO1*, *MpALKBH8B*, *MpALKBH9C*, *MpCPSF30-L1*, and *MpCPSF30-L4*, whereas
708 *MpCPSF30-L5* expression significantly increased. Together, the RNA-seq and qRT-PCR analyses
709 revealed gene- and stress-dependent expression responses among *M. polymorpha* m^6A-related
710 genes.
711
712 **Predicted protein structures are consistent with conserved domain architectures in *M.***
713 ***polymorpha* m^6A-related proteins**

To visualize the structural context of the sequence-based domain annotations, we examined predicted protein models corresponding to the 23 selected transcript models of the 18 *M. polymorpha* m^6A-related loci. Twenty-one models were retrieved from the AlphaFold Protein Structure Database, and the two MpVIR isoforms were modeled using ColabFold (Figure 10). For each of the five loci with two annotated transcript models—MpMTA, MpFIP37, MpVIR, MpFIO1, and MpALKBH1C—the corresponding protein models are displayed separately under a shared gene label. Annotated domains were mapped onto the predicted models and highlighted using a fixed color scheme, whereas displayed residues outside these domains were shown in grey. Residues below the protein-specific pLDDT thresholds were omitted from the visualization. The mapped regions included MT-A70 domains in MpMTA, MpMTB, and MpMTC; Wtap, VIR_N, MTase-10, and RING-HC_HAKAI-like domains in MpFIP37, MpVIR, MpFIO1, and MpHAKAI, respectively; ALKBH-related domains in the eraser candidates; and YTH, CPSF, or RRM-related domains in the reader candidates (Figure 10). These predicted models provide a structural context for the sequence-based domain annotations and complement the phylogenetic and conserved-domain analyses. However, experimental validation will be required to confirm the biochemical activities and RNA-related functions of these proteins in *M. polymorpha*.

A predicted protein-association network links *M. polymorpha* m^6A-related proteins to RNA-processing factors

Finally, we used STRING to examine predicted functional associations among the identified writer, eraser, and reader proteins and between these proteins and additional proteins (Figure 11). The network contained one main connected component and three isolated protein nodes labelled MpALKBH1C, MpCPSF30-L3, and MpCPSF30-L5. Most candidate m^6A-related proteins were

737 incorporated into the main component, although their connectivity and associated partners differed
738 among the three functional groups.
739 The writer-associated portion of the main component included MpMTA, MpMTB, MpMTC,
740 MpFIP37, MpVIR, MpFIO1, and MpHAKAI. Within this component, MpMTA, MpMTB,
741 MpFIP37, MpVIR, and MpHAKAI showed predicted associations with other writer candidates
742 and additional proteins. Several associated proteins had annotations related to RNA binding or
743 processing, including MpTHO2 (Mp1g20320.1; THO complex subunit 2), Mp5g19650.1 (splicing
744 factor 3B subunit 1), and Mp1g27850.1 (heterogeneous nuclear ribonucleoprotein A1/A3). Studies
745 in other species have implicated THO/TREX components in RNA processing and export and
746 established roles for SF3B1 and hnRNP A1 in pre-mRNA splicing [74-76]. Their occurrence
747 within the writer-associated region therefore suggests potential functional connections between the
748 candidate m^6A machinery and broader RNA-processing pathways. MpFIP37 was also connected
749 to a peripheral subnetwork containing Mp5g12770.1 and Mp2g23770.1, both annotated as GPI-
750 anchor transamidase subunit K, together with proteins with hypothetical or unspecified annotations.
751 Reader candidates, including MpECT2, MpECT12, MpCPSF30-L1, and MpCPSF30-L4, were
752 also included in the main connected component. Their predicted associations with writer
753 candidates and proteins annotated as RNA-processing factors suggest possible links between the
754 reader candidates and RNA-processing pathways. However, MpCPSF30-L3 and MpCPSF30-L5
755 appeared as isolated nodes under the network settings used in this analysis.
756 Among the eraser candidates, MpALKBH1A, MpALKBH6, MpALKBH8B, and MpALKBH9C
757 were incorporated into the main component, whereas the MpALKBH1C node was isolated. The
758 connected ALKBH candidates showed predicted associations with one another and with additional
759 proteins annotated as RNA-related factors, suggesting possible links to RNA metabolism. The

absence of predicted edges for MpALKBH1C, MpCPSF30-L3, and MpCPSF30-L5 under the selected network parameters does not demonstrate the absence of biological interactions, because such interactions may not yet be represented in the databases or may occur only under specific developmental or environmental conditions.

The network was inferred using STRING's integrated functional-association evidence and does not establish direct physical interactions in *M. polymorpha* [72]. Collectively, the predicted network links subsets of the *M. polymorpha* candidate m^6A-related proteins to additional proteins annotated as RNA-related factors. These predictions identify candidate associations for subsequent biochemical and genetic testing.

Discussion

Our analysis identifies 18 candidate m^{6}A-related loci in *Marchantia polymorpha* and places them in a comparative framework spanning 122 plant genomes. The candidates encompass seven writer-related proteins, five ALKBH-family proteins, and six YTH-, CPSF30-, or RRM-related proteins, represented by 23 selected transcript models. Their phylogenetic placement and domain composition support the proposed family assignments, while their expression profiles reveal gene-specific patterns across developmental stages and stress conditions (Figures 2, 3, and 6–9). These data provide a framework for studying RNA methylation in *M. polymorpha*.

The detection of MpMTA, MpMTB, MpFIP37, MpVIR, and MpHAKAI is consistent with the conservation of the core mRNA writer machinery described in *Arabidopsis thaliana* (Figures 2 and 3) [1, 8]. This finding extends earlier reports that recognized several writer homologs in *M. polymorpha*, while the present study analyzes them alongside eraser and reader-related candidates [25]. MpMTC and MpFIO1 warrant separate consideration. Although MpMTC contains an MT-A70-related domain, its participation in the canonical writer complex cannot be assumed from that feature. *A. thaliana* FIO1 modifies U6 snRNA and affects splice-site selection, so the annotation of MpFIO1 as writer-related does not establish that it acts on the same RNAs as MpMTA [4]. Across the sampled species, relatively limited copy-number variation among several writer families contrasts with broader variation in ALKBH- and reader-related groups (Figure 2). Family counts reported for flowering plants and gymnosperms show a similar overall diversity, although direct numerical comparisons depend on genome assembly, isoform filtering, and family

806 definitions [14, 21, 22].

807 The eraser and reader-related candidates illustrate why domain-based annotation must be
808 interpreted at the family level. All five MpALKBH proteins contain an ALKBH-related domain,
809 but the present data cannot show that each removes m^6A from mRNA [7, 13, 35]. Likewise, the
810 YTH domains of MpECT2 and MpECT12 provide a rationale for testing m^6A recognition, whereas
811 the CPSF30-like proteins are not all equivalent to the experimentally characterized *A. thaliana*
812 CPSF30-L reader [9, 15]. In particular, MpCPSF30-L3, MpCPSF30-L4, and MpCPSF30-L5 lack
813 a detectable YTH domain in the displayed annotation (Figure 3B). They should therefore remain
814 reader-related candidates until m^6A binding is demonstrated. The predicted structures provide
815 structural context for the sequence-based domain annotations but cannot resolve substrate
816 specificity, catalytic activity, or RNA-binding capacity (Figure 10). Previous heterologous
817 complementation by a *M. polymorpha* YTHDF-like protein supports conservation of some reader-
818 associated molecular function; it does not establish the endogenous functions of the candidates
819 analyzed here [26].

820 Chromosomal comparisons provide additional evidence of evolutionary conservation. Eight of the
821 18 *M. polymorpha* candidate loci, including members of all three functional groups, showed
822 detectable collinearity with *Marchantia quadrata* (Figure 4). These relationships indicate retention
823 of local genomic context and complement the sequence-based phylogenetic analyses. No collinear
824 counterpart was detected for the remaining ten loci under the applied criteria. The relationships
825 indicate retention of local genomic context for a subset of candidates without establishing one-to-
826 one orthology or the direction of any duplication or loss. In this respect, the collinearity analysis
827 complements sequence phylogenies but does not replace locus-specific evolutionary analysis.

828 The putative promoter analysis suggests possible routes by which the candidate genes could

829 respond to developmental and environmental signals. A total of 636 predicted motif occurrences
830 were retained for functional classification across 23 transcript-level promoter entries representing
831 18 loci (Figure 5 and Table 2). Hormone-responsive motifs were the most abundant broad category,
832 whereas the light-responsive element subcategory had the highest individual count. This
833 distribution suggests potential regulation by hormonal and light signals. Functional labels derived
834 from seed-plant studies, such as endosperm- or seed-associated elements, should not be transferred
835 literally to the liverwort life cycle.
836 The expression analyses indicate potential developmental contexts in which these genes could be
837 investigated. Public expression data varied across the sporeling time course, vegetative tissues,
838 reproductive structures, and sporophyte samples (Figure 6). In the independent qRT-PCR
839 experiment, the thallus showed significantly higher transcript abundance than both gemmae and
840 gemma cups for many candidates, including *MpMTC*, *MpFIO1*, *MpHAKAI*, *MpALKBH1C*,
841 *MpALKBH6*, *MpALKBH9C*, and four reader-related candidates (*MpECT12*, *MpCPSF30-L1*,
842 *MpCPSF30-L3*, and *MpCPSF30-L4*) (Figure 7). *MpALKBH1A* transcript abundance was
843 significantly higher in gemmae and gemma cups than in thallus, whereas *MpMTB* showed no
844 significant differences among the three tissues. These differences are consistent with regulated
845 deployment of RNA-related factors across vegetative and asexual tissues, but transcript abundance
846 alone cannot establish a developmental function or a change in m^6A deposition. The broader
847 database profiles and the qRT-PCR experiments sample different developmental materials and
848 should be interpreted as complementary views rather than paired measurements of the same
849 specimens.
850 Abiotic stress also elicited gene-specific responses. Reanalysis of the public RNA-seq data
851 detected changes under single and combined stresses, with some combinations differing from the

corresponding individual treatments (Figure 8). In our qRT-PCR experiments, darkness was associated with increased transcript abundance for several candidates, including *MpMTA*, *MpMTC*, *MpVIR*, *MpFIO1*, *MpALKBH1A*, *MpECT2*, *MpCPSF30-L1*, and *MpCPSF30-L3*; cold, heat, osmotic treatment, and salinity produced different responses across the gene set (Figure 9). This diversity is consistent with the context-dependent expression reported in other plant gene-family surveys [21-23]. The RNA-seq and qRT-PCR analyses provide complementary evidence of stress-responsive expression, although their treatment designs and biological materials differ. Time-resolved methylation profiling and functional assays could clarify how these expression responses relate to RNA metabolism during stress.

The predicted association network provides possible connections to broader RNA metabolism. Several writer-related candidates appear within the main component together with MpTHO2 and proteins annotated as splicing factor 3B subunit 1 (Mp5g19650.1) and heterogeneous nuclear ribonucleoprotein A1/A3 (Mp1g27850.1) (Figure 11). STRING evidence supports predicted functional associations, which cannot be interpreted as proof of direct physical interactions [72]. Predicted cellular localizations also warrant direct testing: for example, the nuclear localization predicted for MpECT2 should be examined in light of the cytoplasmic localization reported for *A. thaliana* ECT2, ECT3, and ECT4 [15]. Localization imaging, biochemical interaction tests, and RNA-binding assays would distinguish these possibilities.

The combined results prioritize experimentally tractable questions. Loss-of-function and complementation analyses could test whether specific candidate genes affect gemma cup development or stress responses; m^6A mapping and measurements on defined RNA classes could then establish which effects depend on methylation. Direct assays of MpMTA- and MpFIO1-related activities, ALKBH-dependent demethylation, YTH-dependent RNA binding, and CPSF30-

associated 3′ end processing would clarify whether the functions described in flowering plants are retained or altered in *M. polymorpha* [4, 7-9, 15]. Together, this study places the m^6A-related gene repertoire of *M. polymorpha* in a comparative framework spanning 122 plant genomes and links family conservation with tissue- and stress-associated expression patterns. These findings provide a foundation for investigating RNA methylation in liverwort development and environmental responses.

Conclusions

This study identifies 18 candidate m^6A-related genes in *Marchantia polymorpha*—seven writer-related, five ALKBH-related, and six reader or reader-related loci—and places them within a survey of 122 plant genomes. Conserved domains and phylogenetic relationships support their family assignments, while collinearity with *Marchantia quadrata* identifies eight candidates with retained genomic context. Promoter analysis identifies putative regulatory elements, while expression analyses reveal variation across tissues, developmental stages, and abiotic stress conditions. Predicted protein structures and associations provide additional context for investigating the functions of the identified proteins. Together, these findings provide a comparative framework and candidate genes for investigating the roles of RNA methylation in liverwort development and stress responses.

List of abbreviations

ABA	abscisic acid
ALKBH	AlkB homolog
ANOVA	analysis of variance
BLAST	Basic Local Alignment Search Tool
CDD	Conserved Domain Database
CDS	coding sequence
CPSF	cleavage and polyadenylation specificity factor
ECT	evolutionarily conserved C-terminal region
FIO1	FIONA1
FIP37	FKBP12 interacting protein 37 kDa
iTOL	Interactive Tree Of Life
m^6A	N^6-methyladenosine
MBEX	MarpolBase Expression database
MeJA	methyl jasmonate
MTA	methyltransferase A
MTB	methyltransferase B
MTC	methyltransferase C
MW	molecular weight
pI	isoelectric point

pLDDT	predicted local distance difference test
PPI	protein–protein interaction
qRT-PCR	quantitative reverse-transcription polymerase chain reaction
RNA-seq	RNA sequencing
RPS-BLAST	Reverse Position-Specific BLAST
RRM	RNA recognition motif
Tak-1	Takaragaike-1
TPM	transcripts per million
UTR	untranslated region
VIR	VIRILIZER
YTH	YT521-B homology

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The public genomic and protein sequence datasets analyzed in this study were obtained from Ensembl Plants release 62, The Arabidopsis Information Resource (TAIR; Araport11), JGI Phytozome (Marchantia quadrata var. MQ v1.1), and MarpolBase (Marchantia polymorpha

MpTak_v7.1). The Tak-1 v5.1r2 reference transcriptome used for RNA-seq quantification was also obtained from MarpolBase. Tissue- and developmental-stage expression data were obtained from the MarpolBase Expression database (MBEX), and the replicate-level TPM values used in this study are provided in Supplementary Table S2 (Additional file 2). The abiotic-stress RNA-seq dataset is available in ArrayExpress through BioStudies under accession E-MTAB-11141. Raw qRT-PCR Cq values and associated quality metrics generated in this study are provided in Supplementary Table S3 (Additional file 4). The custom analysis scripts and associated input files have been deposited in Zenodo (https://doi.org/10.5281/zenodo.23330317).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Suzhou Science and Technology Bureau (grant no. ZSL2024334) and the Startup Fund of Duke Kunshan University (grant no. 00AKUG0122), both awarded to EYK.

Authors' contributions

EYK: Conceptualization, Supervision, Project administration, Funding acquisition, Visualization, Writing – review & editing. JTY: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Validation, Visualization, Writing – original draft. SM: Methodology, Software, Formal analysis, Investigation. YL: Visualization. JTY and SM contributed equally to this work. All authors read and approved the final manuscript.

Acknowledgements

We thank the members of the CReAtE (Climate-Resilient Agricultural Engineering) Lab at Duke Kunshan University for their support throughout this study.

Figure Legends

Figure 1. Plant m^{6}A regulatory pathways and the analytical workflow used in this study.

a Schematic of plant N^6-methyladenosine (m^{6}A) regulation. Nuclear RNA polymerase II (Pol II) transcribes DNA into precursor mRNA (pre-mRNA), and the MTA, MTB, FIP37, VIR, and HAKAI writer complex adds m^{6}A. The long isoform of cleavage and polyadenylation specificity factor 30 (CPSF30-L) is shown as a nuclear reader associated with 3′ end processing and polyadenylation. Evolutionarily conserved C-terminal region (ECT) proteins represent cytoplasmic readers; selected AlkB homolog (ALKBH) proteins represent nuclear or cytoplasmic erasers. Orange, lavender, and pale-blue shapes denote writers, readers, and ALKBH proteins, respectively. Yellow circles denote m^{6}A; white circles labeled A denote unmodified adenosine. The divider separates nuclear and cytoplasmic compartments; arrows indicate RNA processing and transport.

b Examples of RNA processes associated with these proteins: FIONA1 (FIO1), U6 small nuclear RNA (snRNA), and pre-mRNA splicing; CPSF30-L and alternative 3′ end formation; MTA and primary microRNA (pri-miRNA) processing; and ECT proteins and transcript stability or stress-associated decay in cytosolic condensates. The ribosome and pink peptide chain depict translation. DICER-LIKE 1 (DCL1) and HYPONASTIC LEAVES 1 (HYL1) are indicated in the small-RNA pathway. These schematics do not establish corresponding functions in *Marchantia polymorpha*.

c Genome-wide searches across 122 plant genomes identified candidate m^{6}A regulators. *M. polymorpha* candidates were classified as writers (orange), erasers (light blue), or readers

(lavender). Analyses covered phylogeny, gene structure, conserved domains, chromosomal collinearity, *cis*-elements, expression profiles, tissue- and stress-response quantitative reverse-transcription PCR (qRT-PCR), predicted protein structures, and a protein–protein interaction (PPI) network of predicted associations. Pale orange, mint, and lavender panels group gene architecture/collinearity, regulation/expression, and protein structure/interaction analyses, respectively. Grey arrows indicate the workflow.

Figure 2. Distribution of candidate m^{6}A regulator families across 122 plant species.

Each radial column represents a species. From the outermost to innermost ring, the ten rings show copy numbers for RRM_ist3_like (RRM), Cleavage and polyadenylation specificity factor (CPSF), YTH plus CPSF (YTH+CPSF), YTH, 2OG-FeII_Oxy_2 (ALKBH), RING-HC_HAKAI-like, Methyltransf_10 (MTase-10), VIR_N, Wtap, and MT-A70 categories. Within each category, stronger color intensity indicates a higher value according to its separate gradient in the key; black indicates zero detected copies. Colored taxonomic sectors correspond to the lineages identified in the left-hand key, and the central tree depicts relationships among the sampled species. BOP and PACMAD designate the Bambusoideae-Oryzoideae-Pooideae and Panicoideae-Aristidoideae-Chloridoideae-Micrairoideae-Arundinoideae-Danthonioideae grass clades, respectively. One protein isoform per locus was retained for cross-species counts. The central tree and colored sectors represent NCBI taxonomy and do not constitute an m^{6}A-protein phylogeny.

Table 1. Identified candidate m^{6}A-related genes in *M. polymorpha*.

The table lists 18 loci—seven writers, five erasers, and six readers—and their 23 selected transcript entries. Gene names, locus and transcript identifiers, transcript genomic coordinates, strand,

transcript genomic span, predicted protein length, molecular weight (MW), isoelectric point (pI), and predicted subcellular localization are shown. Localization abbreviations are Nu, nucleus; Cp, chloroplast; Mt, mitochondrion; and Cy, cytoplasm. Superscripts a, b, and c identify predictions from CELLO2GO, Plant-mPLoc v2.0, and LOCALIZER v1.0.4, respectively; predictions from different programs are reported separately. bp, base pairs; aa, amino acids; kDa, kilodaltons; Mp, *M. polymorpha*.

Figure 3. Exon–intron structures and conserved protein domains of *M. polymorpha* candidate m^{6}A regulators.

a Separate maximum-likelihood gene trees and transcript-specific gene structures of the writer, eraser, and reader candidates. Each row represents a selected transcript isoform. Yellow boxes indicate coding sequences (CDS), dark-blue boxes untranslated regions (UTRs), and thin black lines introns. The horizontal scale is in kilobases (kb).

b Conserved domains mapped onto the proteins encoded by the transcript models in **a**. Grey bars indicate full protein lengths. Colored blocks indicate MT-A70 (very pale blue), Wtap (pale blue), VIR_N (blue), RING-HC_HAKAI-like (medium blue), MTase-10 (dark blue), 2OG-FeII_Oxy_2 (orange), YTH (yellow-green), CPSF (green), and RRM (dark green) domains, as keyed in the panel. The lower scale is in amino acids (aa).

Figure 4. Chromosomal positions and interspecific collinearity of candidate m^{6}A-related genes in *Marchantia*.

Chromosomes of *M. polymorpha* (MP_, teal bars, top) and *Marchantia quadrata* (MQ_, lime-green bars, bottom) are shown in reference-assembly order. The input data contain 5,560

interspecific collinear gene pairs. Light-grey curves represent 5,551 pairs without highlighted endpoints. Eight colored curves connect an *M. polymorpha* candidate to an *M. quadrata* locus: three writer-associated links in red, three eraser-associated links in green, and two reader-associated links in purple. A ninth colored curve, also green, connects *M. polymorpha* Mp1g16590 to *M. quadrata* Maqua.2G183700. Mp1g16590 is not among the 18 candidates in Table 1; this curve is highlighted because its *M. quadrata* endpoint is also linked to the eraser candidate Mp1g21480. Correspondingly colored triangles mark candidate loci and their detected collinear counterparts. MP_ and MQ_ are species-specific chromosome prefixes; U and V designate the sex chromosomes labeled in the *M. polymorpha* assembly.

Table 2. Predicted *cis*-elements in putative promoters of *M. polymorpha* candidate m^{6}A-related genes.

Predicted motif occurrences in the 2-kb upstream sequences of 23 transcript entries representing 18 loci are grouped by 21 PlantCARE functional subcategories. The Total column gives the number of predicted occurrences of each element across all entries; Writer, Eraser, and Reader give the corresponding counts within each candidate group. Across the listed elements, 636 occurrences were recorded: 282 in writer, 153 in eraser, and 201 in reader promoter entries. These are motif counts, not counts of genes with a motif. When transcript models have identical or overlapping upstream regions, their predictions contribute separately to the transcript-entry totals. MeJA, methyl jasmonate; MYB and MYBHv1, motif names used in the PlantCARE annotations. The functional categories are predictions and do not establish regulatory activity in *M. polymorpha*.

Figure 5. Predicted *cis*-elements in putative promoters of *M. polymorpha* candidate m^{6}A-

related genes.

a Positions of 2,619 named motif occurrences predicted by PlantCARE within 2-kb upstream sequences of 23 selected transcript models representing 18 candidate loci. Each colored tick denotes a predicted motif; its color corresponds to the motif named in the key beneath the plot. Motif positions are shown in base pairs from the distal end of the 2-kb upstream sequence (0 bp) toward the annotated transcript 5′ end (2,000 bp). Separate rows for isoforms of the same locus represent the analyzed transcript-level promoter entries. Colored label backgrounds distinguish writer (pink), eraser (green), and reader (purple) entries.

b Heatmap of 636 functionally classified motif occurrences across 21 annotated functional subcategories for the same 23 promoter entries. Darker red indicates more annotated occurrences, whereas pale or white cells indicate few or none. The colored strip above the heatmap identifies the five major categories.

c Stacked bars aggregate sites into light-responsive (purple), hormone-responsive (yellow), environmental stress-related (green), development-related (light blue), and other (orange) categories. The bar-plot axis gives the number of predicted motif occurrences. These are PlantCARE motif annotations, rather than experimentally established functions in *M. polymorpha*. MeJA, methyl jasmonate; ABA, abscisic acid; MYB, MYB-family binding motif.

Figure 6. Expression of *M. polymorpha* candidate m^{6}A-related genes across tissues and developmental stages.

Writer, eraser, and reader genes are shown in separately clustered blocks, with dendrograms at left. Columns show the spore time course (0–96 h), gametophytic apical cell, thallus, midrib, gemma cup, male and female reproductive tissues or cells, and sporophyte; sample identity, age, and

accession are indicated above the heatmap. Values are condition means of replicate-level $\log_2$ (TPM + 1) values from the MBEX data; the display scale is bounded at −1 and 8. Red indicates higher transformed expression, blue lower expression, and white values near $\log_2$ (TPM + 1) = 1. h, hours; d, days.

Figure 7. Tissue-associated expression of 18 *M. polymorpha* candidate m^{6}A-related genes measured by qRT-PCR.

Individual panels for the seven writer, five eraser, and six reader candidates are arranged in that order from left to right and top to bottom. Thallus is represented by black points, gemmae by blue points, and gemma cups—labeled Cup in the plots—by magenta points. Relative transcript levels were calculated using the 2−ΔΔCt method, with *MpSAND* as the internal reference and thallus samples as the calibrator. Bars represent means, error bars indicate SD, and points represent individual biological replicates (n = 3 per condition). Statistical comparisons used ordinary one-way ANOVA followed by Tukey's multiple-comparisons test for all tissue pairs. Groups without a shared uppercase letter differ at adjusted $P < 0.05$.

Figure 8. Differential expression of *M. polymorpha* candidate m^{6}A-related genes under abiotic stress in RNA-seq data.

a Single-stress conditions. **b** Pairwise stress combinations; "x" in a column label denotes a combined treatment. Writers, erasers, and readers are clustered separately. Cells show the mean $\log_2$ fold change ($\log_2$FC) relative to Control D2 and Control H2 when adjusted $P < 0.05$ and the direction of change agreed in both comparisons; otherwise, the displayed value is zero. Red denotes increased expression, blue decreased expression, and white a display value of zero, which

does not necessarily indicate an estimated log_2FC of zero. The color scale ranges asymmetrically from −10 to +3. N. deficiency, nitrogen deficiency; high light, elevated-light treatment.

Figure 9. Expression of 18 *M. polymorpha* candidate m^{6}A-related genes under abiotic stress measured by qRT-PCR.

Treatments are Normal (black points), cold at 3 °C for 24 h (blue), heat at 33 °C for 24 h (pink), osmotic stress induced by 100 mM mannitol (orange), salt stress induced by 40 mM NaCl (green), and darkness for 3 d (grey). Relative transcript levels were calculated using the $2^{-\Delta\Delta Ct}$ method, with *MpSAND* as the internal reference and the corresponding Normal samples as the calibrator. Bars represent means, error bars indicate SD, and points represent individual biological replicates (n = 3 per condition). Statistical comparisons used ordinary one-way ANOVA followed by Tukey's multiple-comparisons test for all treatment pairs. Groups without a shared uppercase letter differ at adjusted $P < 0.05$.

Figure 10. Predicted three-dimensional structures of *M. polymorpha* candidate m^{6}A-related protein isoforms.

Structures are grouped as writers, erasers, and readers, with protein identifiers and family names beneath the models. Models were obtained from the AlphaFold Protein Structure Database except for the two MpVIR isoforms, which were modeled using ColabFold. Colored ribbon segments indicate annotated domains: MT-A70 (blue), Wtap (orange), VIR_N (green), RING-HC_HAKAI-like (red), MTase-10 (mauve), ALKBH (yellow), YTH (teal), CPSF (pink), and RRM (brown). Displayed regions outside annotated domains are grey; low-confidence regions excluded for visualization are not shown.

Figure 11. STRING-derived protein–protein interaction (PPI) network of *M. polymorpha* candidate m^{6}A-related proteins and predicted functional associates.

Circles denote labeled proteins. The 18 candidate query proteins were represented by one protein per locus; the .1 isoform was used for loci with multiple annotated isoforms. Within the main connected component, candidate writer nodes are salmon, eraser nodes orange, and reader nodes blue. The figure displays three unconnected candidate nodes—MpALKBH1C, MpCPSF30-L3, and MpCPSF30-L5—below the connected component. Edge colors indicate STRING evidence channels: experimentally determined (magenta), gene neighborhood (green), text mining (olive), gene fusion (red), curated databases (blue), gene co-occurrence (dark blue), co-expression (black), and protein homology (lavender). The minimum required STRING association score was 0.4, with up to 20 first-shell and five second-shell additional interactors. Edges represent predicted functional associations.

Supplementary Figure 1. Composite display of maximum-likelihood phylogenies of candidate m^{6}A writer proteins across 122 plant species.

Colored outer sectors delimit MTB (pink), MTA (peach), MTC (pale yellow), FIP37 (pale green), VIR (light blue), FIO1 (periwinkle), and HAKAI (lavender) clades. Within each domain-defined writer group, branches represent IQ-TREE-inferred protein relationships. The arrangement of the writer groups was imposed using a synthetic backbone for visualization and does not represent an inferred relationship among the groups; bootstrap support therefore applies only within each domain-defined subtree. One representative isoform per locus was retained.

Supplementary Figure 2. Maximum-likelihood phylogeny of candidate ALKBH proteins across 122 plant species.

Colored outer sectors identify ALKBH1A (light blue), ALKBH1B (blue), ALKBH1C (lavender), ALKBH1D (light purple), ALKBH2 (light green), ALKBH6 (pale yellow), ALKBH8B (orange), ALKBH8B-2 (teal), and ALKBH9/10 (pink) groups. Group labels follow the Arabidopsis reference-query nomenclature. One representative isoform per locus was retained.

Supplementary Figure 3. Composite display of maximum-likelihood phylogenies of candidate m^{6}A reader-related proteins across 122 plant species.

Colored outer sectors distinguish YTH (pink), combined YTH+CPSF (peach), CPSF (pale yellow), and RRM (pale green) protein/domain categories. Within each domain-defined reader group, branches represent IQ-TREE-inferred protein relationships. The arrangement of the reader groups was imposed using a synthetic backbone for visualization and does not represent an inferred relationship among the groups; bootstrap support therefore applies only within each domain-defined subtree. One representative isoform per locus was retained.

Supplementary Figure 4. qRT-PCR expression of abiotic-stress marker genes under the treatments used in this study.

MpFAD3 is shown under cold; *MpHSP17.8A1* under heat; *MpLEA-like4*, *MpCIPK-B*, and *MpCBL-A* under osmotic stress; *MpLEA-like4*, *MpABI3A*, and *MpABI5B* under salt; and *MpSIG1* and *MpSIG2* under darkness. Relative transcript levels were calculated using the $2^{-\Delta\Delta Ct}$ method, with *MpSAND* as the internal reference and the corresponding Normal samples as the calibrator. Control points are black; cold, heat, osmotic, salt, and dark-treatment points are blue, pink, orange, green,

and grey, respectively. Bars represent means, error bars indicate SD, and points represent individual biological replicates (n = 3 per condition). Brackets indicate comparisons assessed by two-tailed unpaired Welch's t-tests; * indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$, and **** indicates $P < 0.0001$.

Supplementary Table S1. *Arabidopsis thaliana* reference genes used for the identification of candidate m^{6}A-related genes.

Gene names and *A. thaliana* locus identifiers are provided for the reference genes used to select protein queries for homology searches. The reference set comprises seven writer-related genes, 14 ALKBH-family genes, and 17 reader-related genes.

Supplementary Table S2. Transcript abundance of candidate m^{6}A-related genes across tissues and developmental stages in *M. polymorpha*.

Replicate-level transcript abundance values for 18 candidate m^{6}A-related genes across 16 tissue and developmental-stage conditions were obtained from the MarpolBase Expression database (MBEX) and are expressed as transcripts per million (TPM). Gene names and corresponding MarpolBase gene identifiers are provided above the expression columns. Each row represents a biological replicate, with two to four replicates per condition. These data were used for the expression analysis shown in Figure 6.

Supplementary Table S3. Raw qRT-PCR Cq values and associated quality metrics.

The workbook contains nine worksheets: Tissue_1–Tissue_3 for tissue comparisons (Figure 7), Stress_1–Stress_3 for abiotic-stress response analyses of candidate m6A-related genes (Figure 9),

and Reference_1–Reference_3 for stress-marker validation (Supplementary Figure 4). The suffixes 1–3 indicate the three independent biological replicates. Each row represents an individual qRT-PCR well. Well numbering restarts for each qRT-PCR run, and target Cq values were normalized to the *MpSAND* measurements from the same run. The data include well positions, sample and target identifiers, individual-well Cq values, amplification scores, Cq confidence values, and instrument-reported Cq means and standard deviations.

Supplementary Table S4. Primer sequences used for qRT-PCR analysis in *M. polymorpha*.

Forward (F) and reverse (R) primer sequences are provided for the 18 candidate m^6A-related genes, nine abiotic-stress marker genes, and the internal reference gene *MpSAND*. All sequences are presented in the 5′–3′ direction. Sources of previously published primer pairs are indicated by reference numbers corresponding to the manuscript reference list.

References

1. Zhong S, Li H, Bodi Z, Button J, Vespa L, Herzog M, Fray RG: MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell 2008, 20(5):1278-1288.
2. Luo GZ, MacQueen A, Zheng G, Duan H, Dore LC, Lu Z, Liu J, Chen K, Jia G, Bergelson J et al: Unique features of the m6A methylome in Arabidopsis thaliana. Nat Commun 2014, 5:5630.
3. Hou Y, Sun J, Wu B, Gao Y, Nie H, Nie Z, Quan S, Wang Y, Cao X, Li S: CPSF30-L-mediated recognition of mRNA m(6)A modification controls alternative polyadenylation of nitrate signaling-related gene transcripts in Arabidopsis. Mol Plant 2021, 14(4):688-699.
4. Parker MT, Soanes BK, Kusakina J, Larrieu A, Knop K, Joy N, Breidenbach F, Sherwood AV, Barton GJ, Fica SM et al: m(6)A modification of U6 snRNA modulates usage of two major classes of pre-mRNA 5' splice site. Elife 2022, 11:e78808.
5. Song P, Wei L, Chen Z, Cai Z, Lu Q, Wang C, Tian E, Jia G: m(6)A readers ECT2/ECT3/ECT4 enhance mRNA stability through direct recruitment of the poly(A) binding proteins in Arabidopsis. Genome Biol 2023, 24(1):103.
6. Chen T, Greene GH, Motley J, Mwimba M, Luo GZ, Xu G, Karapetyan S, Xiang Y, Liu C, He C et al: m(6)A modification plays an integral role in mRNA stability and translation during pattern-triggered immunity. Proc Natl Acad Sci U S A 2024, 121(33):e2411100121.
7. Duan HC, Wei LH, Zhang C, Wang Y, Chen L, Lu Z, Chen PR, He C, Jia G: ALKBH10B

Is an RNA N(6)-Methyladenosine Demethylase Affecting Arabidopsis Floral Transition. Plant Cell 2017, 29(12):2995-3011.

8. Ruzicka K, Zhang M, Campilho A, Bodi Z, Kashif M, Saleh M, Eeckhout D, El-Showk S, Li H, Zhong S et al: Identification of factors required for m(6) A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. New Phytol 2017, 215(1):157-172.

9. Song P, Yang J, Wang C, Lu Q, Shi L, Tayier S, Jia G: Arabidopsis N(6)-methyladenosine reader CPSF30-L recognizes FUE signals to control polyadenylation site choice in liquid-like nuclear bodies. Mol Plant 2021, 14(4):571-587.

10. Shen L, Liang Z, Gu X, Chen Y, Teo ZWN, Hou X, Cai WM, Dedon PC, Liu L, Yu H: N(6)-Methyladenosine RNA Modification Regulates Shoot Stem Cell Fate in Arabidopsis. Dev Cell 2016, 38(2):186-200.

11. Cai Z, Tang Q, Song P, Tian E, Yang J, Jia G: The m6A reader ECT8 is an abiotic stress sensor that accelerates mRNA decay in Arabidopsis. Plant Cell 2024, 36(8):2908-2926.

12. Bhat SS, Bielewicz D, Gulanicz T, Bodi Z, Yu X, Anderson SJ, Szewc L, Bajczyk M, Dolata J, Grzelak N et al: mRNA adenosine methylase (MTA) deposits m(6)A on pri-miRNAs to modulate miRNA biogenesis in Arabidopsis thaliana. Proc Natl Acad Sci U S A 2020, 117(35):21785-21795.

13. Martinez-Perez M, Aparicio F, Lopez-Gresa MP, Belles JM, Sanchez-Navarro JA, Pallas V: Arabidopsis m(6)A demethylase activity modulates viral infection of a plant virus and the m(6)A abundance in its genomic RNAs. Proc Natl Acad Sci U S A 2017, 114(40):10755-10760.

14. Zhang J, Wu L, Mu L, Wang Y, Zhao M, Wang H, Li X, Zhao L, Lin C, Zhang H et al:

Evolution and post-transcriptional regulation insights of m(6)A writers, erasers, and readers in plant epitranscriptome. Plant J 2024, 120(2):505-525.

15. Arribas-Hernandez L, Bressendorff S, Hansen MH, Poulsen C, Erdmann S, Brodersen P: An m(6)A-YTH Module Controls Developmental Timing and Morphogenesis in Arabidopsis. Plant Cell 2018, 30(5):952-967.

16. Arribas-Hernandez L, Simonini S, Hansen MH, Paredes EB, Bressendorff S, Dong Y, Ostergaard L, Brodersen P: Recurrent requirement for the m(6)A-ECT2/ECT3/ECT4 axis in the control of cell proliferation during plant organogenesis. Development 2020, 147(14):dev189134.

17. Wu X, Su T, Zhang S, Zhang Y, Wong CE, Ma J, Shao Y, Hua C, Shen L, Yu H: N(6)-methyladenosine-mediated feedback regulation of abscisic acid perception via phase-separated ECT8 condensates in Arabidopsis. Nat Plants 2024, 10(3):469-482.

18. Nishat ZS, Hasan MS, Islam MS, Hossain T, Ghosh A: Identification of epitranscriptomic methylation marker genes in Arabidopsis and their expression profiling in response to developmental, anatomical, and environmental modulations. Current Plant Biology 2022, 30:100247.

19. Sun X, Wu W, Yang Y, Wilson I, Shao F, Qiu D: Genome-Wide Identification of m(6)A Writers, Erasers and Readers in Poplar 84K. Genes (Basel) 2022, 13(6):1018.

20. Tang L, Xue J, Ren X, Zhang Y, Du L, Ding F, Zhou K, Ma W: Genome-Wide Identification and Expression Analysis of m6A Writers, Erasers, and Readers in Litchi (Litchi chinensis Sonn.). Genes (Basel) 2022, 13(12):2284.

21. Hasan M, Nishat ZS, Hasan MS, Hossain T, Ghosh A: Identification of m(6)A RNA methylation genes in Oryza sativa and expression profiling in response to different

developmental and environmental stimuli. Biochem Biophys Rep 2024, 38:101677.

22. Yao S, Song Y, Cheng X, Wang D, Li Q, Zhang J, Chen Q, Yu Q, Ji K: Transcriptome-Wide Identification of m(6)A Writers, Erasers and Readers and Their Expression Profiles under Various Biotic and Abiotic Stresses in Pinus massoniana Lamb. Int J Mol Sci 2024, 25(14):7987.

23. Shan C, Dong K, Wen D, Ye Z, Hu F, Zekraoui M, Cao J: Writers, readers, and erasers of N6-Methyladenosine (m6A) methylomes in oilseed rape: identification, molecular evolution, and expression profiling. BMC Plant Biol 2025, 25(1):147.

24. Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F et al: Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell 2017, 171(2):287-304.e15.

25. Pietrykowska H, Sierocka I, Zielezinski A, Alisha A, Carrasco-Sanchez JC, Jarmolowski A, Karlowski WM, Szweykowska-Kulinska Z: Biogenesis, conservation, and function of miRNA in liverworts. J Exp Bot 2022, 73(13):4528-4545.

26. Flores-Tellez D, Tankmar MD, von Bulow S, Chen J, Lindorff-Larsen K, Brodersen P, Arribas-Hernandez L: Insights into the conservation and diversification of the molecular functions of YTHDF proteins. PLoS Genet 2023, 19(10):e1010980.

27. Tsuboyama S, Okumura T, Attri P, Koga K, Shiratani M, Kuchitsu K: Growth control of Marchantia polymorpha gemmae using nonthermal plasma irradiation. Sci Rep 2024, 14(1):3172.

28. Tan QW, Lim PK, Chen Z, Pasha A, Provart N, Arend M, Nikoloski Z, Mutwil M: Cross-stress gene expression atlas of Marchantia polymorpha reveals the hierarchy and regulatory principles of abiotic stress responses. Nat Commun 2023, 14(1):986.

29. Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L et al: Ensembl Genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Res 2020, 48(D1):D689-D695.
30. Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD: Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J 2017, 89(4):789-804.
31. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al: Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 2012, 40(Database issue):D1178-D1186.
32. Potente G, Yasui Y, Shimokawa E, Jenkins J, Walstead RN, Grimwood J, Schmutz J, Leebens-Mack J, Bruna T, Kaur N et al: Insights into convergent evolution of cosexuality in liverworts from the Marchantia quadrata genome. Cell Rep 2025, 44(4):115503.
33. Tanizawa Y, Mochizuki T, Yagura M, Sakamoto M, Fujisawa T, Kawamura S, Shimokawa E, Yamaoka S, Nishihama R, Bowman JL et al: MarpolBase: genome database for Marchantia polymorpha featuring high quality reference genome sequences. Plant Cell Physiol 2026, 67(3):377-388.
34. Hu J, Manduzio S, Kang H: Epitranscriptomic RNA Methylation in Plant Development and Abiotic Stress Responses. Front Plant Sci 2019, 10:500.
35. Leihne V, Kirpekar F, Vagbo CB, van den Born E, Krokan HE, Grini PE, Meza TJ, Falnes PO: Roles of Trm9- and ALKBH8-like proteins in the formation of modified wobble uridines in Arabidopsis tRNA. Nucleic Acids Res 2011, 39(17):7688-7701.
36. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: BLAST+: architecture and applications. BMC Bioinformatics 2009, 10:421.

37. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS et al: CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 2020, 48(D1):D265-D268.

38. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D: The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis 1993, 14(10):1023-1031.

39. Bjellqvist B, Basse B, Olsen E, Celis JE: Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis 1994, 15(3-4):529-539.

40. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF: Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999, 112:531-552.

41. Chou KC, Shen HB: Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. PLoS One 2010, 5(6):e11335.

42. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH: CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS One 2014, 9(6):e99368.

43. Sperschneider J, Catanzariti AM, DeBoer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM: LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep 2017, 7:44598.

44. Katoh K, Standley DM: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013, 30(4):772-780.

45. Wong TKF, Ly-Trong N, Ren H, Demotte P, Banos H, Roger AJ, Susko E, Bielow C, De Maio N, Goldman N et al: IQ-TREE 3: phylogenomic inference software using complex evolutionary models. Mol Biol Evol 2026, 43(5):msag117.
46. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS: ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 2017, 14(6):587-589.
47. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS: UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 2018, 35(2):518-522.
48. Huerta-Cepas J, Serra F, Bork P: ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol 2016, 33(6):1635-1638.
49. Letunic I, Bork P: Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res 2024, 52(W1):W78-W82.
50. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004, 32(5):1792-1797.
51. Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K: MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. Mol Biol Evol 2024, 41(12):msae263.
52. Buchfink B, Reuter K, Drost HG: Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 2021, 18(4):366-368.
53. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H et al: MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 2012, 40(7):e49.
54. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y et al: TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol

Plant 2023, 16(11):1733-1742.

55. Rombauts S, Dehais P, Van Montagu M, Rouze P: PlantCARE, a plant cis-acting regulatory element database. Nucleic Acids Res 1999, 27(1):295-296.

56. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 2002, 30(1):325-327.

57. Kawamura S, Romani F, Yagura M, Mochizuki T, Sakamoto M, Yamaoka S, Nishihama R, Nakamura Y, Yamato KT, Bowman JL et al: MarpolBase Expression: A Web-Based, Comprehensive Platform for Visualization and Analysis of Transcriptomes in the Liverwort Marchantia polymorpha. Plant Cell Physiol 2022, 63(11):1745-1755.

58. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C: Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017, 14(4):417-419.

59. Soneson C, Love MI, Robinson MD: Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 2015, 4:1521.

60. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014, 15(12):550.

61. Saint-Marcoux D, Proust H, Dolan L, Langdale JA: Identification of reference genes for real-time quantitative PCR experiments in the liverwort Marchantia polymorpha. PLoS One 2015, 10(3):e0118678.

62. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001, 25(4):402-408.

63. Takemura M, Hamada T, Kida H, Ohyama K: Cold-induced accumulation of omega-3 polyunsaturated fatty acid in a liverwort, Marchantia polymorpha L. Biosci Biotechnol

Biochem 2012, 76(4):785-790.

64. Ueda M, Takami T, Peng L, Ishizaki K, Kohchi T, Shikanai T, Nishimura Y: Subfunctionalization of sigma factors during the evolution of land plants based on mutant analysis of liverwort (Marchantia polymorpha L.) MpSIG1. Genome Biol Evol 2013, 5(10):1836-1848.

65. Nishihama R, Ishida S, Urawa H, Kamei Y, Kohchi T: Conditional Gene Expression/Deletion Systems for Marchantia polymorpha Using its Own Heat-Shock Promoter and Cre/loxP-Mediated Site-Specific Recombination. Plant Cell Physiol 2016, 57(2):271-280.

66. Tanaka H, Suzuki R, Okabe N, Suzuki T, Kodama Y: Salinity stress-responsive transcription factors in the liverwort Marchantia polymorpha. Plant Biotechnol (Tokyo) 2018, 35(3):281-284.

67. Tansley C, Houghton J, Rose AME, Witek B, Payet RD, Wu T, Miller JB: CIPK-B is essential for salt stress signalling in Marchantia polymorpha. New Phytol 2023, 237(6):2210-2223.

68. Kanazawa T, Ishizaki K, Kohchi T, Hanaoka M, Tanaka K: Characterization of four nuclear-encoded plastid RNA polymerase sigma factor genes in the liverwort Marchantia polymorpha: blue-light- and multiple stress-responsive SIG5 was acquired early in the emergence of terrestrial plants. Plant Cell Physiol 2013, 54(10):1736-1748.

69. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A et al: AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 2022, 50(D1):D439-D444.

70. Steinegger M, Soding J: MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 2017, 35(11):1026-1028.

71. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M: ColabFold: making protein folding accessible to all. Nat Methods 2022, 19(6):679-682.

72. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S et al: The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 2023, 51(D1):D638-D646.

73. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003, 13(11):2498-2504.

74. Tavanez JP, Madl T, Kooshapur H, Sattler M, Valcarcel J: hnRNP A1 proofreads 3' splice site recognition by U2AF. Mol Cell 2012, 45(3):314-329.

75. Francisco-Mangilet AG, Karlsson P, Kim MH, Eo HJ, Oh SA, Kim JH, Kulcheski FR, Park SK, Manavella PA: THO2, a core member of the THO/TREX complex, is required for microRNA production in Arabidopsis. Plant J 2015, 82(6):1018-1029.

76. Butt H, Bazin J, Alshareef S, Eid A, Benhamed M, Reddy ASN, Crespi M, Mahfouz MM: Overlapping roles of spliceosomal components SF3B1 and PHF5A in rice splicing regulation. Commun Biol 2021, 4(1):529.

Supplementary information

Additional file 1 (.xlsx): Supplementary Table S1. A. thaliana reference genes used for the identification of candidate m6A-related genes.

Gene names and Arabidopsis locus identifiers for the reference genes used to select protein queries for homology searches, grouped into writer-related, ALKBH-family, and reader-related categories.

Additional file 2 (.xlsx): Supplementary Table S2. Transcript abundance of candidate m^{6}A-related genes across tissues and developmental stages in *M. polymorpha*.

Replicate-level transcript abundance values, expressed as transcripts per million (TPM), for 18 candidate genes across 16 tissue and developmental-stage conditions. The dataset includes gene names and MarpolBase gene identifiers and was obtained from the MarpolBase Expression database (MBEX) for the analysis shown in Figure 6.

Additional file 3 (.pptx): Supplementary Figures 1–4. Phylogenetic analyses and validation of abiotic-stress responses.

Supplementary Figure 1. Composite display of maximum-likelihood phylogenies of candidate m^{6}A writer proteins across 122 plant species.

Supplementary Figure 2. Maximum-likelihood phylogeny of candidate ALKBH proteins across 122 plant species.

Supplementary Figure 3. Composite display of maximum-likelihood phylogenies of candidate

1519 m^6A reader-related proteins across 122 plant species.

1520 **Supplementary Figure 4.** qRT-PCR expression of abiotic-stress marker genes under the

1521 treatments used in this study.

1522

1523 **Additional file 4 (.xlsx): Supplementary Table S3. Raw qRT-PCR Cq values and associated**

1524 **quality metrics.**

1525 Individual-well Cq values and associated quality metrics for tissue comparisons, abiotic-stress

1526 response analyses, and stress-marker validation, corresponding to Figures 7 and 9 and

1527 Supplementary Figure 4. The nine worksheets are organized by experiment and biological replicate.

1528

1529 **Additional file 5 (.xlsx): Supplementary Table S4. Primer sequences used for qRT-PCR**

1530 **analysis in *M. polymorpha*.**

1531 Forward and reverse primer sequences for the 18 candidate m6A-related genes, nine abiotic-stress

1532 marker genes, and the internal reference gene *MpSAND*. Reference numbers identify the sources

1533 of previously published primer pairs.

Figure 1

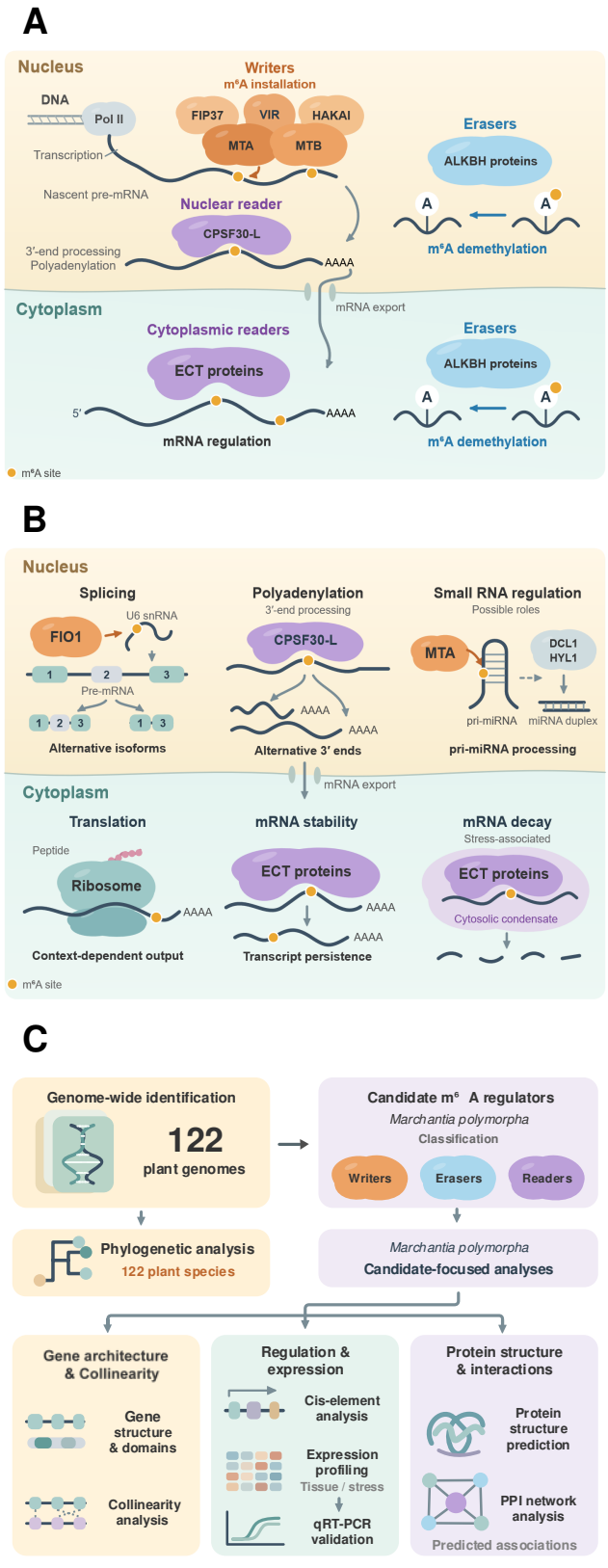

A
Nucleus
Writers
m⁶A installation
DNA
Pol II
FIP37
VIR
HAKAI
MTA
MTB
Transcription
Nascent pre-mRNA
Nuclear reader
CPSF30-L
3′-end processing
Polyadenylation
AAAA
Erasers
ALKBH proteins
m⁶A demethylation
mRNA export
Cytoplasm
Cytoplasmic readers
ECT proteins
5′
mRNA regulation
Erasers
ALKBH proteins
m⁶A demethylation
m⁶A site
B
Nucleus
Splicing
U6 snRNA
FIO1
Pre-mRNA
Alternative isoforms
Polyadenylation
3′-end processing
CPSF30-L
Alternative 3′ ends
Small RNA regulation
Possible roles
MTA
DCL1
HYL1
pri-miRNA
miRNA duplex
pri-miRNA processing
mRNA export
Cytoplasm
Translation
Peptide
Ribosome
Context-dependent output
mRNA stability
ECT proteins
Transcript persistence
mRNA decay
Stress-associated
ECT proteins
Cytosolic condensate
m⁶A site
C
Genome-wide identification
122
plant genomes
Candidate m⁶A regulators
Marchantia polymorpha
Classification
Writers
Erasers
Readers
Phylogenetic analysis
122 plant species
Marchantia polymorpha
Candidate-focused analyses
Gene architecture & Collinearity
Gene structure & domains
Collinearity analysis
Regulation & expression
Cis-element analysis
Expression profiling
Tissue / stress
qRT-PCR validation
Protein structure & interactions
Protein structure prediction
PPI network analysis
Predicted associations

Figure 2

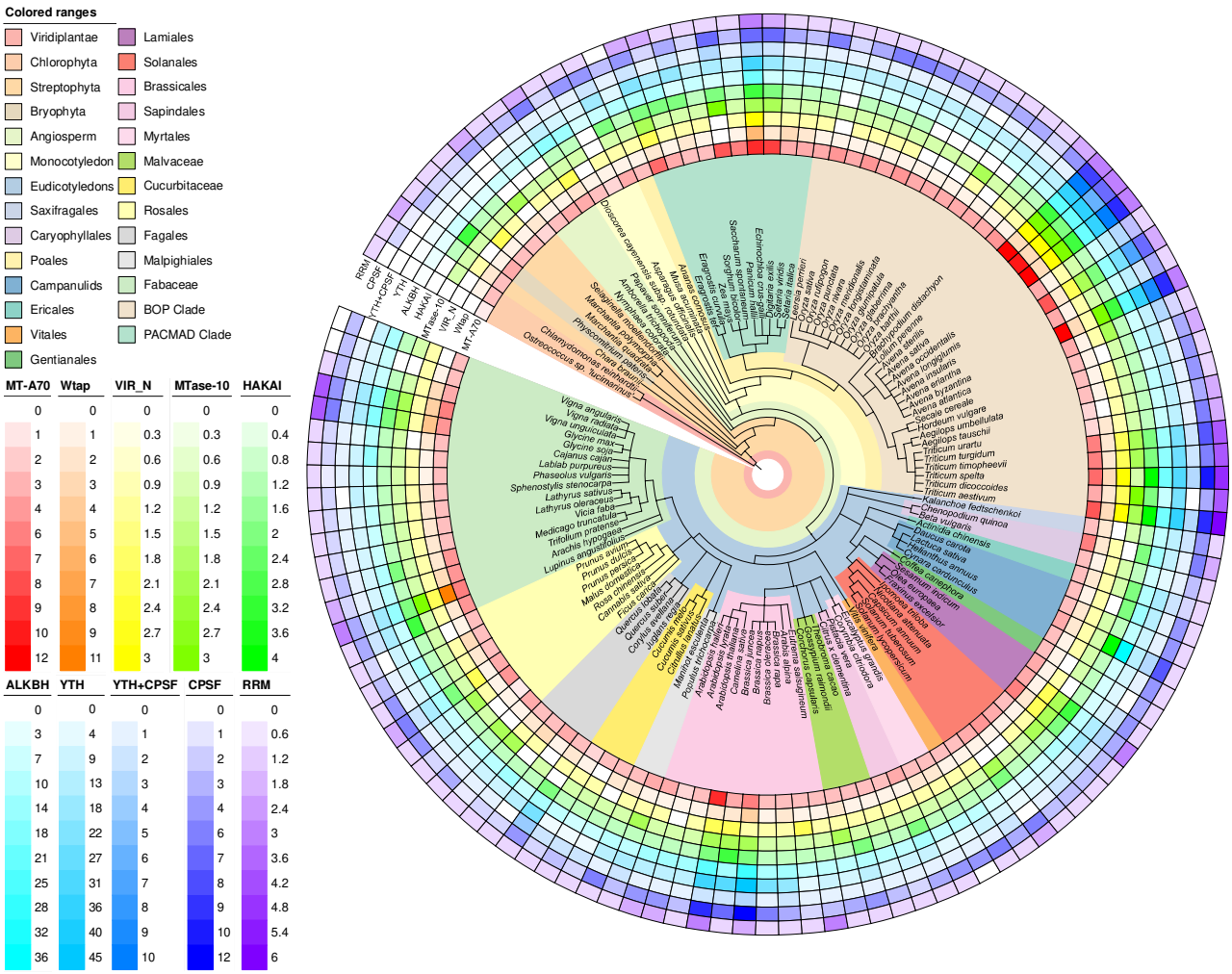

Colored ranges
Viridiplantae
Chlorophyta
Streptophyta
Bryophyta
Angiosperm
Monocotyledon
Eudicotyledons
Saxifragales
Caryophyllales
Poales
Campanulids
Ericales
Vitales
Gentianales
Lamiales
Solanales
Brassicales
Sapindales
Myrtales
Malvaceae
Cucurbitaceae
Rosales
Fagales
Malpighiales
Fabaceae
BOP Clade
PACMAD Clade
MT-A70
Wtap
VIR_N
MTase-10
HAKAI
ALKBH
YTH
YTH+CPSF
CPSF
RRM

Figure 3

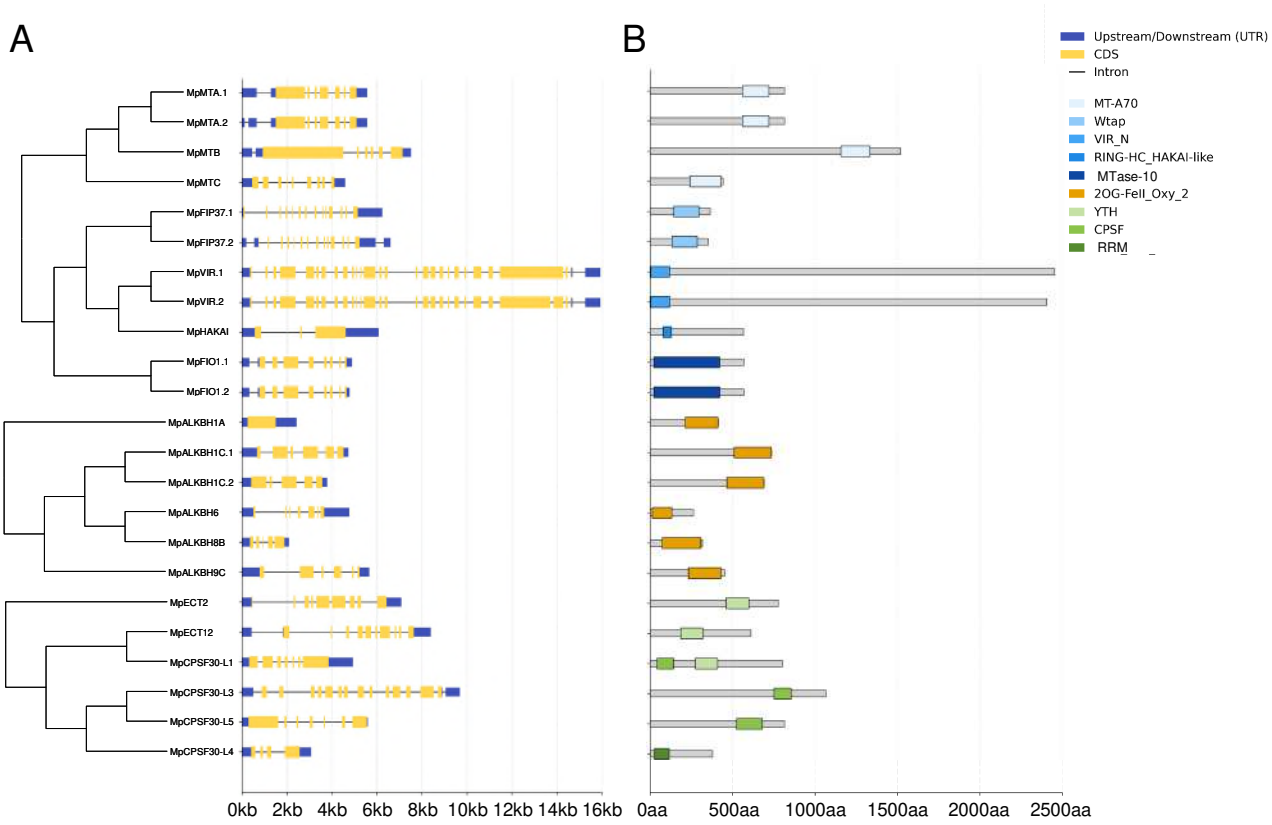

A
B
MpMTA.1
MpMTA.2
MpMTB
MpMTC
MpFIP37.1
MpFIP37.2
MpVIR.1
MpVIR.2
MpHAKAI
MpFIO1.1
MpFIO1.2
MpALKBH1A
MpALKBH1C.1
MpALKBH1C.2
MpALKBH6
MpALKBH8B
MpALKBH9C
MpECT2
MpECT12
MpCPSF30-L1
MpCPSF30-L3
MpCPSF30-L5
MpCPSF30-L4
0kb 2kb 4kb 6kb 8kb 10kb 12kb 14kb 16kb
0aa 500aa 1000aa 1500aa 2000aa 2500aa
Upstream/Downstream (UTR)
CDS
Intron
MT-A70
Wtap
VIR_N
RING-HC_HAKAI-like
MTase-10
2OG-FeII_Oxy_2
YTH
CPSF
RRM

Figure 4

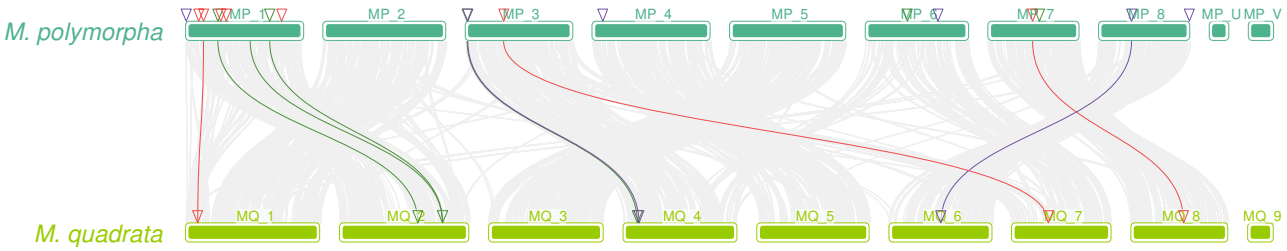

M. polymorpha
MP_1
MP_2
MP_3
MP_4
MP_5
MP_6
MP_7
MP_8
MP_U
MP_V
M. quadrata
MQ_1
MQ_2
MQ_3
MQ_4
MQ_5
MQ_6
MQ_7
MQ_8
MQ_9

Figure 5

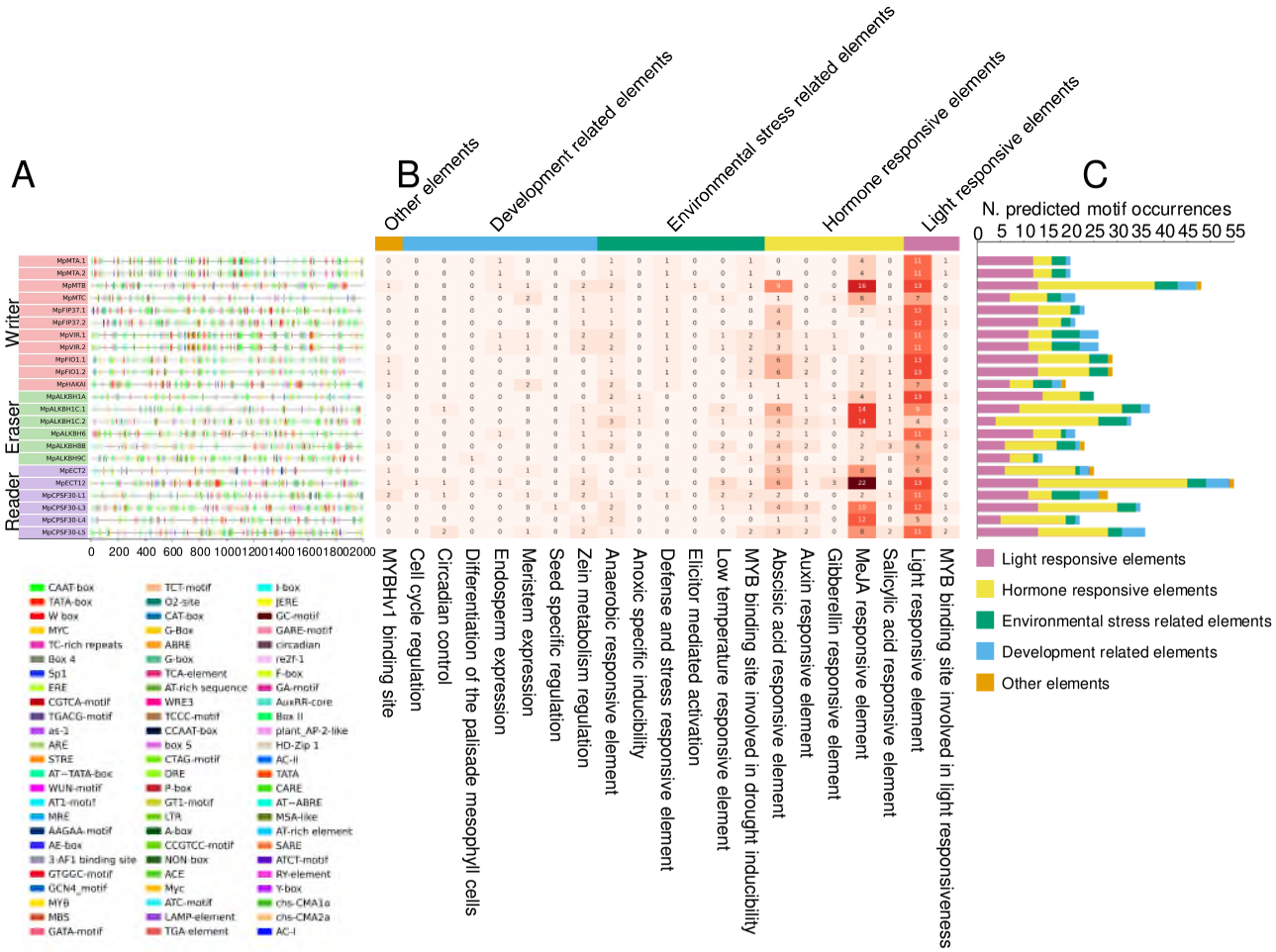

A
B
C
Writer
Eraser
Reader
MpMTA.1
MpMTA.2
MpMTB
MpMTC
MpFIP37.1
MpFIP37.2
MpVIR.1
MpVIR.2
MpFIO1.1
MpFIO1.2
MpHAKAI
MpALKBH1A
MpALKBH1C.1
MpALKBH1C.2
MpALKBH6
MpALKBH8B
MpALKBH9C
MpECT2
MpECT12
MpCPSF30-L1
MpCPSF30-L3
MpCPSF30-L4
MpCPSF30-L5
0 200 400 600 800 1000 1200 1400 1600 1800 2000
CAAT-box
TATA-box
W box
MYC
TC-rich repeats
Box 4
Sp1
ERE
CGTCA-motif
TGACG-motif
as-1
ARE
STRE
AT~TATA-box
WUN-motif
AT1-motif
MRE
AAGAA-motif
AE-box
3-AF1 binding site
GTGGC-motif
GCN4_motif
MYB
MBS
GATA-motif
TCT-motif
O2-site
CAT-box
G-Box
ABRE
G-box
TCA-element
AT-rich sequence
WRE3
TCCC-motif
CCAAT-box
box 5
CTAG-motif
ORE
P-box
GT1-motif
LTR
A-box
CCGTCC-motif
NON-box
ACE
Myc
ATC-motif
LAMP-element
TGA-element
I-box
JERE
GC-motif
GARE-motif
circadian
re2f-1
F-box
GA-motif
AuxRR-core
Box II
plant_AP-2-like
HD-Zip 1
AC-II
TATA
CARE
AT~ABRE
MSA-like
AT-rich element
SARE
ATCT-motif
RY-element
Y-box
chs-CMA1a
chs-CMA2a
AC-I
Other elements
Development related elements
Environmental stress related elements
Hormone responsive elements
Light responsive elements
MYBHv1 binding site
Cell cycle regulation
Circadian control
Differentiation of the palisade mesophyll cells
Endosperm expression
Meristem expression
Seed specific regulation
Zein metabolism regulation
Anaerobic responsive element
Anoxic specific inducibility
Defense and stress responsive element
Elicitor mediated activation
Low temperature responsive element
MYB binding site involved in drought inducibility
Abscisic acid responsive element
Auxin responsive element
Gibberellin responsive element
MeJA responsive element
Salicylic acid responsive element
Light responsive element
MYB binding site involved in light responsiveness
N. predicted motif occurrences
0 5 10 15 20 25 30 35 40 45 50 55
Light responsive elements
Hormone responsive elements
Environmental stress related elements
Development related elements
Other elements

Figure 6

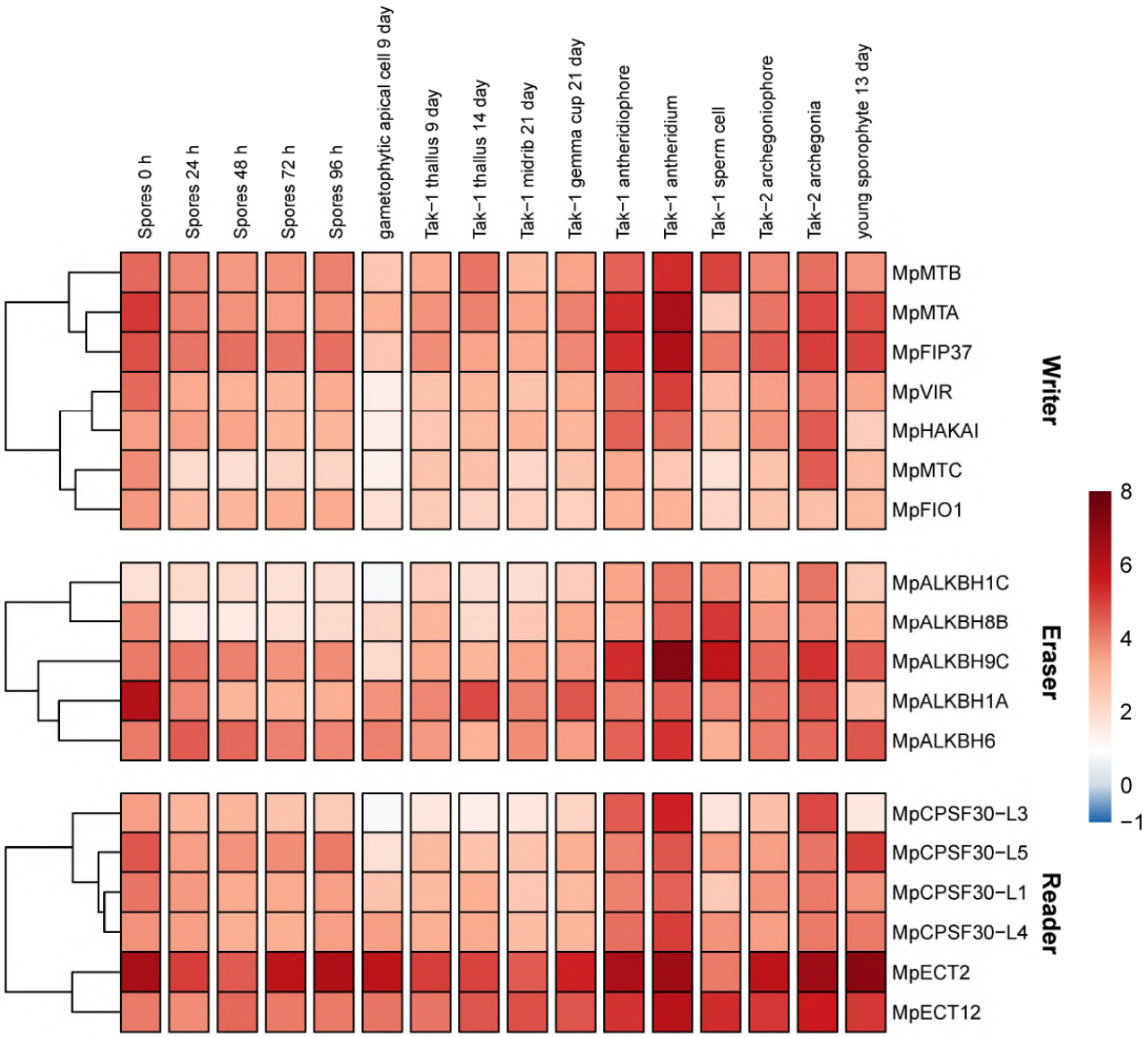

Spores 0 h
Spores 24 h
Spores 48 h
Spores 72 h
Spores 96 h
gametophytic apical cell 9 day
Tak−1 thallus 9 day
Tak−1 thallus 14 day
Tak−1 midrib 21 day
Tak−1 gemma cup 21 day
Tak−1 antheridiophore
Tak−1 antheridium
Tak−1 sperm cell
Tak−2 archegoniophore
Tak−2 archegonia
young sporophyte 13 day
MpMTB
MpMTA
MpFIP37
MpVIR
MpHAKAI
MpMTC
MpFIO1
Writer
MpALKBH1C
MpALKBH8B
MpALKBH9C
MpALKBH1A
MpALKBH6
Eraser
MpCPSF30−L3
MpCPSF30−L5
MpCPSF30−L1
MpCPSF30−L4
MpECT2
MpECT12
Reader
8
6
4
2
0
−1

Figure 7

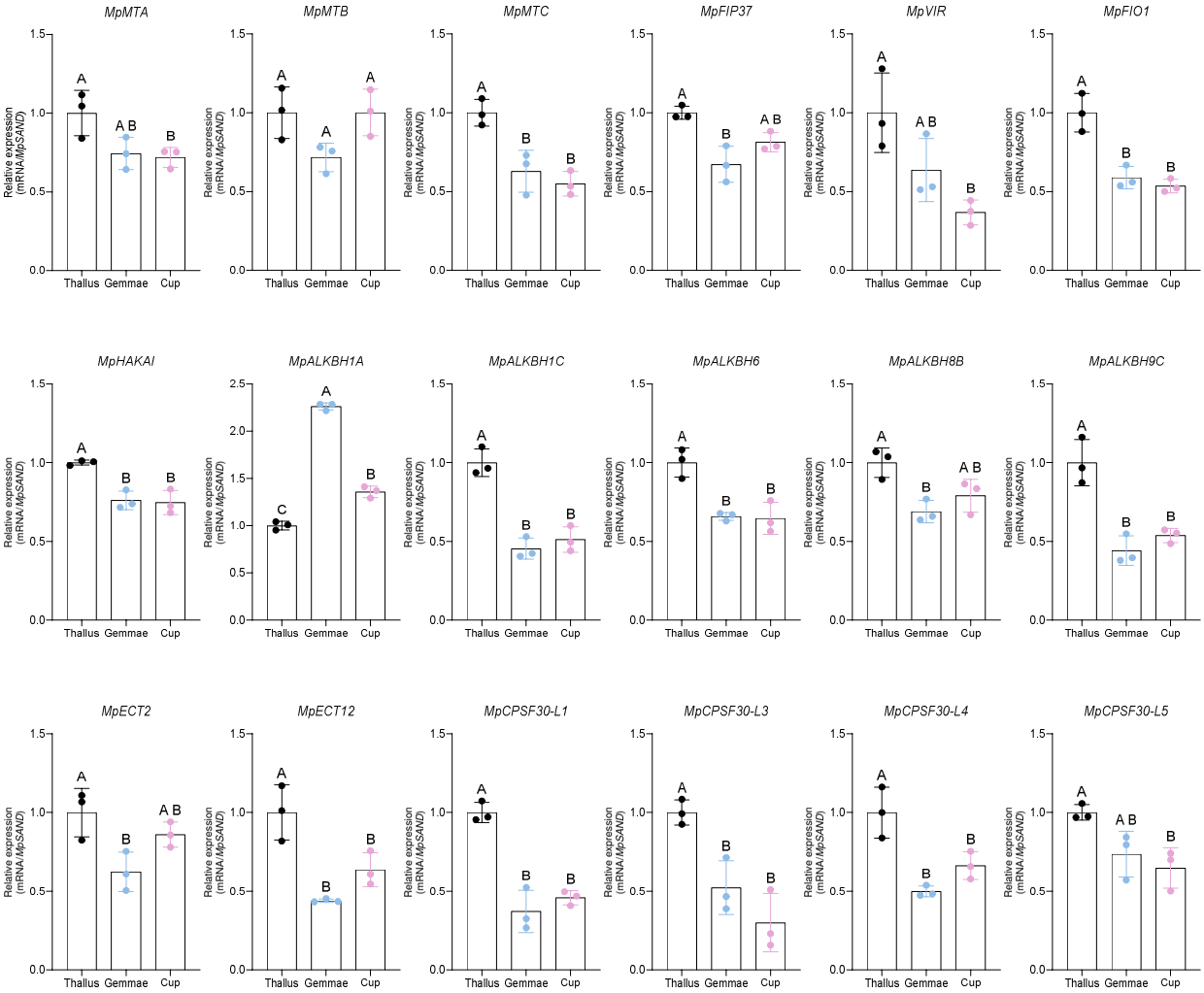

MpMTA
MpMTB
MpMTC
MpFIP37
MpVIR
MpFIO1
MpHAKAI
MpALKBH1A
MpALKBH1C
MpALKBH6
MpALKBH8B
MpALKBH9C
MpECT2
MpECT12
MpCPSF30-L1
MpCPSF30-L3
MpCPSF30-L4
MpCPSF30-L5
Relative expression (mRNA/MpSAND)
Thallus
Gemmae
Cup

Figure 8

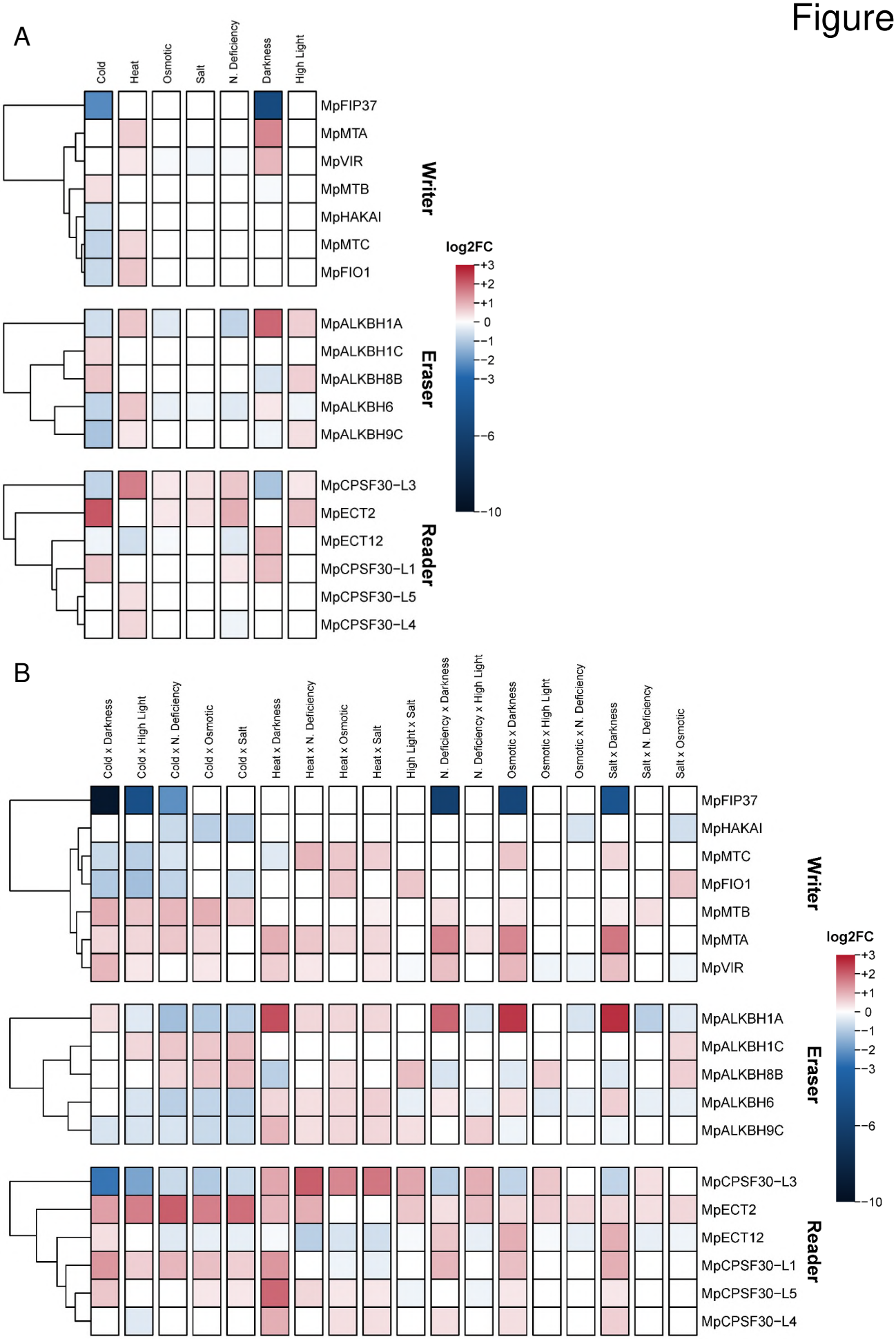

A
Cold
Heat
Osmotic
Salt
N. Deficiency
Darkness
High Light
MpFIP37
MpMTA
MpVIR
MpMTB
MpHAKAI
MpMTC
MpFIO1
Writer
MpALKBH1A
MpALKBH1C
MpALKBH8B
MpALKBH6
MpALKBH9C
Eraser
MpCPSF30−L3
MpECT2
MpECT12
MpCPSF30−L1
MpCPSF30−L5
MpCPSF30−L4
Reader
log2FC
+3
+2
+1
0
−1
−2
−3
−6
−10
B
Cold x Darkness
Cold x High Light
Cold x N. Deficiency
Cold x Osmotic
Cold x Salt
Heat x Darkness
Heat x N. Deficiency
Heat x Osmotic
Heat x Salt
High Light x Salt
N. Deficiency x Darkness
N. Deficiency x High Light
Osmotic x Darkness
Osmotic x High Light
Osmotic x N. Deficiency
Salt x Darkness
Salt x N. Deficiency
Salt x Osmotic
MpFIP37
MpHAKAI
MpMTC
MpFIO1
MpMTB
MpMTA
MpVIR
Writer
MpALKBH1A
MpALKBH1C
MpALKBH8B
MpALKBH6
MpALKBH9C
Eraser
MpCPSF30−L3
MpECT2
MpECT12
MpCPSF30−L1
MpCPSF30−L5
MpCPSF30−L4
Reader
log2FC
+3
+2
+1
0
−1
−2
−3
−6
−10

Figure 9

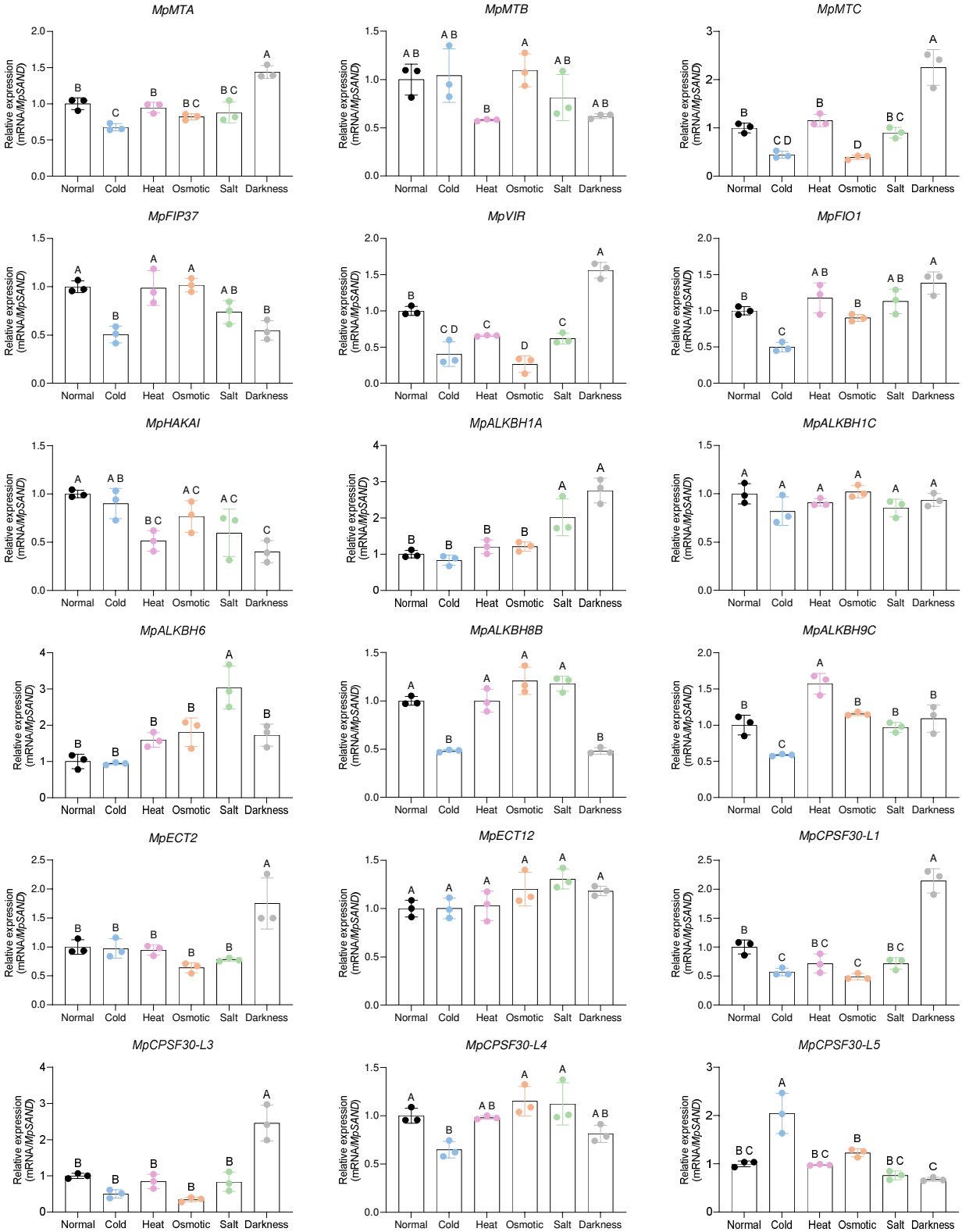

MpMTA
MpMTB
MpMTC
MpFIP37
MpVIR
MpFIO1
MpHAKAI
MpALKBH1A
MpALKBH1C
MpALKBH6
MpALKBH8B
MpALKBH9C
MpECT2
MpECT12
MpCPSF30-L1
MpCPSF30-L3
MpCPSF30-L4
MpCPSF30-L5
Relative expression (mRNA/MpSAND)
Normal
Cold
Heat
Osmotic
Salt
Darkness

Figure 10

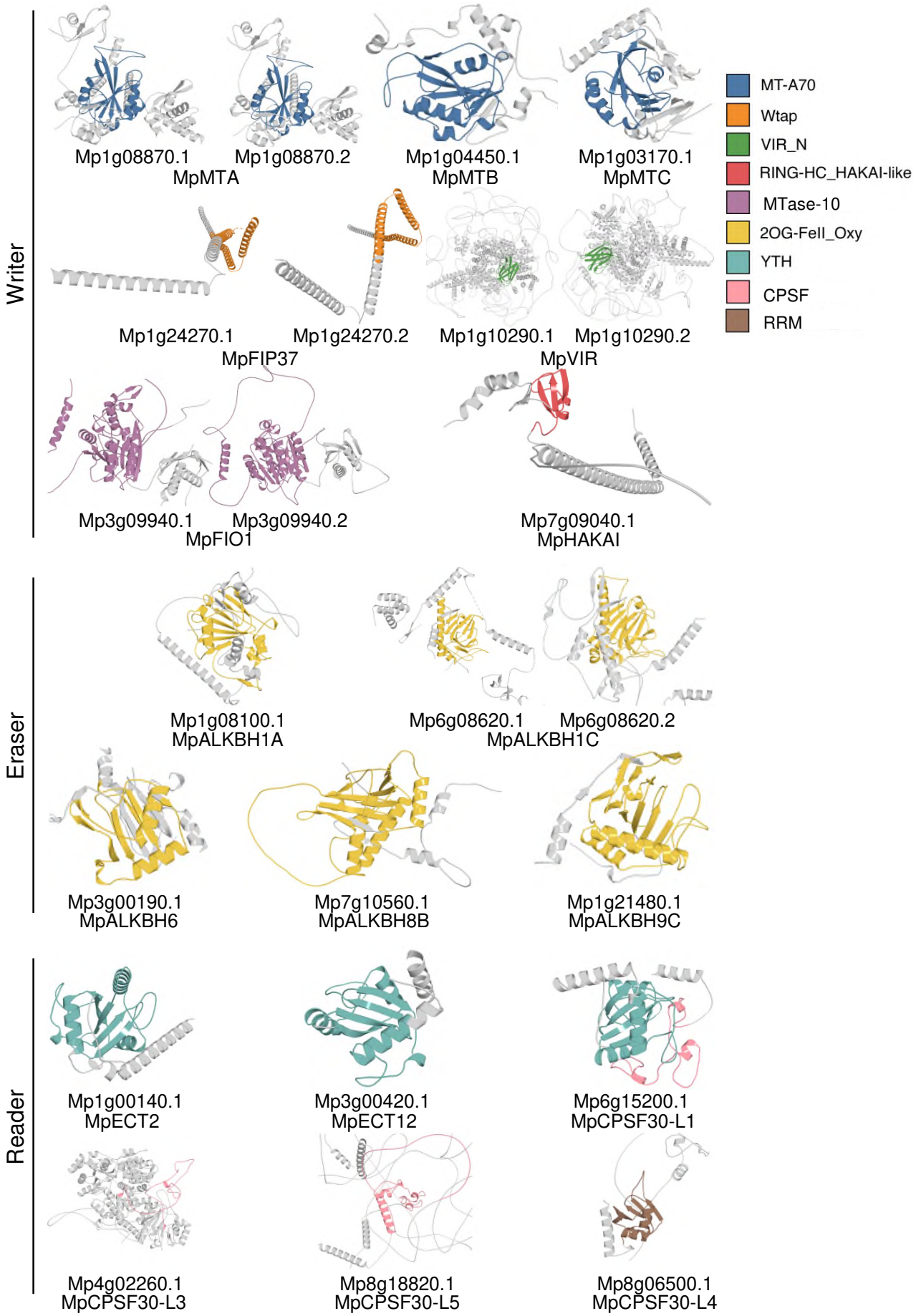
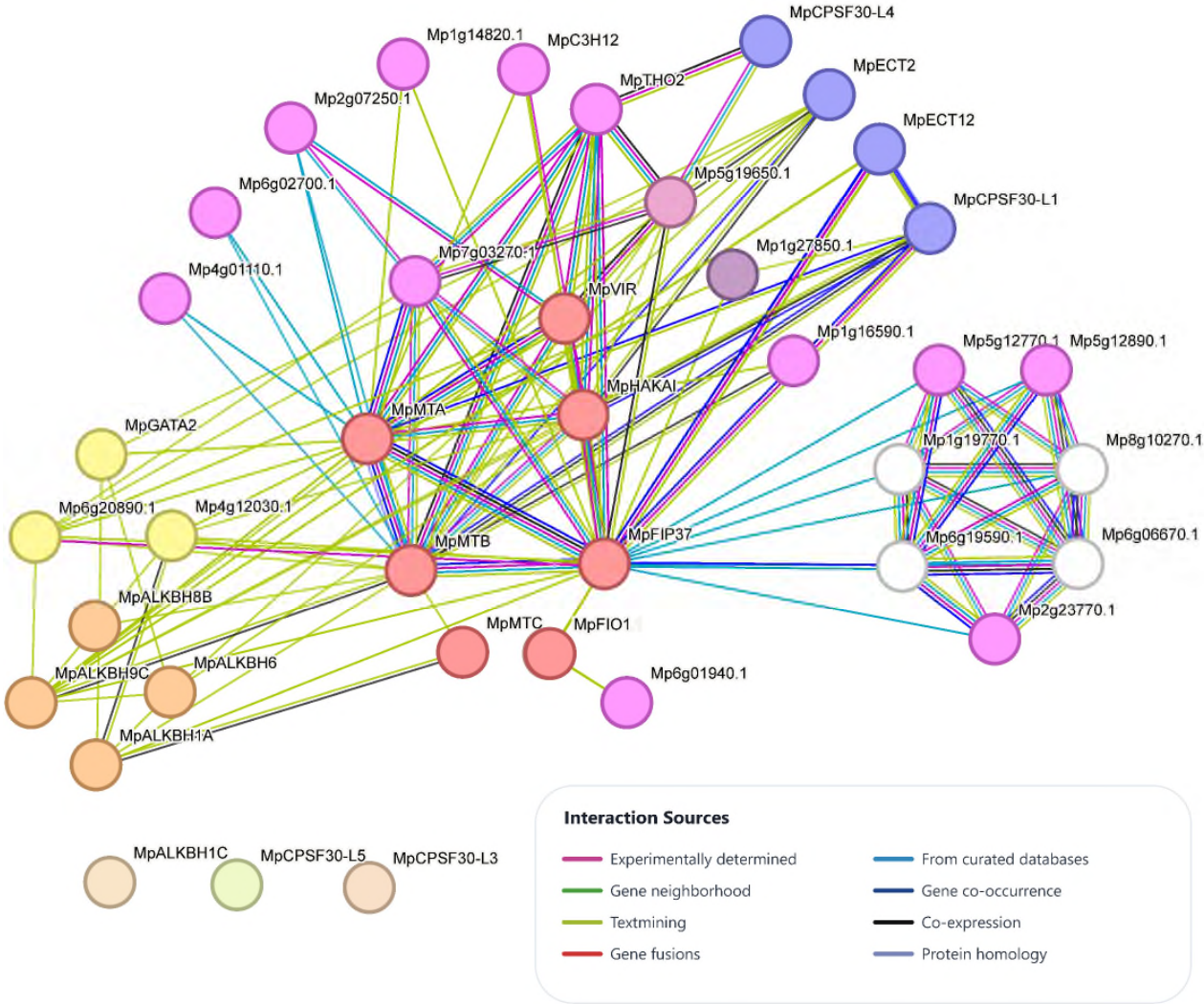

Writer
Eraser
Reader
MT-A70
Wtap
VIR_N
RING-HC_HAKAI-like
MTase-10
2OG-FeII_Oxy
YTH
CPSF
RRM
Mp1g08870.1
Mp1g08870.2
MpMTA
Mp1g04450.1
MpMTB
Mp1g03170.1
MpMTC
Mp1g24270.1
Mp1g24270.2
MpFIP37
Mp1g10290.1
Mp1g10290.2
MpVIR
Mp3g09940.1
Mp3g09940.2
MpFIO1
Mp7g09040.1
MpHAKAI
Mp1g08100.1
MpALKBH1A
Mp6g08620.1
Mp6g08620.2
MpALKBH1C
Mp3g00190.1
MpALKBH6
Mp7g10560.1
MpALKBH8B
Mp1g21480.1
MpALKBH9C
Mp1g00140.1
MpECT2
Mp3g00420.1
MpECT12
Mp6g15200.1
MpCPSF30-L1
Mp4g02260.1
MpCPSF30-L3
Mp8g18820.1
MpCPSF30-L5
Mp8g06500.1
MpCPSF30-L4

Figure 11


MpCPSF30-L4
Mp1g14820.1
MpC3H12
MpTHO2
MpECT2
Mp2g07250.1
MpECT12
Mp6g02700.1
Mp5g19650.1
MpCPSF30-L1
Mp7g03270.1
Mp1g27850.1
Mp4g01110.1
MpVIR
Mp1g16590.1
Mp5g12770.1
Mp5g12890.1
MpHAKAI
MpMTA
MpGATA2
Mp1g19770.1
Mp8g10270.1
Mp6g20890.1
Mp4g12030.1
MpMTB
MpFIP37
Mp6g19590.1
Mp6g06670.1
MpALKBH8B
MpMTC
MpFIO1
Mp2g23770.1
MpALKBH9C
MpALKBH6
Mp6g01940.1
MpALKBH1A
MpALKBH1C
MpCPSF30-L5
MpCPSF30-L3
Interaction Sources
Experimentally determined
From curated databases
Gene neighborhood
Gene co-occurrence
Textmining
Co-expression
Gene fusions
Protein homology

Supplementary Files

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

- Table1.pptx
- Table2.pptx
- SupplementaryTableS1.xlsx
- SupplementaryTableS2.xlsx
- SupplementaryFigures14.pptx
- SupplementaryTableS3.xlsx
- SupplementaryTableS4.xlsx