OXFORD

PRMD: an integrated database for plant RNA modifications

Xiaoqiang Lang[2,3,†], Chunyan Yu[4,†], Mengyuan Shen [1,†], Lei Gu[5], Qian Qian [1], Degui Zhou[1], Jiantao Tan[1], Yiliang Li[6], Xin Peng[1], Shu Diao[7], Zhujun Deng[8], Zhaohui Ruan[9], Zhi Xu[10], Junlian Xing[1], Chen Li [1], Runfeng Wang[11], Changjun Ding[2,*], Yi Cao[3,*] and Qi Liu [1,*]

[1]Rice Research Institute, Guangdong Academy of Agricultural Sciences, Key Laboratory of Genetics and Breeding of High Quality Rice in Southern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Guangdong Key Laboratory of New Technology in Rice Breeding, Guangdong Rice Engineering Laboratory, Guangzhou, 510640, China
[2]State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China
[3]Microbiology and Metabolic Engineering Key Laboratory of Sichuan Province, College of Life Science, Sichuan University, Chengdu, Sichuan, 610041, China
[4]Frontiers Science Center for Disease-related Molecular Network, Laboratory of Omics Technology and Bioinformatics, West China Hospital, Sichuan University, Chengdu, Sichuan, 610041, China
[5]Epigenetics Laboratory, Max Planck Institute for Heart and Lung Research & Cardiopulmonary Institute (CPI). Parkstr.1 61231 Bad Nauheim Germany
[6]Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization/Guangdong Academy of Forestry, Guangzhou, Guangdong 510520, China
[7]Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou, China
[8]Precision Medicine Center, Precision Medicine Key Laboratory of Sichuan Province, West China Hospital, Sichuan University, Chengdu, Sichuan, 610041, China
[9]Sun Yat-sen University Cancer Center, State Key Laboratory Oncology in South China, Collaborative Innovation Center of Cancer Medicine, 510060, Guangzhou, China
[10]Guangxi Key Laboratory of Images and Graphics Intelligent Processing, Guilin University of Electronics Technology, Guilin, 541004, China
[11]Guangdong Provincial Key Laboratory of Crop Genetic Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China

*To whom correspondence should be addressed. Tel: +86 020 34275794; Email: changjunding@caf.ac.cn
Correspondence may also be addressed to Yi Cao. Email: geneium@scu.edu.cn
Correspondence may also be addressed to Qi Liu. Email: rnainfor@gmail.com
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

Abstract

The scope and function of RNA modifications in model plant systems have been extensively studied, resulting in the identification of an increasing number of novel RNA modifications in recent years. Researchers have gradually revealed that RNA modifications, especially N^6-methyladenosine (m^6A), which is one of the most abundant and commonly studied RNA modifications in plants, have important roles in physiological and patho-logical processes. These modifications alter the structure of RNA, which affects its molecular complementarity and binding to specific proteins, thereby resulting in various of physiological effects. The increasing interest in plant RNA modifications has necessitated research into RNA modifications and associated datasets. However, there is a lack of a convenient and integrated database with comprehensive annotations and intuitive visualization of plant RNA modifications. Here, we developed the Plant RNA Modification Database (PRMD; http://bioinformatics.sc.cn/PRMD and http://rnainformatics.org.cn/PRMD) to facilitate RNA modification research. This database contains information regarding 20 plant species and provides an intuitive interface for displaying information. Moreover, PRMD offers multiple tools, including RMlevelDiff, RMplantVar, RNAmodNet and Blast (for functional analyses), and mRNAbrowse, RNAlollipop, JBrowse and Integrative Genomics Viewer (for displaying data). Furthermore, PRMD is freely available, making it useful for the rapid development and promotion of research on plant RNA modifications.

Graphical abstract
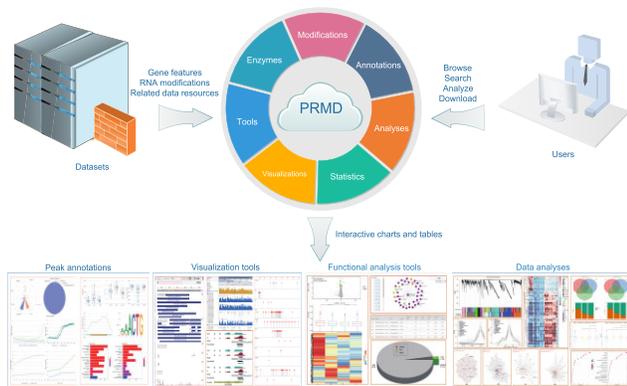
Introduction

RNA modifications provide diverse RNA molecules with an additional layer of information during epitranscriptome-mediated post-transcriptional regulation (1). More than 160 types of RNA modifications have been identified, including N^1-methyladenosine (m^1A), C^5-methylcytosine (m^5C), N^6-methyladenosine (m^6A), N^7-methylguanosine (m^7G), N^4-acetylcytidine (ac^4C), 2′-O-Me and pseudouridine, which can regulate the RNA secondary structure, expression, splicing, stability and translation (2–4). Specifically, m^6A, which is one of the most prevalent internal modifications of diverse RNA molecules, has been detected in messenger RNAs (mRNAs) and accounts for approximately 50% of the methylated nucleotides in RNA (5–7). Numerous studies have shown that RNA modifications are important for gene regulation and affect various aspects of plant development, stress responses and evolution (8,9). For example, the inactivation of a key *Arabidopsis thaliana* m^6A methyltransferase (mRNA adenosine methylase) can lead to embryo lethality (10). In rice, a lack of the m^5C methyltransferase OsNSUN2 increases susceptibility to heat stress (11). The evolution of m^6A is at least in part related to genome replication events in complex polyploid plant genomes (12).

To date, many studies have developed new sequencing methods to identify different RNA modifications (e.g. m^6A-seq, m^5C-RIP-seq and m^1A-seq) (13,14). For example, the methylated RNA immunoprecipitation sequencing (MeRIP-seq) analysis, which combines immunoprecipitation and next-generation sequencing, has been widely used to quantitatively explore RNA modifications throughout the transcriptome. Using these methods, substantial amounts of sequencing data have been obtained for a variety of species, including mammals, yeast and plants. Several databases and web servers, such as CVm6A (15), SRAMP (16) and WHISTLE (17), have been constructed to integrate the existing sequencing data for mammals. In contrast, there are only a few web servers, such as AthMethPre (18) and RFAthM6A (19), for predicted RNA modifications in plants. There are other databases that include RNA modifications from a few plant species. For example, RMBase v2.0 (20), MeT-DB v2.0 (21) and REPIC (22) contain information for only one plant species (*A. thaliana*), whereas m6A-Atlas 2.0 (23) and ENCORE (https://rna.sysu.edu.cn/encore/) provide details regarding 6 and 10 plant species, respectively. Additionally, RNAmod (24) consists of annotated RNA modifications from 21 species, but

only a few of these species are plants (24). Both PEA, which is an integrated R toolkit for plant epitranscriptome analyses, and its updated version deepEA, which is a containerized web-based platform for interactive analyses of epitranscriptome sequencing data, focus exclusively on m^6A peak analyses and annotations, with no information regarding other RNA modifications (25,26). Because of the considerable accumulation of datasets for RNA modifications and related annotation datasets in different plant tissues at various developmental stages and under diverse stress conditions, there is an urgent need for a comprehensive database that integrates these large-scale datasets and their extensive annotations.

We herein present the Plant RNA Modification Database (PRMD), a comprehensive plant RNA modification database that integrates the RNA modifications from up to 20 diverse plant species. As well as the m^6A modification, we collected other diverse known RNA modifications (m^1A, m^5C, m^7G, ac^4C, 2′-O-Me and pseudouridine) and related datasets (e.g. RNA secondary structures and RBP binding sites) from published research articles and related plant resources. Additionally, PRMD provides various useful tools for functional analyses and data visualization, such as RMlevelDiff, RMplantVar, RNAmodNet and Blast (for data analyses) as well as mRNAbrowse, RNAlollipop, JBrowse and Integrative Genomics Viewer (IGV) (for intuitively visualizing the data in PRMD). Using PRMD, we determined that RNA modifications are common in plants. Moreover, the genes encoding transcripts with the m^6A modification were under purifying selection during evolution. Details regarding the significant effects of m^6A modifications on alternative polyadenylation (APA), exon usage, gene expression and translational efficiency (TE) are available in PRMD. In summary, PRMD is an integrated and intuitive database developed specifically for research on RNA modifications in plants, making it relevant for the rapidly developing plant epitranscriptomics-related research fields.

Materials and methods

Data sources and reference genomes

Reference genome sequences and gene annotation files of plant species were obtained from the EnsemblPlants database (27), with the exception of the reference genomes of five species, namely *Gossypium arboretum, Gossypium hirsutum, Fragaria vesca, Nicotiana benthamiana* and *Paulownia*

fortunei, which were retrieved from the Cotton Functional Genomics Database (CottonFGD) (28), Genome Database for Rosaceae (GDR) (29), Phytozome (30), Solanaceae Genomics Network (SGN) (31) and National Center for Biotechnology Information (NCBI: https://www.ncbi.nlm.nih.gov/), respectively. For the species that have only GFF3 format gene annotation files, the Gffread program (32) was used to convert the GFF format to the GTF format, which was used for annotating RNA modification. The methylated RNA immunoprecipitation sequencing (MeRIP-seq) datasets were collected from the Sequence Read Archive (SRA) database of NCBI (https://www.ncbi.nlm.nih.gov/sra) and the Genome Sequence Archive (GSA) database (33) of the National Genomics Data Center (NGDC) (34) (Supplementary Table S1).

Other types of known RNA modifications, such as m^1A, m^5C, m^7G, ac^4C, 2′-O-Me and pseudouridine, were collected from published research articles and other RNA modification resources (Supplementary Tables S2 and S3). The modification sites in nanopore sequencing datasets were obtained from the DirectRMDB database (35), whereas the nucleotide-resolution modification sites in miCLIP datasets were obtained from the m^6A-Altas2.0 database. We also integrated additional related datasets, including expression quantitative trait locus (eQTL) datasets from the AtMAD (36) and Rice-eQTL (http://riceqtl.ncpgr.cn/) databases, genome-wide association study (GWAS) datasets from the GWAS Atlas database (37), RNA G-quadruplex (rG4) structural information datasets from the G4Atlas database (38), RBP binding datasets from the POSTAR3 database (39), RNA loop information datasets from the R-loopAtlas database (40), transcriptome-scale RNA secondary structure probing datasets from the RNA Atlas of Structure Probing (RASP) database (41), APA site datasets from the PlantAPAdb database (42), small open reading frame (sORF) datasets from the PsORF database (http://psorf.whu.edu.cn), conservation datasets from the PlantRegMap database (43) and known RNA modification-related enzyme datasets from the Modomics database (44).

Plant material, RNA extraction and sequencing for wild rice and cultivated rice

The MeRIP-seq analysis was performed for *Oryza rufipogon* (DXW81; common wild rice) and the following two cultivated rice subspecies: *Oryza sativa* ssp. *indica* (WSSM) and *Oryza sativa* ssp. *japonica* (ZH11).

1) Plant material and growth conditions: DXW81, WSSM and ZH11 were grown in the experimental field of the Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China. Young leaves at the tillering stage and panicles at the booting stage were collected for the MeRIP-seq analysis. All tissues were immediately frozen in liquid nitrogen and stored at −80°C prior to the RNA extraction.

2) RNA extraction and sequencing: Total RNA was isolated and purified using the TRIzol reagent (Invitrogen, CA, USA) following the manufacturer's procedure. The RNA quantity and purity were determined using the NanoDrop ND-1000 spectrophotometer (NanoDrop, USA), whereas RNA integrity (RNA integrity number > 7.0) was assessed using the Bioanalyzer 2100 system (Agilent, CA, USA) and confirmed by denaturing agarose gel electrophoresis. An A-base was added

to the blunt ends of each strand to facilitate the ligation to the index adapters, which had a T-base overhang. Single- or dual-index adapters were ligated to the fragments, after which a size selection step was completed using AMPureXP beads. After the heat-labile UDG enzyme (cat.m0280; NEB, USA) treatment of the U-labeled double-stranded DNA fragments, the ligated products were amplified by PCR under the following conditions: initial denaturation at 95°C for 3 min; eight cycles of denaturation at 98°C for 15 s, annealing at 60°C for 15 s and extension at 72°C for 30 s; and then final extension at 72°C for 5 min. The average insert size for the final cDNA library was 300 ± 50 bp. Finally, we performed a paired-end sequencing analysis (2 × 150 bp) using the Illumina NovaSeq™ 6000 system (LC-Bio Technology Co., Ltd, Hangzhou, China).

Data preprocessing

The MeRIP-seq datasets in the SRA format were converted to the FASTQ format using sratoolkit v2.11.0 (https://github.com/ncbi/sra-tools) and then the default parameters of fastp v0.20.0 (45) were used to trim the adapter sequences and low-quality bases. After filtering the data for quality, STAR 2.7.4a (46) was used to map clean reads to the corresponding reference genomes. The parameter settings were as follows: –outSAMtype BAM SortedByCoordinate –outFilterMultimapNmax 1 –outFilterMismatchNmax 2 –runThreadN 16 – readFilesCommand zcat –outFileNamePrefix.

Peak calling and m^6A level calculation

Using the BAM format genome mapping files of the IP and the input samples obtained as described above, we applied two peak calling strategies. The following parameters of MACS2 (47) were used to identify methylation peaks: –nomodel –extsize 150 -B -n -q 0.05. The R package exomePeak2 (48) (https://github.com/ZW-xjtlu/exomePeak2) was used for peak calling, which involved the same BAM files and GTF files. The parameters were set as follows: fragment_length = 100 binding_length = 25 step_length = 25 pc_count_cutoff = 5 bg_count_cutoff = 50 p_cutoff = 1e-5 peak_calling_mode = exon. The m^6A level for each peak was calculated on the basis of RNASeQC v2.4.2 (49), bedtools v2.17.0 (50) and our in-house pipelines as:

$$m^6A\ level = \frac{\text{Peak reads (IP)} \times \text{Total reads (Input)}}{\text{Peak reads (Input)} \times \text{Total reads (IP)}}$$

Prediction of m^6A sites

The cDNA and ncRNA sequences of the following plant species were downloaded from EnsemblPlants: Aegilops tauschii, A. thaliana, Brassica rapa, Glycine max, Malus domestica, O. sativa, Physcomitrium patens, Populus trichocarpa, Phaseolus vulgaris, Sorghum bicolor, Solanum lycopersicum, Triticum aestivum, Triticum dicoccoides and Zea mays. The cDNA and ncRNA sequences of the following five species were obtained from the databases mentioned above: *F. vesca, G. arboreum, G. hirsutum, N. benthamiana* and *P. fortunei.* The SRAMP (16) software was used to predict all potential m^6A sites.

Identification of orthologous genes among plant species and analysis of evolution

To analyze orthologous groups (OGs), the longest protein was identified by examining the protein-coding sequences. The OrthoFinder v2.4.0 software (51) was used to construct OGs among 20 plant species on the basis of the longest protein. We also investigated the evolutionary patterns of m^6A methylation divergence. The ParaAT v2.0 software (52) was used for the sequence alignment analysis involving the longest protein sequences, after which KaKs_calculator v2.0 (53) (with the Model Averaging method) was used to calculate the nonsynonymous substitutions-to-synonymous substitutions (Ka/Ks) ratio.

Analyses of alternative polyadenylation, expression levels, differential usage of exons and translational efficiency

The BAM files for the MeRIP-seq inputs (RNA-seq) were used to analyze gene expression, APA and exon usage. The featureCounts program in the Subread package (54) was used to calculate gene read counts, which were then normalized to RPKM (reads per kilobase per million mapped reads) values. DaPars2 was used to infer the dynamic APA for each sample (55,56). The DEXSeq (57) package was used to examine differential exon usage on the basis of RNA-seq exon counts between samples examined using different experimental methods. The TE of each gene was calculated as 'RPKM (translation level)/RPKM (expression level)' on the basis of the ribosome profiling (Ribo-seq) datasets of four species (*A. thaliana*, *O. sativa*, *S. lycopersicum* and *Z. mays*), which were obtained from the NCBI SRA database.

Construction and analysis of co-methylated m^6A gene network

We evaluated the variant m^6A peaks to construct the m^6A gene co-methylation network using the WGCNA package (58) implemented in R. We constructed an adjacency matrix to describe the extent of the correlation of the m^6A levels between genes using the 'adjacency' function (scale-free R2 > 0.9). Subsequently, a topological overlap matrix (TOM) was derived from the adjacency matrix using the 'TOMsimilarity' function:

$$TOM_{ij} = \frac{\sum_{u \neq i,j}(a_{iu} * a_{uj}) + a_{ij}}{\min\{k_i, \; k_j\} + 1 - a_{ij}}$$

Finally, we calculated the eigengene, hierarchically clustered the modules and merged similar modules.

Trait ontology enrichment analysis of the m^6A gene co-methylation module

We conducted a trait ontology (TO) enrichment analysis to verify the gene modules that were enriched with agronomic trait-associated genes after obtaining the gene co-methylation modules. The p-value was calculated according to the hypergeometric distribution as previously described (59):

$$p\left(TO, \; module\right) \; = \; \sum_{x=m}^{\min(n,k)} \frac{\binom{k}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

where *N* represents the total number of genes, *K* represents the total number of genes assigned a TO term among all genes, *n* represents the number of genes in the module and *m* represents the number of genes assigned a TO term in the module. The modules with adjusted p-values less than 1e-3 were considered to be gene modules associated with the corresponding agronomic trait.

Nucleotide-binding leucine-rich repeat gene family identification and classification

We used previously described methods (60) to identify and classify nucleotide-binding leucine-rich repeat (NLR) gene families. Briefly, InterProScan (61) (with E-value < 1e-4) was used to identify the genes containing the NB-ARC domain as candidate genes. Next, NLR-Parser (version 3.0) (62) was used to determine the LRR domain, and the results were cross-verified with InterProScan. Finally, the candidate genes with any of motif 9, motif 11, or motif 19 in NLR-Parser were designated as NLR genes in rice. The NLR gene family members were divided into the following three categories according to their distribution on chromosomes: singleton genes, gene pairs and gene clusters.

Identification of genes encoding plant RNA modification-related enzymes

All genes encoding known RNA modification-related enzymes were obtained from the Modomics database. The encoded enzyme sequences in 20 plant species were extracted from the respective protein reference sequences. The Blastp program of blast-2.2.26 (63) was used to align sequences and then the best bidirectional hits were used to obtain the orthologous enzyme-encoding genes for the 20 plant species. In addition, the expression profiles of the identified RNA modification-related enzyme-encoding genes in different tissues were obtained from the Plant Public RNA-seq Database (http://ipf.sustech.edu.cn/pub/plantrna/) and displayed using the Plotly package (https://github.com/plotly/plotly.R). Moreover, protein–protein interactions among the enzymes encoded by these genes were determined using the STRING database (https://cn.string-db.org/) and displayed using RNAmodNet developed in this study. Because of the lack of related data, the current version of PRMD provides the RNA modification-related gene expression patterns for only six species and the protein–protein interactions for only three species.

Annotation of RNA modifications

RNAmod (24), which is an interactive and freely available platform for the annotation and visualization of RNA modifications, was used to annotate the RNA modifications in PRMD. First, RNAmod extracted gene features, such as promoter regions, 5′ and 3′ untranslated regions (UTRs), start codon regions, coding sequence (CDS) regions and stop codon regions, from different annotated reference genomes and then examined gene characteristics, including the GC content, length and minimum free energy. Second, RNAmod mapped all modification sites to different RNA features and calculated coverage values and analyzed metagenes and other annotations. The modified genes were functionally characterized on the basis of Gene Ontology (GO) (64) and KEGG pathways (65) using the clusterProfiler package (66) and according to Reactome pathways (67) using ReactomePA packages (68).

Database and web interface implementation

MySQL was used for storing and managing metadata in PRMD. The PHP/Apache environment in a Linux system equipped with 512 GB RAM and four Octa-core AMD processors (2.6 GHz each) was used to host the PRMD web server. The back-end workflow and plots were implemented using the Python/Perl language and R packages. Additionally, JQuery, DataTable, Highchart and igv.js were embedded in the system for dynamic and interactive data visualizations ([69],[70]). A web-based genome browser was implemented using JBrowse ([71]). mRNAbrowse was designed for visualizing RNA modifications on the basis of igv.js (https://github.com/igvteam/igv.js), whereas RNAlollipop was designed to develop lollipop views of RNA modifications.

Database content and usage

Overview of the PRMD content

We developed PRMD as a comprehensive resource for information regarding RNA modifications. For m^{6}A, 693 MeRIP-seq samples were carefully collected from the SRA and GSA databases for the following 19 plant species: *Ae. tauschii* (6), *A. thaliana* (259), *B. rapa* (24), *F. vesca* (18), *G. max* (24), *G. arboreum* (6), *G. hirsutum* (18), *M. domestica* (32), *N. benthamiana* (18), *O. sativa* (88), *P. fortunei* (8), *P. patens* (6), *P. trichocarpa* (26), *P. vulgaris* (6), *S. bicolor* (22), *S. lycopersicum* (38), *T. aestivum* (22), *T. dicoccoides* (6) and *Z. mays* (66). In addition, 12 MeRIP-seq samples were generated in this study for *O. rufipogon* (DXW81), *O. sativa* ssp. *indica* (WSSM) and *O. sativa* ssp. *japonica* (ZH11). The overall statistical analysis of all datasets indicated the m^{6}A methylation ratios were negatively correlated with the genome size and the number of genes in these plant species (Supplementary Figure S1).

Moreover, PRMD integrated datasets of other types of RNA modifications, such as m^{1}A, m^{5}C, m^{7}G, ac^{4}C, 2′-O-Me and pseudouridine, as well as additional related datasets, including those for eQTLs, SNVs, GWAS sites, rG4 structures, sORFs, RBP binding sites, RNA loops, RNA secondary structures, conservations and APA site information. All datasets were processed through our uniform pipelines. The information was deposited in a MySQL database and displayed in convenient web modules in PRMD. Furthermore, we designed PRMD to enable users to easily visualize and analyze the data in the database (Figure 1).

Web interface modules developed in PRMD

1) The 'Browse' module contains the following five tables: (i) browse by genes; (ii) browse by samples; (iii) other modifications; (iv) predicted m^{6}A sites and (v) RNA modification-related enzymes. The 'Browse by genes' table contains information about the genes with m^{6}A modifications (with hyperlinks for gene details), list of supported samples, list of m^{6}A orthologous genes and other visualization tools (e.g. mRNAbrowse and JBrowse). The 'Browse by samples' table consists of information about the MeRIP-seq samples (with hyperlinks for sample details in the original database), references and the RNA modification annotation module. The 'Other modifications' table contains other types of RNA modifica-

tions (Supplementary Table S4), with hyperlinks for the original resource and modification annotation module. The 'Predicted m^{6}A sites' table displays the predicted m^{6}A sites for 20 species (Supplementary Table S5), with hyperlinks for the annotation module for each species. The 'RNA modification-related enzymes' table includes a list of the genes encoding the RNA modification-related enzymes identified in 20 plant species (with hyperlinks for gene details), protein–protein interaction network comprising RNA modification-related enzymes and expression profiles of the genes encoding RNA modification-related enzymes in different tissues.

2) The 'Search' module in PRMD allows users to quickly obtain comprehensive information by submitting gene IDs, transcript IDs, gene symbols, sample IDs, study IDs, or PubMed IDs. We also added a quick-search function that supports one query on the PRMD homepage.

3) The 'Annotation' module provides details regarding RNA modifications, including the following: (a) overall statistics for the modifications across different RNA features and RNA biotypes; (b) density/histogram map of RNA modifications at RNA exon splice junctions, RNA transcription start/end sites and RNA translation start and stop codons; (c) identified motifs at RNA modification sites; (d) RNA modification site heatmap; (e) enriched GO terms and KEGG pathways among the modified genes and (f) metagene analysis in a specific gene context (Figure 2A, B). This module can be used to present RNA modification features in each sample. For example, m^{6}A sites were mainly located in the CDS and 3′ UTRs, with a density peak near the translation end sites. In contrast, the other types of modifications, including m^{5}C and pseudouridine, were similarly distributed, with density peaks near the translation start sites (Supplementary Figure S2).

4) The 'Visualization' module contains the following four visualization tools: (a) JBrowse embedded in a fast, scalable genome browser built using JavaScript and HTML5 to visualize all RNA modification sites and other publicly available annotations (genome scale) (Figure 2C); (b) mRNAbrowse, which intuitively displays RNA modifications (transcript scale) (Figure 2D); (c) RNAlollipop, which displays the lollipop view of merged datasets containing the transcript locations (Figure 2E) and (d) IGV, which visualizes the read distribution of modification peaks on the basis of the BigWig format files (genome scale).

5) The 'Tools' module contains the following four convenient analytical tools: (a) RMlevelDiff for analyzing RNA modification levels and differential modifications; (b) RMplantVar for detecting potential deleterious variant effects on RNA modifications; (c) RNAmodNet for analyzing gene co-methylation networks and visualizing m^{6}A co-modification gene networks (Figure 2F) and (d) Blast for identifying potential RNA modification-related enzymes among the sequences uploaded by users.

6) The other modules include the following: (a) the 'Statistics' module, which contains the overall statistics for RNA modifications among species; (b) the 'Download' module, which provides the datasets of different types of RNA modifications and different species, with users able to download data with one click depending on the species and type of RNA modification; (c) the 'Links'
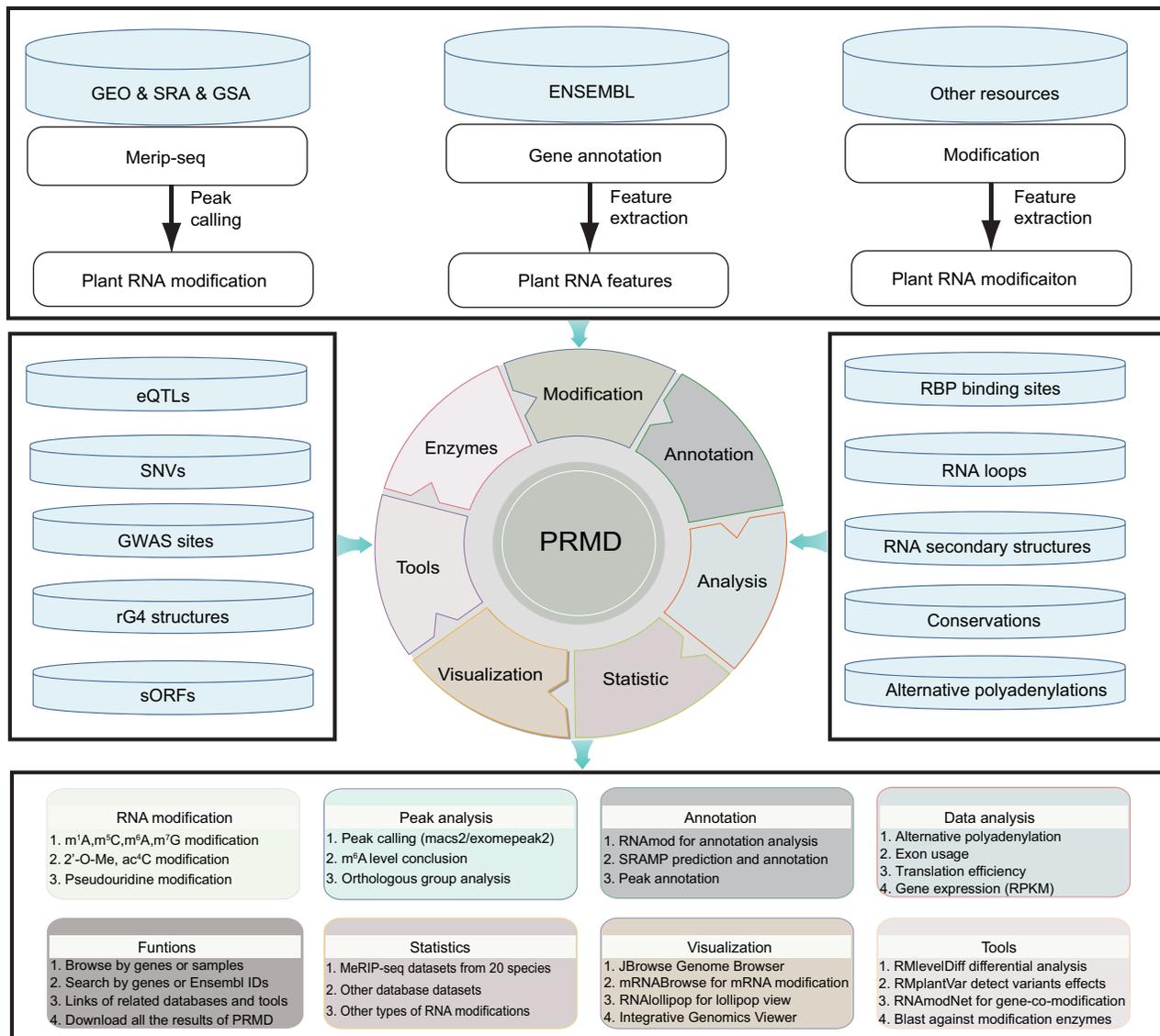
Figure 1. Overall workflow of PRMD. PRMD provides comprehensive information about RNA modifications. All of the data generated by PRMD were deposited in a MySQL database and displayed in several convenient modules on web pages. In addition to the m^{6}A modification, PRMD also contains other types of known RNA modifications (m^{1}A, m^{5}C, m^{7}G, ac^{4}C, 2'-O-Me and pseudouridine) and additional related datasets derived from published research articles and related plant resources. Furthermore, PRMD provides several convenient tools for visualizing and analyzing data.

module, which provides dozens of software and database links for RNA modifications and (d) the 'Document' module, which provides a detailed explanation for the data processing.

Visualization tools developed in PRMD

mRNAbrowse was designed for the intuitive visualization of RNA modifications and related datasets (transcript scale), including modification sites determined by nanopore sequencing and miCLIP-seq, sequence conservation, GWAS sites, miRNA target sites, APA sites, RBP binding sites, RNA secondary structures, rG4 structures, R-loop elements, sORFs and other types of modifications. In mRNAbrowse, users can zoom in and out using buttons in the upper right corner to visualize the modification site sequence context.

RNAlollipop was designed to establish lollipop views of the modifications and other datasets in PRMD. For each dataset

type, all of the datasets were merged to enable users to intuitively compare the RNA modifications with other elements at the same location from different sources. JBrowse was integrated to visualize the modification sites and other information (genome scale). In addition, IGV, with hyperlinks provided in the table with specific details and information, was integrated to visualize the distribution of the modification peaks on the basis of the BigWig files of the input and IP samples in the genome. Users can intuitively check the read coverage of the modification peaks called by MACS2 and exomePeak2 among different samples.

Web-based analytical tools developed in PRMD

RMlevelDiff was designed to analyze the differences in m^{6}A levels; the parameter settings are provided in Supplementary Figure S3. When users select a species, all of the MeRIP-seq samples for that species will be displayed. Users must divide

Figure 2. Main functions of the PRMD database. (**A**) Outputs of statistical analyses, including peak gene feature distribution, gene type distribution, coverage plots for different features, boundary coverage for transcription start/end sites, boundary coverage for translation start/end sites and boundary coverage for splice sites. (**B**) Outputs of annotation analyses, including RNA characteristics of genes with modifications, mRNA metagene plot, enriched motifs, enriched Gene Ontology terms among the modified genes and enriched functional pathways among the modified genes. (**C**) Visualization of JBrowse data. JBrowse displays tracks of RNA modifications and other related annotations with genomic coordinates. (**D**) Visualization of mRNAbrowse data. mRNAbrowse shows tracks of RNA modifications and other related annotations. (**E**) Visualization of RNAlollipop data. RNAlollipop was designed to produce lollipop views of RNA modifications. (**F**) Outputs of the gene co-methylation network analysis. Dots represent gene names, whereas lines represent protein–protein interactions. Because of the limited number of samples, only four species are currently available for gene co-methylation analyses.

their samples of interest into different groups to perform the analysis. Notably, each group should contain at least one sample. The outputs include the following: (a) a volcano plot of the differences in modification levels, with red and green points representing up-regulated and down-regulated modification sites, respectively; (b) a box plot and a heatmap plot showing the distribution of m^6A levels among samples; (c) a detailed list of m^6A modification peaks in selected samples and the $\log_2$(fold-change) value, *P*-value and false discovery rate and (d) a detailed list of the means and fold-changes of the two groups of selected samples. Using *O. sativa* (Accession: SAMN19341036) as an example, the shoot and root tissue samples were divided into two groups, with each group comprising two pairs of FTO transgenic and wild-type samples. The RMlevelDiff analysis indicated that the shoot (FTO transgenic versus wild-type) had 10 up-regulated and 44 down-regulated peaks, whereas the root (FTO transgenic versus wild-type) had 7 up-regulated and 43 down-regulated peaks. For the shoot and the root, the FTO transgenic samples had a significantly higher m^6A methylation ratio (*P*-value < 2.2e − 16) than the wild-type samples (Figure 3A–D).

RMplantVar was designed to detect potential variants affecting RNA modifications. A variant file in the VCF format and the objective sample list from PRMD are used as the inputs for RMplantVar (Supplementary Figure S3). After users submit the data, the queue system will provide a job ID, which can be used to check the progress of the analysis and retrieve the results. Moreover, users have the option of being notified by email when their job has been completed. The results include a pie chart presenting the distribution of the m^6A lost, m^6A gained and unchanged variants as well as a table with specific details, including the location of variants, transcript IDs, peak IDs, reference motifs, mutated motifs and the score used to determine whether the variants were lost or gained. Using the variations from the Rice SNP-Seek Database (https://snp-seek.irri.org/_download.zul) as an example, RMplantVar analyzed the sequence characteristics of the sites with mutated RNA modifications. Approximately 1.3% of the mutation sites were designated as m^6A lost variants, whereas 0.1% of the mutation sites were designated as m^6A gained variants. Accordingly, these mutation sites altered the DRACH motif structure, which may influence the binding of writers/readers/erasers (Figure 3E and Supplementary Table S6).

RNAmodNet was developed to visualize the m^6A co-modification gene network. PRMD currently supports four species with many samples (>30), including *A. thaliana*, *O. sativa*, *S. lycopersicum* and *Z. mays*. Users are required to select a species and a gene from the input parameters. All genes associated with the selected gene will be displayed, with different colors used to represent different network modules. Additionally, gene names and their modules will be displayed on the left panel. Specific gene details can be obtained by clicking the gene name.

Basic Local Alignment Search Tool (Blast) was integrated in PRMD to screen for genes encoding known RNA modification-related enzymes. Users can upload protein sequences in the FASTA format and input other related Blast parameters. The results consist of a table that lists query sequence IDs, gene IDs of known RNA modification-related enzymes, percentage of identity, alignment length, e-values and the start/end of the query and subject. These genes encoding RNA modification-related enzymes can be used to search for particular RNA modifications and perform downstream functional studies.

Case studies

In rice, *IDEAL PLANT ARCHITECTURE INTERACTING PROTEIN 1 (IPI1)* is an important gene that encodes a protein that modulates IPA1 protein levels, thereby regulating the plant architecture (72). The two transcripts derived from this gene differ in terms of two RNA modifications (m^6A and m^5C) in PRMD. The 'Browse' module revealed that the m^6A of the transcript Os01t0350900-01 is supported by 27 MACS2 and 15 exomePeak2 samples, whereas the m^6A in the other transcript (Os01t0350900-02) is supported by 37 MACS2 and 32 exomePeak2 samples. Additionally, mRNAbrowse displayed the tracks of the RNA modification peaks and different related sites (transcript scale), including m^5C modification sites, predicted m^6A sites and an APA site in the PlantAPAdb database. RNAlollipop presented the distribution of the modification sites and other related information for whole transcripts, whereas JBrowse showed the tracks of the m^6A RNA modifications and other related sites (genome scale), including m^5C modification sites and predicted m^6A sites. Furthermore, 43 orthologous m^6A-modified genes were detected among 20 plant species (Figure 4A–E).

The m^6A modification sites were highly enriched in the CDS. The GC content was higher in the CDS than in the UTRs, whereas the opposite trend was detected for the minimum free energy. The metagene analysis of RNA modifications indicated that m^6A was enriched near the stop codon and the consensus motif 'DRACH' was enriched (Figure 4F, G). We further analyzed the distribution of the m^6A peaks in IGV using the genome mapping files, which revealed that m^6A modifications occurred at the same location in the genomes of different samples. The *IPI1* homolog in *A. thaliana* (AT3G05545) had a similar distribution in IGV (Supplementary Figure S4), indicative of conserved m^6A modifications between rice and *A. thaliana*.

Comparative analyses using the data in PRMD

Orthogroup gene identification, evolution analyses

In total, 47671 orthogroups were identified in 20 plant species. We subsequently selected four species with more than 30 samples (*A. thaliana*, *O. sativa*, *S. lycopersicum* and *Z. mays*) to perform a comparative analysis, which detected 1126 (*A. thaliana*), 1252 (rice), 1218 (tomato) and 1608 (maize) species-specific orthogroups and 9715 common orthogroups. The comparison between *A. thaliana* and rice identified 2881 one-to-one orthogroups with m^6A modifications. Synonymous and nonsynonymous substitutions among these orthogroup genes revealed that the evolutionary rate was significantly lower for both m^6A-modified genes than for the non-m^6A-modified genes (*P*-value < 2.2e−16), indicating that the genes with m^6A modifications were under purifying selection during evolution (Supplementary Figure S5A–C).

Gene co-methylation network analyses of rice and *arabidopsis thaliana*

We obtained 20 and 16 modules in the m^6A co-methylation networks of *O. sativa* and *A. thaliana*, respectively, on the basis of the correlations between m^6A indices (Figure 5A, B
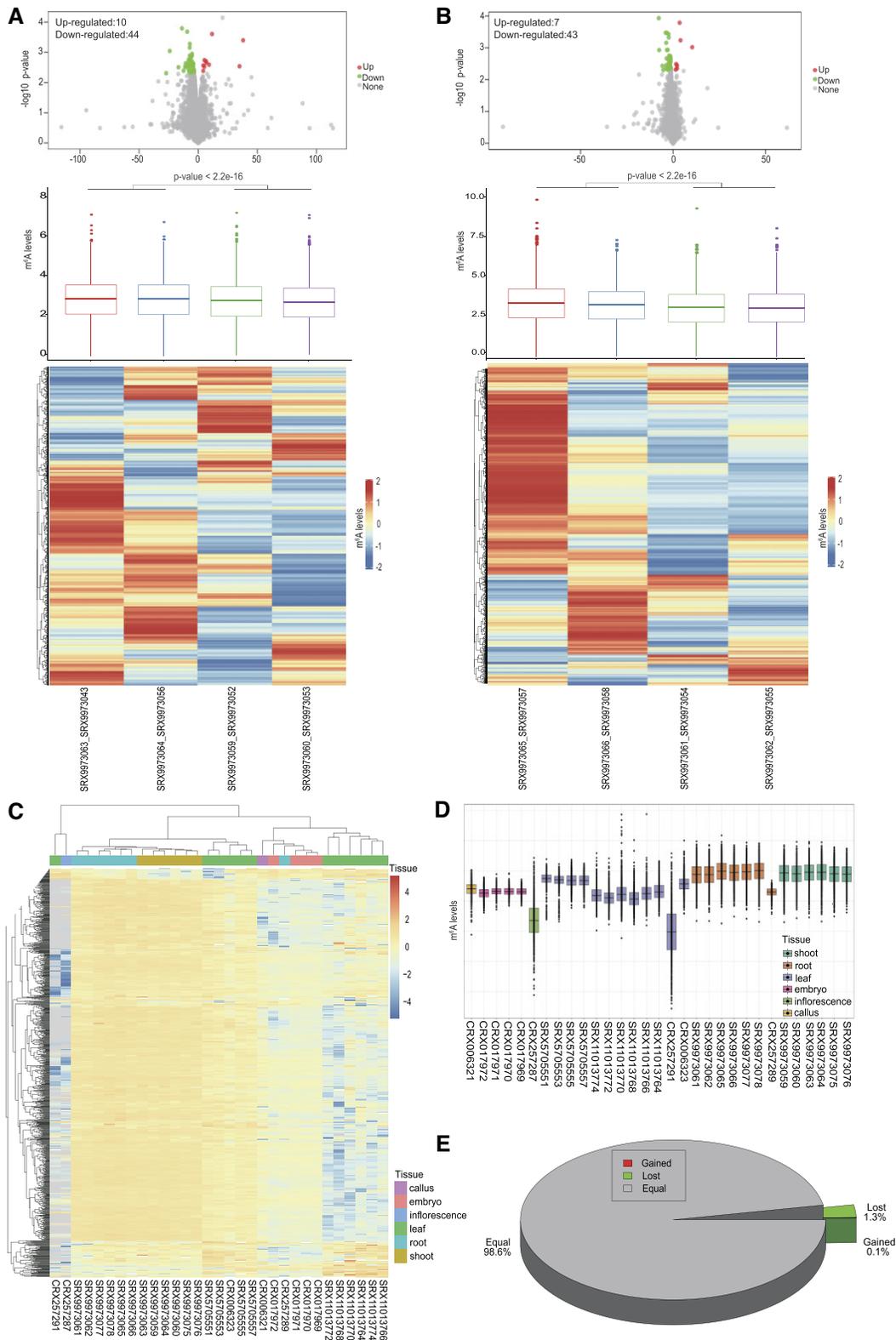
Figure 3. Comparative analysis of m^{6}A modifications in *Oryza sativa*. (**A**, **B**) RMlevelDiff comparison of m^{6}A levels between samples. (**C**) Pheatmap of m^{6}A levels. (**D**) Boxplot of m^{6}A levels. (**E**) RMplantVar outputs, including a pie chart and a table.

Figure 4. Comprehensive analysis of *IPA1* in PRMD. (**A**) Outputs of the 'Search' module in PRMD revealed that *IPA1* produces two transcripts with differing RNA modifications (m⁶A and m⁵C). (**B**) mRNAbrowse displays the tracks of mRNA modification peaks derived from different methods and different sources. (**C**) Lollipop view of RNA modifications. (**D**) JBrowse shows the tracks of m⁶A RNA modifications and other related annotations. (**E**) Details regarding *IPA1*. (**F**) Various annotations, including peak gene feature and type distributions, coverage plots for different features and boundary coverage for transcription start/end sites. (**G**) RNA characteristics of genes with modifications, RNA metagene plots, enriched motifs, enriched pathways and gene annotation information are presented in a tabular form.
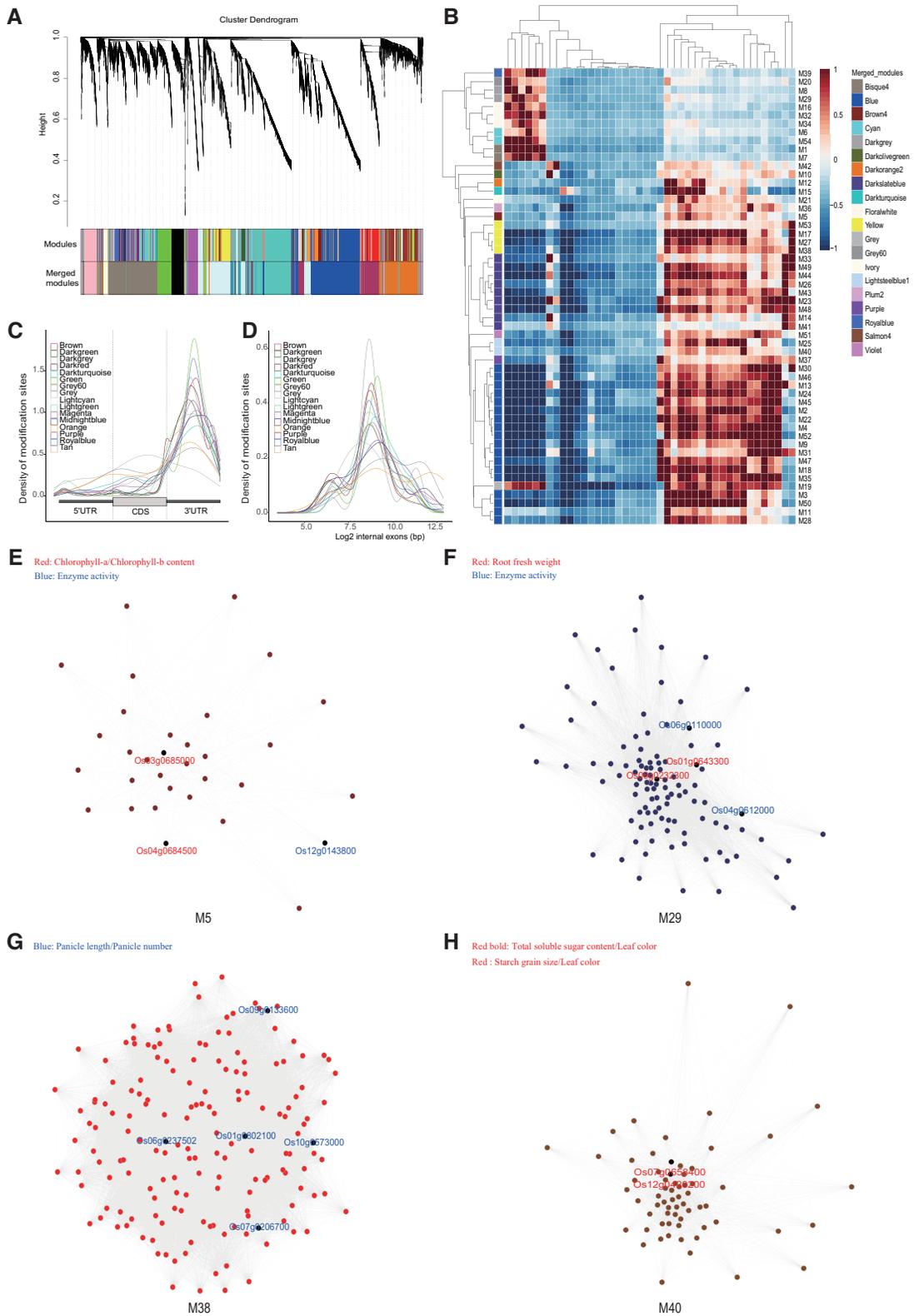
Figure 5. Classification and analyses of the co-methylation m⁶A modules in *Oryza sativa*. (**A**) Classification of co-methylated m⁶A modules. (**B**) Heatmap presenting the m⁶A indices of all co-methylation modules across all rice samples. (**C**) Density distributions of m⁶A peaks in different combined modules across the 5′ UTR, CDS and 3′ UTR. (**D**) Density distributions of log-transformed lengths of the internal exons with m⁶A peaks in different combined co-methylation modules. (**E–H**) In *Oryza sativa*, the following modules were associated with several traits: M5 (chlorophyll content and enzyme activity), M29 (root fresh weight and enzyme activity), M38 (panicle length/panicle number) and M40 (total soluble sugar content, leaf color and starch grain size).

and Supplementary Figure S6A, B). The metagene analysis of the m⁶A peak distribution in different modules indicated that m⁶A peaks were strongly enriched in the 3′ UTR in rice, which was consistent with the results of an earlier study by Yu *et al*. (73). The lengths of the internal exons with m⁶A were similarly distributed in the co-methylation modules of *O. sativa* (Figure 5C, D). In contrast, in *A. thaliana*, m⁶A was enriched in the CDS and 3′ UTR and the internal exon length distribution varied among the co-methylation modules (Supplementary Figure S6C, D), suggesting the m⁶A methylomes had species-specific dynamic topologies.

We also explored the distribution of m⁶A-modified genes that were associated with different traits in *O. sativa*. More than 50% of the m⁶A-modified genes were related to responses to biotic and abiotic stresses as well as quality and yield (Supplementary Figure S5D), reflecting the influence of RNA modifications on these important traits. We performed TO analyses to explore the m⁶A co-methylation modules significantly associated with these important traits. In rice, several modules closely related to specific traits were identified, including M5 (chlorophyll content and enzyme activity), M29 (root fresh weight and enzyme activity), M38 (panicle length/panicle number) and M40 (total soluble sugar content, leaf color and starch grain size) (Figure 5E–H). Notably, M40 was a broadly expressed module associated with three different traits, indicative of the functional diversity of the genes in this module. However, in *A. thaliana*, modules M2, M10, M48 and M55 were related to panicle size, male sterility/starch content, sucrose content/sugar content and male sterility, respectively (Supplementary Figure S6E–H), indicative of the conservation of the m⁶A co-methylation modules for these traits. Interestingly, M38 of *O. sativa* and M2 of *A. thaliana* were associated with the panicle, implying different gene modules may have similar functions in different species.

Comparative analysis of *oryza rufipogon* and two cultivated rice subspecies (*oryza sativa* ssp. *Indica* and *oryza sativa* ssp. *Japonica*)

The MeRIP-seq datasets derived from our sequencing analyses of *O. rufipogon* (DXW81), *O. sativa* ssp. *indica* (ZH11) and *O. sativa* ssp. *japonica* (WSSM) were selected for the comparative analysis. The m⁶A peak density distributions were consistent with the earlier observations that m⁶A accumulated substantially around the stop codon and within the 3′ UTR and that the m⁶A methylation ratio among all genes was higher for both WSSM and ZH11 than for DXW81 (Supplementary Figure S7A, B). Orthologous group analyses detected 19620 orthogroups that had a high m⁶A methylation ratio in DXW81, WSSM and ZH11. The DXW81 *vs* WSSM and DXW81 versus ZH11 comparisons indicated that the m⁶A methylation ratio was significantly higher for the one-to-one orthologous gene pairs (DXW81–WSSM and DXW81–ZH11) than for the species-specific m⁶A-modified genes (*P*-value < 2.2e−16; Fisher's exact test). Synonymous and non-synonymous substitutions among these orthogroup genes suggested that m⁶A-modified genes were likely under purifying selection during evolution (Figure 6A–C). The common enriched GO terms (DXW81 and WSSM as well as DXW81 and ZH11) among the orthologous genes with m⁶A modifications included metalloendopeptidase activity, DNA-templated transcription initiation and galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase activity, suggestive of the rela-

tive similarity in these m⁶A-modified genes between the two cultivated rice subspecies (Figure 6D).

We also identified NLR gene families in DXW81, WSSM and ZH11, and then divided them into three categories (singleton genes, gene pairs and gene clusters). The m⁶A methylation ratios were higher for the gene pairs and gene clusters than for the singleton genes. Moreover, the m⁶A methylation ratios for the gene pairs and gene clusters were higher for WSSM and ZH11 than for DXW81 (Supplementary Figure S7C). These results imply m⁶A methylations may have been important for rice domestication.

Comparison with other databases

Several databases and web servers that integrate the existing sequencing data have been constructed to help researchers identify diverse RNA modifications. However, most of these databases were focused on mammalian species, especially human and mouse (Table 1). For example, m6AVar (74) and RMVar (75) focus on m⁶A-associated variants that potentially affect RNA modifications in human and mouse. The MeT-DB v2.0, REPIC and RMBase v2.0 databases contain only one type of RNA modification (m⁶A) for one plant species (*A. thaliana*). The m6A-Atlas database (76) comprises 442162 m⁶A sites identified from the epitranscriptome profiles of several species (e.g. human, mouse, fly, zebrafish, rat, yeast, *A. thaliana* and virus); the updated version of m6A-Atlas (i.e. v2.0) (23) includes only six plant species (*A. thaliana*, black cottonwood, maize, tomato, rice and wild strawberry), with 172 MeRIP-seq samples for *A. thaliana* (78), *F. vesca* (18), *O. sativa* (8), *S. lycopersicum* (18) and *Z. mays* (24). The ENCORE database is an updated version of RMBase that includes 10 plant species, but m⁶A modifications are provided only for *A. thaliana* and there are relatively few samples for different modifications. In addition, there are single-species databases, such as AthMethPre (18) and RFAthM6A (19), which are only used for predicting m⁶A sites in *A. thaliana*. Most importantly, the available databases lack comprehensive annotations and cannot intuitively visualize plant RNA modifications and other related data.

Increasing numbers of recent studies in diverse plant species have profiled several RNA modifications in different tissues, developmental stages and stress conditions, which may be relevant for constructing an integrated plant RNA modification database. We developed PRMD by collecting and processing MeRIP-seq data generated in previous studies on 20 plant species. Moreover, we incorporated several convenient tools, such as RMlevelDiff, RMplantVar, RNAmodNet and Blast (for data analyses) as well as JBrowse, mRNAbrowse, RNAlollipop and IGV (for visualizing diverse datasets). Additionally, PRMD provides information regarding annotations and may be used to analyze APA, exon usage, TE and differential expression. Moreover, PRMD intuitively displays the related datasets from previous studies and other resources (e.g. eQTLs, SNVs, GWAS, sORFs, RNA loops, rG4 structures, RBP binding sites, RNA secondary structures and APAs).

Perspectives

There is increasing evidence of the links between aberrant RNA modifications and many key biological processes in plants. In particular, m⁶A, which is the most prevalent and abundant RNA modification, reversibly regulates RNA pro-
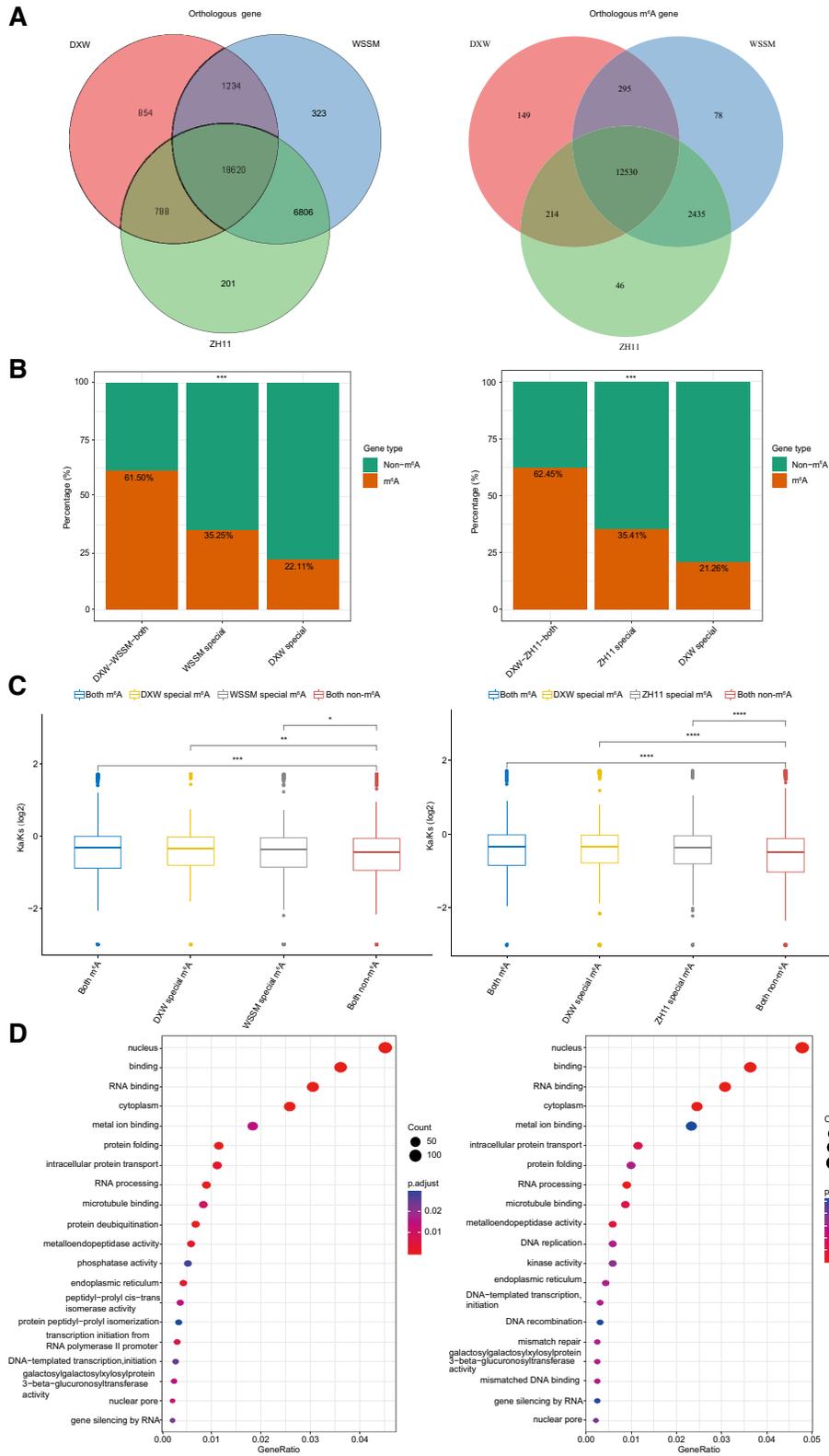
Figure 6. Analysis of wild rice (*Oryza rufipogon*) and two cultivated rice subspecies (*Oryza sativa* ssp. *indica* and *Oryza sativa* ssp. *japonica*). (**A**) Number of orthologous genes and m^{6}A-modified genes among *O. rufipogon, O. sativa* ssp. *indica* and *O. sativa* ssp. *japonica*. (**B**) m^{6}A methylation ratios for the one-to-one orthologous gene pairs of DXW81 versus WSSM and DXW81 *vs* ZH11. (**C**) Ka/Ks values for the one-to-one orthologous gene pairs of DXW81 versus WSSM and DXW81 *vs* ZH11. (**D**) GO enrichment analysis of the one-to-one orthologous gene pairs of DXW81 versus WSSM and DXW81 versus ZH11.

Table 1. Comparison with other integrated RNA modification related databases

Name	PRMD	m^{6}A-Atlas 2.0	MeT-DB 2.0	REPIC	RMBase v2.0	ENCORE
Data resources of plants						
Species	20	6	1	1	1	10
Modification types	m^{1}A, m^{5}C, m^{6}A, m^{7}G, ac^{4}C, 2′O-Me, Pseudo	m^{6}A	m^{6}A	m^{6}A	m^{6}A	m^{1}A, m^{5}C, m^{6}A, m^{7}G, 2′O-Me, Pseudo
m^{6}A-seq	693	156	42	36	16	16
Other types modification sites	79973	No	No	No	No	2762
Nanopore-seq	Yes	No	No	No	No	No
miCLIP-seq	Yes	Yes	Yes	No	No	Yes
Other plant specific regulation datasets						
RBP binding sites	Yes	Yes	No	No	No	Yes
RNA Structures	Yes	No	No	No	No	No
micro-RNA target sites	Yes	No	No	No	No	No
Single nucleotide polymorphisms (SNPs)	Yes	Yes	No	No	No	No
Expression Quantitative Trait Loci (eQTLs)	Yes	No	No	No	No	No
GWAS sites	Yes	No	No	No	No	No
Small open reading frames (sORFs)	Yes	No	No	No	No	No
Tools						
Differentially methylated analysis	Yes	Yes	No	No	No	No
Variation effects on modifications	Yes	No	No	No	No	No
Analysis						
Predicted m^{6}A sites	Yes	No	Yes	Yes	Yes	Yes
Peak calling and annotations	Yes	Yes	Yes	Yes	Yes	Yes
Orthologous analysis	Yes	No	No	No	No	No
Co-methylation analysis	Yes	No	No	No	No	No
Post-transcription analysis	Yes	No	No	No	No	No
Blast to search RNA modification enzymes	Yes	No	No	No	No	No
Visualizations						
mRNA coordinate view	Yes	No	Yes	No	No	No
Genomic coordinate view	Yes	Yes	Yes	Yes	No	No
m^{6}A visualizations	Yes	Yes	Yes	Yes	No	No
Gene-co-methylation Netviewer	Yes	No	No	No	No	No
Gene feature/types distribution	Yes	No	Yes	Yes	No	No
Coverage plots for different features	Yes	No	No	No	No	No
Boundary coverage for translation start/end and splice sites	Yes	No	No	No	No	No
mRNA metagene plot	Yes	No	No	Yes	No	No
Enriched motifs	Yes	No	No	Yes	Yes	Yes
Peak heatmaps near transcription/translation start and end sites	Yes	No	No	No	No	No
Gene Ontology functional enrichment for modified genes	Yes	No	No	No	No	No
Functional Pathway enrichment for modified genes	Yes	No	No	No	No	No
RNA characteristics of genes with modifications	Yes	No	No	No	No	No
The ways to query datasets						
Gene ID	Yes	Yes	Yes	Yes	Yes	Yes
Gene Name	Yes	No	Yes	Yes	No	No
Transcript ID	Yes	No	No	No	No	No
Sample accession	Yes	No	No	No	No	No
Study ID	Yes	No	No	No	No	No
PubMed ID	Yes	No	No	No	No	No
Links	http://bioinformatics.sc.cn/PRMD	http://rnamd.org/m6a/index.php	http://compgenomics.utsa.edu/MeTDB/	https://repicmod.uchicago.edu/repic	http://rna.sysu.edu.cn/rmbase	https://rna.sysu.edu.cn/encore/
Ref		(23)	(21)	(22)	(20)	

'Other plant specific regulation datasets' represents datasets from other relevant data sources. 'Post-transcription analysis' contain the data analysis of alternative polyadenylation, exon usage, translational efficiency and differential expression. Most existed RNA modifications databases focus on mammal species without plants, especially human and mouse, such as m^{6}Avar (74), RMVar (75) and WHISTLE (17). Compared with previous databases in the table which contain plant species, there has no database specifically covering RNA modification in multiple plant species.

cessing and metabolism related to plant development, evolution and pathological processes. Although several databases have been constructed to integrate the existing sequencing datasets and different types of RNA modifications, a comprehensive and easy-to-use database comprising information regarding diverse RNA modifications in plant species is still lacking. Therefore, we developed PRMD primarily to facilitate research on plant RNA modifications. Currently, PRMD includes six RNA modification types in 20 plant species. One limitation of this database is that there are relatively few MeRIP-seq samples for several of the plant species (5 of 20) and relevant datasets were not included for all species. For example, eQTL datasets are available for only *A. thaliana* and *O. sativa*, while R-loop datasets are available for only *A. thaliana*, *G. max*, *O. sativa* and *Z. mays*. Because of the continual development of high-throughput sequencing technology, there will likely be a rapid increase in the number of available sequencing datasets, which will greatly enhance research on RNA modifications. We will regularly update PRMD and add sequencing and related datasets to cover more plant species and RNA modifications. Furthermore, PRMD includes several convenient visualization and functional analysis tools. As a long-term research project, we will develop more helpful tools for functional analyses of PRMD datasets. We believe that PRMD is a very convenient and comprehensive database of plant RNA modifications.

Data availability

PRMD is a comprehensive and convenient interactive online database available at http://bioinformatics.sc.cn/PRMD/ and http://rnainformatics.org.cn/PRMD. The raw-sequencing MeRIP-seq data of *Oryza rufipogon* and two cultivated rice subspecies have been deposited in the Genome Sequence Archive (GSA) database of the National Genomics Data Center (NGDC; https://ngdc.cncb.ac.cn/gsa/), which can be obtained via BioProject accession PRJCA018161. The pipeline code for Merip-seq data processing used in this study have been deposited in github (https://github.com/rnainformatics/PRMD) and figshare (https://doi.org/10.6084/m9.figshare.24139029).

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

Funding

Conflict of interest statement

None declared.

References

1. Yu,X., Sharma,B. and Gregory,B.D. (2021) The impact of epitranscriptomic marks on post-transcriptional regulation in plants. *Brief. Funct. Genomics*, **20**, 113–124.
2. Kadumuri,R.V. and Janga,S.C. (2018) Epitranscriptomic code and its alterations in Human disease. *Trends Mol. Med.*, **24**, 886–903.
3. Dominissini,D. and Rechavi,G. (2018) Epitranscriptome regulation. *Nat. Struct. Mol. Biol.*, https://doi.org/10.1038/s41594-018-0140-7.
4. Shao,Y., Wong,C.E., Shen,L. and Yu,H. (2021) N(6)-methyladenosine modification underlies messenger RNA metabolism and plant development. *Curr. Opin. Plant Biol.*, **63**, 102047.
5. Li,S. and Mason,C.E. (2014) The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genomics Hum. Genet.*, **15**, 127–150.
6. Meyer,K.D. and Jaffrey,S.R. (2014) The dynamic epitranscriptome: n6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.*, **15**, 313–326.
7. Yue,H., Nie,X., Yan,Z. and Weining,S. (2019) N6-methyladenosine regulatory machinery in plants: composition, function and evolution. *Plant Biotechnol. J.*, **17**, 1194–1208.
8. Miao,Z., Zhang,T., Xie,B., Qi,Y. and Ma,C. (2022) Evolutionary implications of the RNA N6-methyladenosine methylome in plants. *Mol. Biol. Evol.*, **39**, msab299.
9. Li,B., Li,D., Cai,L., Zhou,Q., Liu,C., Lin,J., Li,Y., Zhao,X., Li,L., Liu,X., *et al.* (2023) Transcriptome-wide profiling of RNA N(4)-cytidine acetylation in Arabidopsis thaliana and Oryza sativa. *Mol Plant*, **16**, 1082–1098.
10. Wang,S., Wang,H., Xu,Z., Jiang,S., Shi,Y., Xie,H., Hua,J. and Wu,Y. (2023) m6A mRNA modification promotes chilling tolerance and modulates gene translation efficiency in Arabidopsis. *Plant Physiol.*, **192**, 1466–1482.
11. Tang,Y., Gao,C.C., Gao,Y., Yang,Y., Shi,B., Yu,J.L., Lyu,C., Sun,B.F., Wang,H.L., Xu,Y., *et al.* (2020) OsNSUN2-Mediated 5-methylcytosine mRNA modification enhances rice adaptation to high temperature. *Dev. Cell*, **53**, 272–286.
12. Miao,Z., Zhang,T., Qi,Y., Song,J., Han,Z. and Ma,C. (2020) Evolution of the RNA N (6)-methyladenosine methylome mediated by genomic duplication. *Plant Physiol.*, **182**, 345–360.
13. Helm,M. and Motorin,Y. (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.*, **18**, 275–291.
14. Meyer,K.D., Saletore,Y., Zumbo,P., Elemento,O., Mason,C.E. and Jaffrey,S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, **149**, 1635–1646.
15. Han,Y., Feng,J., Xia,L., Dong,X., Zhang,X., Zhang,S., Miao,Y., Xu,Q., Xiao,S., Zuo,Z., *et al.* (2019) CVm6A: a visualization and exploration database for m(6)As in cell lines. *Cells*, **8**, 168.
16. Zhou,Y., Zeng,P., Li,Y.H., Zhang,Z. and Cui,Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.*, **44**, e91.

17. Chen,K., Wei,Z., Zhang,Q., Wu,X., Rong,R., Lu,Z., Su,J., de Magalhaes,J.P., Rigden,D.J. and Meng,J. (2019) WHISTLE: a high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach. *Nucleic Acids Res.*, **47**, e41.

18. Xiang,S., Yan,Z., Liu,K., Zhang,Y. and Sun,Z. (2016) AthMethPre: a web server for the prediction and query of mRNA m(6)A sites in Arabidopsis thaliana. *Mol. Biosyst.*, **12**, 3333–3337.

19. Wang,X. and Yan,R. (2018) RFAthM6A: a new tool for predicting m(6)A sites in Arabidopsis thaliana. *Plant Mol. Biol.*, **96**, 327–337.

20. Xuan,J.J., Sun,W.J., Lin,P.H., Zhou,K.R., Liu,S., Zheng,L.L., Qu,L.H. and Yang,J.H. (2018) RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res.*, **46**, D327–D334.

21. Liu,H., Wang,H., Wei,Z., Zhang,S., Hua,G., Zhang,S.W., Zhang,L., Gao,S.J., Meng,J., Chen,X., *et al.* (2018) MeT-DB V2.0: elucidating context-specific functions of N6-methyl-adenosine methyltranscriptome. *Nucleic Acids Res.*, **46**, D281–D287.

22. Liu,S., Zhu,A., He,C. and Chen,M. (2020) REPIC: a database for exploring the N(.6)-methyladenosine methylome. *Genome Biol.*, **21**, 100.

23. Liang,Z., Ye,H., Ma,J., Wei,Z., Wang,Y., Zhang,Y., Huang,D., Song,B., Meng,J., Rigden,D.J., *et al.* (2023) m6A-Atlas v2.0: updated resources for unraveling the N6-methyladenosine (m6A) epitranscriptome among multiple species. *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gkad691.

24. Liu,Q. and Gregory,R.I. (2019) RNAmod: an integrated system for the annotation of mRNA modifications. *Nucleic Acids Res.*, **47**, W548–W555.

25. Zhai,J., Song,J., Cheng,Q., Tang,Y. and Ma,C. (2018) PEA: an integrated R toolkit for plant epitranscriptome analysis. *Bioinformatics*, **34**, 3747–3749.

26. Zhai,J., Song,J., Zhang,T., Xie,S. and Ma,C. (2021) deepEA: a containerized web server for interactive analysis of epitranscriptome sequencing data. *Plant Physiol.*, **185**, 29–33.

27. Howe,K.L., Achuthan,P., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R., Bhai,J., *et al.* (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.

28. Zhu,T., Liang,C., Meng,Z., Sun,G., Meng,Z., Guo,S. and Zhang,R. (2017) CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol.*, **17**, 101.

29. Jung,S., Lee,T., Cheng,C.H., Buble,K., Zheng,P., Yu,J., Humann,J., Ficklin,S.P., Gasic,K., Scott,K., *et al.* (2019) 15 years of GDR: new data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Res.*, **47**, D1137–D1145.

30. Goodstein,D.M., Shu,S., Howson,R., Neupane,R., Hayes,R.D., Fazo,J., Mitros,T., Dirks,W., Hellsten,U., Putnam,N., *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.

31. Fernandez-Pozo,N., Menda,N., Edwards,J.D., Saha,S., Tecle,I.Y., Strickler,S.R., Bombarely,A., Fisher-York,T., Pujar,A., Foerster,H., *et al.* (2015) The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. *Nucleic Acids Res.*, **43**, D1036–D1041.

32. Pertea,G. and Pertea,M. (2020) GFF utilities: gffRead and GffCompare. *F1000Res*, **9**, ISCB Comm J-304.

33. Chen,T., Chen,X., Zhang,S., Zhu,J., Tang,B., Wang,A., Dong,L., Zhang,Z., Yu,C., Sun,Y., *et al.* (2021) The genome sequence archive Family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics*, **19**, 578–583.

34. CNCB-NGDC Members and Partners (2023) Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. *Nucleic Acids Res.*, **51**, D18–D28.

35. Zhang,Y., Jiang,J., Ma,J., Wei,Z., Wang,Y., Song,B., Meng,J., Jia,G., de Magalhaes,J.P., Rigden,D.J., *et al.* (2023) DirectRMDB: a database of post-transcriptional RNA modifications unveiled from direct RNA sequencing technology. *Nucleic Acids Res.*, **51**, D106–D116.

36. Lan,Y., Sun,R., Ouyang,J., Ding,W., Kim,M.J., Wu,J., Li,Y. and Shi,T. (2021) AtMAD: arabidopsis thaliana multi-omics association database. *Nucleic Acids Res.*, **49**, D1445–D1451.

37. Liu,X., Tian,D., Li,C., Tang,B., Wang,Z., Zhang,R., Pan,Y., Wang,Y., Zou,D., Zhang,Z., *et al.* (2023) GWAS Atlas: an updated knowledgebase integrating more curated associations in plants and animals. *Nucleic Acids Res.*, **51**, D969–D976.

38. Yu,H., Qi,Y., Yang,B., Yang,X. and Ding,Y. (2023) G4Atlas: a comprehensive transcriptome-wide G-quadruplex database. *Nucleic Acids Res.*, **51**, D126–D134.

39. Zhao,W., Zhang,S., Zhu,Y., Xi,X., Bao,P., Ma,Z., Kapral,T.H., Chen,S., Zagrovic,B., Yang,Y.T., *et al.* (2022) POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res.*, **50**, D287–D294.

40. Li,K., Wu,Z., Zhou,J., Xu,W., Li,L., Liu,C., Li,W., Zhang,C. and Sun,Q. (2023) R-loopAtlas: an integrated R-loop resource from 254 plant species sustained by a deep-learning-based tool. *Mol Plant*, **16**, 493–496.

41. Li,P., Zhou,X., Xu,K. and Zhang,Q.C. (2021) RASP: an atlas of transcriptome-wide RNA secondary structure probing data. *Nucleic Acids Res.*, **49**, D183–D191.

42. Zhu,S., Ye,W., Ye,L., Fu,H., Ye,C., Xiao,X., Ji,Y., Lin,W., Ji,G. and Wu,X. (2020) PlantAPAdb: a comprehensive database for alternative polyadenylation sites in plants. *Plant Physiol.*, **182**, 228–242.

43. Tian,F., Yang,D.C., Meng,Y.Q., Jin,J. and Gao,G. (2020) PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.*, **48**, D1104–D1113.

44. Boccaletto,P., Stefaniak,F., Ray,A., Cappannini,A., Mukherjee,S., Purta,E., Kurkowska,M., Shirvanizadeh,N., Destefanis,E., Groza,P., *et al.* (2022) MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.*, **50**, D231–D235.

45. Chen,S., Zhou,Y., Chen,Y. and Gu,J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.

46. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

47. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.

48. Meng,J., Lu,Z., Liu,H., Zhang,L., Zhang,S., Chen,Y., Rao,M.K. and Huang,Y. (2014) A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/bioconductor package. *Methods*, **69**, 274–281.

49. DeLuca,D.S., Levin,J.Z., Sivachenko,A., Fennell,T., Nazaire,M.D., Williams,C., Reich,M., Winckler,W. and Getz,G. (2012) RNA-SeQC: rNA-seq metrics for quality control and process optimization. *Bioinformatics*, **28**, 1530–1532.

50. Quinlan,A.R. (2014) BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics*, **47**, 11–34.

51. Emms,D.M. and Kelly,S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, **20**, 238.

52. Zhang,Z., Xiao,J., Wu,J., Zhang,H., Liu,G., Wang,X. and Dai,L. (2012) ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.*, **419**, 779–781.

53. Wang,D., Zhang,Y., Zhang,Z., Zhu,J. and Yu,J. (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*, **8**, 77–80.

54. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

55. Feng,X., Li,L., Wagner,E.J. and Li,W. (2018) TC3A: the cancer 3′ UTR atlas. *Nucleic Acids Res.*, **46**, D1027–D1030.

56. Li,L., Huang,K.L., Gao,Y., Cui,Y., Wang,G., Elrod,N.D., Li,Y., Chen,Y.E., Ji,P., Peng,F., *et al.* (2021) An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.*, **53**, 994–1005.

57. Anders,S., Reyes,A. and Huber,W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.*, **22**, 2008–2017.

58. Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.*, **9**, 559.

59. Zhang,Y., Han,E., Peng,Y., Wang,Y., Geng,Z., Xu,Y., Geng,H., Qian,Y. and Ma,S. (2022) Rice co-expression network analysis identifies gene modules associated with agronomic traits. *Plant Physiol.*, **190**, 1526–1542.

60. Shang,L., Li,X., He,H., Yuan,Q., Song,Y., Wei,Z., Lin,H., Hu,M., Zhao,F., Zhang,C., *et al.* (2022) A super pan-genomic landscape of rice. *Cell Res.*, **32**, 878–896.

61. Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A., Nuka,G., *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.

62. Steuernagel,B., Jupe,F., Witek,K., Jones,J.D. and Wulff,B.B. (2015) NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics*, **31**, 1665–1667.

63. Mount,D.W. (2007) Using the basic local alignment search tool (BLAST). *CSH Protoc.*, **2007**, pdb top17.

64. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

65. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

66. Yu,G., Wang,L.G., Han,Y. and He,Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.

67. Gillespie,M., Jassal,B., Stephan,R., Milacic,M., Rothfels,K., Senff-Ribeiro,A., Griss,J., Sevilla,C., Matthews,L., Gong,C., *et al.* (2022) The reactome pathway knowledgebase 2022. *Nucleic Acids Res.*, **50**, D687–D692.

68. Yu,G. and He,Q.Y. (2016) ReactomePA: an R/bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.*, **12**, 477–479.

69. Kerpedjiev,P., Hammer,S. and Hofacker,I.L. (2015) Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377–3379.

70. Robinson,J.T., Thorvaldsdottir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

71. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.

72. Wang,J., Yu,H., Xiong,G., Lu,Z., Jiao,Y., Meng,X., Liu,G., Chen,X., Wang,Y. and Li,J. (2017) Tissue-specific ubiquitination by IPA1 INTERACTING PROTEIN1 modulates IPA1 protein levels to regulate plant architecture in rice. *Plant Cell*, **29**, 697–707.

73. Yu,Q., Liu,S., Yu,L., Xiao,Y., Zhang,S., Wang,X., Xu,Y., Yu,H., Li,Y., Yang,J., *et al.* (2021) RNA demethylation increases the yield and biomass of rice and potato plants in field trials. *Nat. Biotechnol.*, **39**, 1581–1588.

74. Zheng,Y., Nie,P., Peng,D., He,Z., Liu,M., Xie,Y., Miao,Y., Zuo,Z. and Ren,J. (2018) m6AVar: a database of functional variants involved in m6A modification. *Nucleic Acids Res.*, **46**, D139–D145.

75. Luo,X., Li,H., Liang,J., Zhao,Q., Xie,Y., Ren,J. and Zuo,Z. (2021) RMVar: an updated database of functional variants involved in RNA modifications. *Nucleic Acids Res.*, **49**, D1405–D1412.

76. Tang,Y., Chen,K., Song,B., Ma,J., Wu,X., Xu,Q., Wei,Z., Su,J., Liu,G., Rong,R., *et al.* (2021) m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Res.*, **49**, D134–D143.