This work was funded by European Research Council (DOUBLE-TROUBLE 833522).

Convergent and/or parallel evolution of RNA-binding proteins in angiosperms after polyploidization

Liangyu Guo[1]* , Shuo Wang[1]*, Xi Jiao[1]*, Xiaoxue Ye[2] , Deyin Deng[1], Hua Liu[1], Yan Li[1] , Yves Van de Peer[3,4,5] and Wenwu Wu[1]

[1]State Key Laboratory of Subtropical Silviculture, School of Forestry and Biotechnology, Zhejiang A&F University, Lin'an, Hangzhou, 311300, China;
[2]Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, 571101, China;
[3]Department of Plant Biotechnology and Bioinformatics, VIB – UGent Center for Plant Systems Biology, Ghent University, B-9052, Ghent, Belgium;
[4]College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing, 210095, China;
[5]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, 0028, South Africa

* Equal contribution

Authors for correspondence: Yan Li (Email: liyan2016@zafu.edu.cn), Yves Van de Peer (Email: yves.vandepeer@psb.ugent.be), Wenwu Wu (Email: wwwu@zafu.edu.cn)

Key words: alternative splicing, C-repeat binding factors, cold stress, convergent evolution, global cooling, K-Pg boundary, RNA-binding proteins, whole-genome duplication.

Summary

- Increasing studies suggest that the biased retention of stress-related transcription factors (TFs) after whole-genome duplications (WGDs) could rewire gene transcriptional networks, facilitating plant adaptation to challenging environments. However, the role of posttranscriptional factors (e.g. RNA-binding proteins, RBPs) following WGDs has been largely ignored.
- Uncovering thousands of RBPs in 21 representative angiosperm species, we integrate genomic, transcriptomic, regulatomic, and paleotemperature datasets to unravel their evolutionary trajectories and roles in adapting to challenging environments.
- We reveal functional enrichments of RBP genes in stress responses and identify their convergent retention across diverse angiosperms from independent WGDs, coinciding with global cooling periods. Numerous RBP duplicates derived from WGDs are then identified as coldinduced. A significant overlap of 29 orthogroups between WGD-derived and cold-induced RBP genes across diverse angiosperms highlights a correlation between WGD and cold stress. Notably, we unveil an orthogroup (Glycine-rich RNA-binding Proteins 7/8, GRP7/8) and relevant TF duplicates (CCA1/LHY, RVE4/8, CBF2/4, etc.), co-retained in different angiosperms post-WGDs. Finally, we illustrate their roles in rewiring circadian and cold-regulatory networks at both transcriptional and posttranscriptional levels during global cooling.
- Altogether, we underline the adaptive evolution of RBPs in angiosperms after WGDs during global cooling, improving our understanding of plants surviving periods of environmental turmoil.

Introduction

The rapid rise and diversification of angiosperms (flowering plants) was coined by Charles Darwin as an 'abominable mystery' (Friedman, 2009). Polyploidization, the result of wholegenome duplication (WGD), is rampant in angiosperms and can profoundly alter genome size, gene content, and genetic diversity, and has therefore been regarded as a critical evolutionary force driving angiosperm speciation, adaptation, and diversification (Soltis et al., 2009; Van de Peer et al., 2021).

Despite the shared polyploid ancestry among extant angiosperms and the occurrence of additional WGD events in various angiosperm lineages, the intervals between successive WGD events span tens of millions of years (Carretero-Paulet & Van de Peer, 2020). This indicates that the survival and establishment of WGD events are formidable challenges. Newly formed polyploids face many issues, such as genomic instability, abnormal chromosome division, and the competitive edge of surrounding diploids, leading to minority cytotype exclusion (Comai, 2005). The successful fixation of WGD events is thus exceedingly rare (Van de Peer et al., 2017; Qiao et al., 2019). However, polyploids may have an increased resistance in adverse environments. This is supported by their prevalence in colder regions of high latitude or elevation (Schinkel et al., 2016; Rice et al., 2019). Additionally, a wave of WGD events have been identified in different angiosperm lineages around the Cretaceous-Paleogene (K-Pg) boundary (Fawcett et al., 2009; Vanneste et al., 2014; Van de Peer et al., 2017; Wu et al., 2020). The K-Pg boundary is marked by mass extinctions triggered by an asteroid impact that caused prolonged global cooling and darkness on Earth (Pope et al., 1997; Vellekoop et al., 2014; Junium et al., 2022).

The increase in the amount of genetic material from WGDs and subsequent gene retention and loss, as well as their functional divergence, potentially facilitate plant adaptation to environmental changes (Seoighe & Gehring, 2004; Song et al., 2020; Wu et al., 2020). Biased retention of genes, particularly those linked to abiotic stress, following WGD events may have contributed to the rewiring of stress-related regulatory networks (De Smet et al., 2017; Song et al., 2020; Wu et al., 2020; Guo et al., 2022; Nie et al., 2022). For example, key components in major coldtolerance regulatory networks, such as C-repeat binding factors (CBFs), inducers of CBF expression (ICEs), CAMTAs, SIZs, OSTs, and EINs, were convergently retained in different angiosperm lineages after WGDs around the K-Pg boundary (Song et al., 2020; Wu et al., 2020; Nie et al., 2022). This retention potentially increased the complexity and robustness of coldtolerance regulatory networks, contributing to plant adaptation during global cooling periods.

Increasing 'evidence' for the contribution of polyploidy to plant adaptive evolution came from the retention of stress-related transcription factor (TF) gene families post-WGDs, impacting gene transcriptional networks (Freeling, 2009; De Smet et al., 2017; Wu et al., 2020; Guo et al., 2022; Nie et al., 2022). However, posttranscriptional regulation is equally crucial in orchestrating gene expression, primarily through RNA splicing, transport, stability, and translation (Nakaminami & Seki, 2018). Alternative splicing (AS), a form of RNA splicing, generates diverse transcripts from the same gene, enhancing transcriptome and proteome diversity (Reddy et al., 2013; Chaudhary et al., 2019). RNA-binding proteins (RBPs), including splicing factors, play a key role in regulating AS events and other RNA processes such as export, translation, and turnover. Despite the validation of hundreds of RBPs in Arabidopsis thaliana (Reichel et al., 2016; Zhang et al., 2016; Bach-Pages et al., 2020), their involvement in plant adaptation to challenging environments remains poorly understood. Our recent findings highlighted the biased retention of the RNA recognition motif (RRM)- containing gene family, one of the most abundant RBPs in plants, ranking among the top 20 gene families preserved in eudicots after WGDs (Guo et al., 2022). This study explores whether biased retention of RBPs occurred in angiosperms post-WGDs and investigates how retained RBPs were linked to co-retained TFs in cooperatively rewiring gene regulatory networks for plant adaptation to harsh environments.

Hereby, we trace the evolutionary history of RBPs in 21 selected species from different angiosperm families, such as Arecaceae, Poaceae, Brassicaceae, Juglandaceae, Salicaceae, Solanaceae, Fabaceae, Euphorbiaceae, Convolvulaceae, and Nelumbonaceae (Fig. 1a). First, we identify all RBP genes and analyze their functional roles in the selected species. Subsequently, we delve into the expansion of RBP genes in the context of paleotemperature fluctuations. Additionally, through experimental investigation, we evaluate the involvement of retained RBP genes from WGD events in

cold response, identifying highly retained and conserved cold-induced RBP orthogroups across diverse angiosperms. Finally, through a comparative analysis of present and predicted ancestral gene regulatory networks, we aim to reconstruct the involvement of stressrelated RBP duplicates and TF duplicates, co-retained post-WGDs during global cooling periods, in rewiring circadian and coldtolerance regulatory networks at both transcriptional and posttranscriptional levels.

Materials and Methods

Identification of RBP genes

The protein sequences of 21 selected angiosperms were downloaded from publicly available databases (Supporting Information Table S1). For RBP genes, we compiled experimentally confirmed RBPs in Arabidopsis thaliana (L.) Heynh. from three independent studies (Reichel et al., 2016; Zhang et al., 2016; Bach-Pages et al., 2020; Table S2). To ensure reliability, we considered only RBPs identified in at least two of the three studies, resulting in 403 RBPs (Fig. S1a). Subsequently, we utilized the 405 RBPs as a query library to retrieve closely related RBP homologs in A. thaliana using stringent BLASTP parameters (E-value = 1e–10, identity = 60%; Altschul et al., 1990). This process identified 557 RBPs in A. thaliana. Moreover, we employed these A. thaliana RBPs to retrieve homologous RBPs in each of 20 other species with the same BLASTP parameters (E-value = 1e–10, identity = 60%; Fig. S1b). The resulting list of retrieved RBP genes is provided in Table S3.

Gene family classification of RBP genes

First, we downloaded all hidden Markov model (HMM) profiles (c. 16 000) of protein domains from the Pfam database (up to July 2021; Mistry et al., 2021). Subsequently, we conducted hmmsearch (E-value = 1e–5; Eddy, 2011) to query the HMM profiles against the previously retrieved RBPs in each species. Based on the identified domains, we classified RBP genes into different gene families (Table S3).

Gene Ontology enrichment analysis

The Gene Ontology (GO) term annotations for A. thaliana genes were downloaded from the website TAIR10 (Rhee et al., 2003). Annotations for genes in other species were assigned based on the highest alignment score matches against A. thaliana proteins from BLASTP searches (E-value = 1e–5). For the gene set of interest, GO term enrichment analysis was performed using the OmicShare tools (www.omicshare.com/tools). The resulting enrichment levels (P-value) were adjusted for multiple testing using Benjamini–Hochberg false discovery rate (FDR) and represented by enrichment scores (_$\log_{10}$ FDR).

Duplication modes of RBP duplicates

In each of the 21 selected species, we first performed a selfcomparison of protein sequences using the BLASTP program (E- value = 1E–10; Altschul et al., 1990) and extracted the top five hits for each query. Based on the hits and their coordinates on chromosomes, we identified intragenomic collinear blocks using the program MCSCANX (Wang et al., 2012a) with parameters (MATCH_ SCORE = 50, GAP_PENALTY =_1, MATCH_SIZE = 5, Evalue = 1e–5), where at least five paired homologous genes define a synteny block. Subsequently, we conducted the script duplicate_- gene_classifier (Wang et al., 2012a) to categorize genes into five duplication modes: WGD, tandem duplication (TD), proximal duplication (PD), dispersed duplication (DD), and Singleton.

Many PD-derived duplicates likely resulted from ancient TD events but were interrupted by gene insertions (Wang et al., 2012b). Further classification involved labeling PD-derived duplicates separated by three or fewer genes as TD-derived duplicates and the remaining as DD-derived duplicates. Additionally, some DD-derived duplicates might arise from DNA- or RNAbased transposed duplication (TSP) events (Wang et al., 2012b). We employed the program DUPGEN_FINDER (Qiao et al., 2019) to distinguish potential duplicates from TSP. Thus, we obtained and summarized duplication modes of all coding genes, including RBP and non-RBP genes, in the 21 selected species (Tables S3, S4).

Identification of 494 gene-rich families

As described previously (Guo et al., 2022), we initially searched HMM profiles (> 16 000) against all proteins in the 21 selected species using the hmmsearch (E-value = 1e–5; Eddy, 2011). Subsequently, we classified the genes sharing a common domain into the same family, totaling 4594 families. We then extracted generich families, which have 10 or more members in a minimum of half of the selected species. In cases of partial gene redundancy between families, we retained the larger family when two shared > 60% of their genes. Finally, we identified 494 gene-rich families in angiosperms (Table S5).

Compilation and periodization of WGD events

The phylogeny of the 21 selected species was sourced from the TimeTree database (Kumar et al., 2017), and their historical WGD events were obtained from the publications (Table S6). We categorized these WGD events into four periods: ancient f/e-WGD events (> 181 million years ago (Ma), shared by all angiosperms; Jiao et al., 2011; Ruprecht et al., 2017), ancient c/r/s-WGD events (c. 104–120 Ma, shared by core eudicots or monocots; Vekemans et al., 2012; Wu et al., 2020), and recent R1-WGD (c. 66Ma around the K-Pg boundary) and R2-WGD (< 25 Ma in the Late Cenozonic Icehouse; Vanneste et al., 2014; Wu et al., 2020).

Assignment of WGD-derived duplicates to specific WGD events

Whole-genome duplications and subsequent gene retention/loss can result in intricate collinearities and gene duplicates (Bowers et al., 2003). Typically, a WGD event generates synteny blocks containing gene duplicates with similar synonymous substitution rates (Ks), characterized by a Gaussian distribution. Using ParaAT v.2.0 (Zhang et al., 2012) and KaKs_Calculator v.2.0 (Wang et al., 2010), we calculated Ks value for each pair of collinear duplicates and plotted the Ks distribution for each species (Fig. S2). The analysis of the Ks distribution revealed an association between Ks ranges and WGD events in the respective species (Table S6). Certain WGD-specific Ks ranges were corroborated by previous studies (Jain et al., 2013; Plomion et al., 2018; Wu et al., 2020). Notably, simultaneous WGD events may exhibit somewhat different Ks ranges among species, due to variations in their generation times.

Subsequently, based on Ks values, we assigned collinear blocks to the corresponding WGD event. Recognizing the potential for redundancy from recent WGDs leading to intricate collinearities and duplicate combinations for ancient WGDs (Bowers et al., 2003; Seoighe & Gehring, 2004), we introduced a rectified step. In brief, when a recent WGD event produced two duplicates, and both were paired with a collinear homolog from an ancient WGD event, only one pair was assigned to the ancient WGD event. Based on the WGD-assigned collinear blocks, we identified RBP duplicates from each WGD event (Table S7).

Paleotemperature data

As described previously (Guo et al., 2022), data of global average temperature (GAT) fluctuations over the past 160 million years (Myr) were collected from the reference (Scotese et al., 2021). These data aimed to infer the paleoclimate temperature conditions during the occurrences of WGD events.

Design of RNA-seq experiments in eight cold-treated species

To investigate cold responses of RBP genes, we selected four eudicots (Glycine max (L.) Merr., Populus trichocarpa Torr. & Gray, Carya illinoinensis (Wangenh.) K.Koch, and A. thaliana) and four monocots (Setaria italica (L.) Beauv., Zea mays L., Oryza sativa L., and Phyllostachys edulis (Carri_ere) J.Houz.) from the 21 selected species. As described previously (Wang et al., 2023a,b), these species were subjected to cold treatment (4°C) for 0, 2, 24, and 168 h for RNA-seq experiments (Fig. S3a). Before cold-stress treatments, all plant seedlings were cultured at 25°C with a 16h:8h, light :dark photoperiod. For A. thaliana, 3-wk-old seedlings were used for cold-stress treatments. In other seven species, young seedlings (c. 30 cm in height) were employed. In each species, to ensure all samples harvesting on the same day at the same developmental stages, Group 1 (cold treatment for 168 h) were first cultured at 4°C wk^{-1} before harvest. Groups 2 and 3 (cold treatment for 24 and 2 h,

respectively) were successively cultured at 4°C for appropriate durations. Group 4 (control) were consistently cultured at 25°C (0h).

After cold stress (a week later) for each species, the fourth expanded leaves of seedlings in the four groups were collected simultaneously (12 : 00 at noon) on the same day. Three biological replicates were collected for each group. Total RNA was isolated and purified using TRIzol reagent (Invitrogen). RNA quantity and purity were assessed by NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) and Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). Poly(A)- RNA was then purified using poly-T oligo-attached magnetic beads to generate strand-specific cDNA libraries with inserts of c. 150–200 bp. The libraries were sequenced on the Illumina NovaSeq 6000 system (2 9 150 bp paired-end reads) at LC-Bio Technology CO., Ltd (Hangzhou, China).

Analysis of RNA-seq datasets

Clean reads of the above RNA-seq samples were aligned to their respective genome (Table S1) using HISAT2-2.1.0 with the parameters (--min-intronlen 20 --max-intronlen 5000 --rna-strandness RF; Kim et al., 2015). Subsequently, gene expression was quantified by STRINGTIE v.2.0.3 (Pertea et al., 2015) with default parameters. Differentially up- and downregulated genes were then identified by DESEQ2 (Love et al., 2014) using the criteria: a FDR < 0.01, fold change > 1.5 between normal (0 h) and either cold treatment (2, 24, and 168 h), and a minimum expression value exceeding 20 in at least one condition. Genes upregulated under cold stress were designated as cold-induced genes, including cold-induced RBP genes (Table S8). Additionally, five types of AS events, including intron retention (IR), alternative $5_0/3_0$ splice site (A5SS/A3SS), skipping exon (SE), and mutually exclusive exons (MXE), were identified by RMATS v.4.0.2 (Shen et al., 2014). Differential AS events (DASEs) between normal (0 h) and cold conditions (2, 24, or 168 h) was identified based on an FDR ≤ 0.05 (Table S9).

Orthogroup analysis of cold-induced RBP genes across eight cold-treated species

To reveal cold-induced RBP orthologs, we conducted orthogroup analysis in the abovementioned eight cold-treated species. We initially categorized all protein-coding genes into 30 812 orthogroups using ORTHOFINDER v.2.3.8 (Emms & Kelly, 2019) with the parameters: -M dendroblast -S diamond -A mafft -I 1.5 -T fasttree. Retaining the orthogroups that encompass all eight species, we obtained a total of 9621 orthogroups, including 273 RBP orthogroups.

Considering the potential issues of paralogs arising before the monocot-eudicot divergence within the same orthogroup, we conducted a phylogenetic analysis on the 273 RBP orthogroups. For each RBP orthogroup, we performed multiple sequence alignment with MAFFT v.7.508 (Katoh & Standley, 2013) and constructed the maximum likelihood tree using IQ-TREE v.2 (Minh et al., 2020) with 1000 bootstraps. The phylogenetic tree was then manually examined. Of the 273 orthogroups, 242 with monophyletic clades in the species were directly retained. Additionally, 27 orthogroups with multiphyletic clades generated before the monocot-eudicot divergence were separated into 41 recategorized orthogroups. During the process, the orthogroups lacking ortholog in either of the eight species were excluded. Two examples are provided to illustrate the strategy (Fig. S4). In total, 282 RBP orthogroups were reidentified in eight cold-treated species (Table S10).

For each species, duplication modes for RBP orthogroup members were obtained from above duplication mode analysis (Table S7). Meanwhile, their expression values under cold stress were extracted from above RNA-seq analysis (Table S8). Finally, we employed Fisher's exact test to evaluate the significance of the overlap between R-WGD-derived RBP orthogroups and coldinduced RBP orthogroups in each of the eight cold-treated species.

Integrative analysis of Arabidopsis iCLIP-, RNA-, and ChIP/DAP-seq datasets

The genome-wide analysis of 858 glycine-rich RNA-binding protein 7 (GRP7) binding transcripts was conducted using Arabidopsis individual nucleotide resolution crosslinking and immunoprecipitation sequencing (iCLIP-seq; Meyer et al., 2017). Among the GRP7-binding transcripts, 433 were identified as cold-induced based on above Arabidopsis RNA-seq analysis (Table S11). To predict TFs responsible for cold induction of these 433 genes, we screened a library of 540 TFs, each with a

genome-wide binding profile determined by DNA affinity purification (DAP-seq; O'Malley et al., 2016) or chromatin immunoprecipitation (ChIP-seq) sequencing experiments (Zhang et al., 2013; Nagel et al., 2015; Kamioka et al., 2016). Among the 540 TFs, 108 were cold-induced (Table S12) and considered as potential TFs regulating the expression of the 433 cold-induced genes. The ChIP/DAP-seq datasets of the 108 TFs were downloaded from the NCBI BioProject (PRJNA257556, PRJNA281176, PRJNA288957, and PRJNA170235).

To identify the genome-wide binding sites of the 108 TFs, we aligned their respective ChIP/DAP-seq reads to the A. thaliana genome (TAIR10; Rhee et al., 2003) using Bowtie2 v.2.3.5 with default parameters (Langmead & Salzberg, 2012). Subsequently, we performed peak calling usingMacs2 v.2.7.1 with the parameters (--keep-dup 1 -g 119 481 543 -p 0.05) and annotated the peaks with nearest genes using ChIPseeker v.1.20 (Yu et al., 2015). Among the 433 cold-induced genes, those with at least one binding peak in their upstream promoters (1-kb) were considered as direct target genes of the corresponding TF (Table S13).

Statistical analysis

The difference between samples or groups was assessed by the paired t-test. The significance of enrichment analysis was evaluated by Fisher's exact test. The FDR was computed using the Benjamini and Hochberg method. Pearson's correlation coefficient was employed to evaluate the correlation between the two groups. All statistical tests were performed in R.

Results

Genome-wide identification and functional analysis of RBP genes

Based on experimentally confirmed 403 Arabidopsis RBPs (Reichel et al., 2016; Zhang et al., 2016; Bach-Pages et al., 2020), we conducted a comprehensive analysis to identify RBPs across 21 selected angiosperm species (Fig. S1). The identified RBPs exhibit a range from 435 to 1043 (with a median of 604), revealing their prevalence in angiosperms (Fig. 1a; Table S3). Subsequently, we employed HMM profiles (> 16 000;Mistry et al., 2021) to discern their containing protein domains. Our result highlights significant enrichments of RNA-binding domains, especially RRM (Fig. 1b; Table S3), aligning with experimentally identified results (Reichel et al., 2016; Zhang et al., 2016; Bach-Pages et al., 2020).

Moreover, we performed GO term enrichment analysis to elucidate their functional roles (Fig. 1c). As expected, RBPs display a predominant enrichment in fundamental posttranscriptional processes, such as mRNA binding, RNA helicase activity, RNA splicing, and regulation of RNA stability and translation. Notably, several GO terms related to abiotic stress, including responses to cold and cold acclimation, are among the most enriched terms (Fig. 1c). These findings suggest that RBP genes are predominant in some specific families and play crucial roles in fundamental posttranscriptional processes, with a notable role in responses to abiotic stresses, particularly cold response.

Generation of RBP duplicates through WGDs

To elucidate the origin and diversification of RBP duplicates, we used MCScanX (Wang et al., 2012a) and DupGen_finder (Qiao et al., 2019) to categorize the generation of RBP genes into five duplication modes, including WGD, TSP, TD, DD, and Singleton (Table S3). For simplicity, we arbitrarily considered Singleton as one 'duplication' mode. Among these modes, WGD emerges as the most dominant mode to produce RBP duplicates (c. 29–78% with an average of c. 55%; Fig. 2a). TSP serves as a noteworthy supplementary mode in producing RBP duplicates. By contrast, TD seems to play a limited role (producing c. 7% RBP duplicates), despite being documented to generate c. 16% of Arabidopsis genes and c. 14% of rice genes (Rizzon et al., 2006). Additionally, a substantial number of RBP duplicates were observed to originate from DD, possibly resulting from degenerate collinearity or gene transpositions obscured by a long evolutionary history (Freeling, 2009). These findings highlight the predominant derivation of RBP duplicates from WGDs, despite potential underestimation.

RBP duplicates significantly retained after WGD events

A comprehensive statistical analysis on protein-coding genes across the 21 selected species reveals that the proportion of WGD-derived genes is significantly greater in RBP genes than in non-RBP genes (Fig. 2b; Table S4). Despite a substantial proportion of TSP-derived genes in RBP genes, their contribution to the expansion of RBP genes is considerably smaller than that of WGD. Conversely, the TD, DD, and Singleton modes show a significant depletion in the expansion of RBP genes (Fig. 2b).

We further explored the extent to which the retention of RBP genes is compared with that of other gene families in angiosperms after WGD events. Among the 494 identified gene-rich families in the 21 selected species, we showed the top 20 retained gene families after WGDs (Fig. 2c; Table S14). Consistent with previous reports (Wu et al., 2020; Guo et al., 2022), genes encoding signal transducers and/or proteins involved in multiprotein complexes (e.g. Ras, IQ, DUF4228, AUX_IAA, and EF-hand family genes) and stress-related TFs, such as MYB, AP2/ERF, WRKY, bHLH, bZIP, and NAC genes were over-retained in angiosperms after WGDs (Fig. 2c). Notably, the RBP gene superfamily is also among the top 20 retained families, suggesting a preferential retention of posttranscriptional regulators, particularly RRM-containing RBP family genes, after WGDs. The findings suggest a preferential retention of RBP genes in angiosperms following WGDs, which implies that the retained RBP genes might have played a role in plant adaptation.

Retention of RBP duplicates during periods of global cooling

There are multiple rounds of WGDs in the evolution of angiosperms. Ancient WGD events (f- and/or e-WGD) likely preceded angiosperm radiation (Jiao et al., 2011; Ruprecht et al., 2017), while a hexaploidy event (c-WGT) and two WGDs (r- and s- WGD) characterized the evolution of early core eudicots and monocots, respectively (Vekemans et al., 2012; Ming et al., 2015; Wu et al., 2020). Recent WGD events (R-WGD) independently took place in various angiosperm lineages, clustering around the K-Pg boundary (R1-WGD) and the Late Cenozoic Icehouse (R2-WGD) with global cooling (Vanneste et al., 2014; Wu et al., 2020; Fig. 3a; Table S6).

We analyzed the proportions of RBP duplicates retained from these ancient and recent WGD events (Table S7). As shown in Fig. 3(b), R-WGD events, especially R1-WGD, are the most significant sources of extant RBP duplicates. For example, in N. nucifera, over 95% of WGD-derived RBP duplicates originated from R1-WGD around the K-Pg boundary. Similarly, in P. edulis, c. 33 and 56% of WGD-derived RBP duplicates were derived from R1-WGD around the K-Pg boundary and R2-WGD in the Late Cenozoic Icehouse, respectively. The analyses exhibit that numerous RBP duplicates were consistently retained across different angiosperm lineages from independent WGD events, coinciding with periods of global cooling.

Cold responses of R-WGD-derived RBP duplicates

Like stress-related TFs (Wu et al., 2020; Guo et al., 2022), the retention of RBP duplicates might have conferred a fitness advantage for plants during global cooling. To reveal the potential of RBP duplicates in cold responsiveness, we conducted RNA-seq analyses before and after cold treatments (0, 2, 24, and 168 h) in eight selected species (Fig. S3a). In each species, RNA-seq analysis demonstrated relatively similar expression levels of leaf development-related genes across the cold treatments, ensuring comparable developmental stages of these samples (Fig. S3b; Table S15). Subsequently, we identified cold-induced genes in each of the eight cold-treated species, revealing a significantly higher proportion of RBP genes (average c. 36%) undergoing cold induction compared with non-RBP genes (average c. 13.8%; Fig. 4a). All eight cold-treated species, especially A. thaliana, C. illinoinensis, G. max, and S. italica, exhibited significant high proportions of cold-induced RBP genes.

We then assessed the number and proportion of cold-induced RBPs from WGD, TSP, TD, DD, and Singleton (Fig. 4b; Table S16). Interestingly, WGD-derived RBPs exhibited a significantly higher proportion in cold induction than other model-derived RBPs in these cold-treated species. Further analysis indicated that cold-induced RBPs primarily originated from recent WGDs (R-WGD including R1- and R2-WGD), while ancient f/e- and c/r/s-WGDs made limited contributions to extant cold-

induced RBP duplicates (Fig. 4c; Table S17). Moreover, expression heatmaps consistently confirmed increased expression of R-WGD-derived RBP duplicates after cold stress (2, 24, and/or 168 h) in the eight cold-treated species, although some RBP duplicates evolved different expression patterns in cold responsiveness (Fig. 4d). Like RBP duplicates, cold-induced TF duplicates were preferentially retained after R-WGD events (Fig. S5). To validate the reliability and repeatability of our results, we collected 18 additional datasets of cold-responsive transcriptomes from independent publications with differences in tissues, durations of cold treatment, and developmental stages (Table S18). All collected experiments consistently demonstrated that a significant high proportion of R-WGD-derived RBP genes were cold-upregulated, irrespective of tissues, durations of cold treatment, and developmental stages (Fig. S6). These findings strongly support the notion that many RBP duplicates retained in angiosperms following R-WGD amidst global cooling are indeed cold-upregulated. **The outcome of posttranscriptional alternative splicing** The cold induction of RBP genes may trigger posttranscriptional regulations, including AS. To investigate this, we conducted transcriptome-wide AS analysis in the eight cold-treated species, identifying thousands to tens of thousands of cold-induced differentially AS genes (DASGs) which cover substantial proportions (12–29%) of plant multi-exon genes (Fig. 4e). Interestingly, the number and proportion of DASGs in Arabidopsis is significantly smaller than those in seven other species, possibly due to shorter introns in Arabidopsis (Fig. S7). This observation aligns with the assumption that AS tends to occur around exons interrupted by long introns (Fox-Walsh et al., 2005; Du et al., 2023). Additionally, the distributions of DASEs across the eight cold-treated species exhibited a consistent pattern: more DASEs occurred at the middle and late stages of cold stress (24 and 168 h) than at the early stage (2 h; Fig. 4f). This pattern is consistent with increased expression of cold-induced RBPs at corresponding time points (24 and 168 h; Fig. 4d). Among the five types of DASEs, such as IR, SE, A5SS/A3SS, and MXE (Ast, 2004; Du et al., 2023), IR emerged as the most common AS mode in angiosperms (Fig. 4f), consistent with prior studies (Calixto et al., 2018). Further GO enrichment analysis showed that these DASGs participate in biological processes of stress response, circadian rhythm, growth and development, transcriptional regulation, etc. (Fig. S8). Considering the induction of RBP genes and the widespread differential AS genes under cold stress, we propose that cold-induced RBP genes likely play a role in posttranscriptional processes, including AS. **Associations between R-WGD-derived RBPs and coldinduced RBPs** As described in our prior study (Wang et al., 2023a,b), we sought to determine whether the retained and cold-induced RBP duplicates across different angiosperms are co-orthologs, that is a set of genes in one species orthologous to one or more genes in another species due to lineage-specific gene duplications (Kristensen et al., 2011). Using OrthoFinder (Emms & Kelly, 2019) coupled with phylogenetic analysis and manual verification, we identified 282 RBP orthogroups in the eight cold-treated species (see Materials and Methods section). Sorting them based the retention of R-WGD-derived RBP duplicates revealed 52 Highly Retained R-WGD-derived Orthogroups (HRROs), convergently retained from independent R-WGD events in at least six of the eight species (Fig. 5a). By contrast, 48 orthogroups, including 4, 11, and 33 in none or two fewer species, were designated as Lowly Retained R-WGD-derived Orthogroups (LRROs). Similarly, based on cold responsiveness, we categorized 108 orthogroups as Highly Conserved Cold-induced Orthogroups (HCCOs), induced in at least six species (Fig. 5b). Conversely, 61 orthogroups, including 11, 23, and 27 in none or two fewer species, were designated as Lowly Conserved Cold-induced Orthogroups (LCCOs).

Significantly, HRROs and HCCOs exhibited an overlap ($P = 3.6\ 9\ 10_{-3}$), indicating that highly retained R-WGDderived RBP orthogroups are mostly cold-induced in angiosperms (Fig. 5c). Conversely, LRROs and LCCOs also showed a significant overlap ($P = 2.9\ 9\ 10_{-5}$), suggesting that lowly retained R-WGD-derived RBP orthogroups are less coldinduced. The limited overlap between HRROs and LCCOs (or LRROs and HCCOs) further supports the association between R-WGD-derived RBP orthogroups and cold-induced RBP orthogroups. Between HRROs and HCCOs, the overlapping 29 orthogroups were associated with various RNA-binding domain families, with the well-known RRM-containing family being the most prominent, including seven orthogroups (GRP7/8, RBP45, RBP47, RSZ, PABN, CP29, and ALY; Fig. 5d; Table S19). Notably, all the 29 orthogroups exhibited a conserved upregulation pattern under cold stress across different angiosperm species (Fig. 5d). These results suggest that the highly retained R-WGDderived RBP duplicates and the highly conserved cold-induced RBP duplicates converge as co-orthologs across diverse angiosperms, highlighting a robust correlation between the retention of RBP duplicates and their cold responsiveness.

An example orthogroup: GRP7/8 retained from R1-WGD

Among the 29 orthogroups, certain well-known members, including Arabidopsis GRP7/8, also known as Cold and Circadian-Regulated 2 and 1 (CCR2/1), are recognized for their significant roles in responding to cold stress. GRP7/8 actively contribute to circadian timekeeping by orchestrating the posttranscriptional processing of a group of transcripts (Kojima et al., 2011; Streitner et al., 2012). Our analysis unveiled that Arabidopsis GRP7 was duplicated from GRP8 by the b-WGD event (R1-WGD in our study) before the divergence of Brassicaceae and Cleomaceae (Cheng et al., 2013), supported by the phylogenetic analysis and collinear orthologous analyses (Fig. S9). Phylogenetic analysis demonstrated that their orthologous counterparts in other species originated independently from R1-WGD events around the K-Pg boundary (Fig. 5e), corroborated by collinear orthologous relationship and Ks analysis (Fig. 5f; Table S20). Moreover, all co-orthologous GRP duplicates exhibited conserved cold-induced patterns in eudicots and monocots (Fig. 5g).

A role of R1-WGD-derived GRP7 in cold response

In this section, we chose R1-WGD-derived GRP7 as an example and explored its potential role in Arabidopsis under cold stress. First, we obtained 858 GRP7 genome-wide binding transcripts from iCLIP-seq (Meyer et al., 2017). GO analysis showed that GRP7-binding transcripts are enriched in abiotic and light stimuli responses, especially in cold response (Fig. 6a). Transcriptome analysis demonstrated that a substantial number (433) of GRP7 targets are cold-induced (Fig. 6b). These targets exhibited three distinct expression patterns across three time points (2, 24, and 168 h) of cold stress (Fig. 6c) and their average expression displayed a significant positive Pearson correlation (R_2 = 0.71, P = 5.9 9 10 _4) with GRP7 expression (Fig. S10). The findings imply that the binding of GRP7 to the 433 targets might be critical for their expression.

Moreover, by exploring transcriptome changes in GRP7- overexpressing and/or loss-of-function plants (Meyer et al., 2017), we confirmed that 294 of the 433 cold-induced GRP7 targets are differentially regulated, strongly suggesting that GRP7 directly regulates these cold-induced genes with high confidence (Fig. S11). Intriguingly, among the 433 targets, only 34 are differentially alternatively spliced genes under cold induction (Fig. S12). Considering the reported role of GRP7 in alternative polyadenylation (Meyer et al., 2017), these findings suggest that GRP7 plays a role in regulating the expression of cold-induced genes not only through alternative splicing but also through other posttranscriptional processes, such as alternative polyadenylation and RNA stability involved in cold stress.

Co-retention of RBP and TF duplicates and their role in cold response

The underlying transcriptional regulation of the 433 coldinduced transcripts is likely orchestrated by TFs. To determine these TFs, we assembled a library of 540 Arabidopsis TFs, subjected to experimental validation for their genome-wide binding profiles using DAP-seq (O'Malley et al., 2016) and ChIP-seq (Zhang et al., 2013; Nagel et al., 2015; Kamioka et al., 2016). Concurrently, we identified 108 cold-induced TFs out of the pool (Fig. 6d; Table S12). Subsequently, analyzing Arabidopsis genome-wide binding profiles of the 108 TFs, we identified the top 30 TFs with enriched binding peaks in the upstream promoters (1-kb) of the 433 cold-induced genes (Fig. 6e; Table S13). For example, LHY and CCA1, well-known TFs in cold stress and circadian rhythm (Kidokoro et al., 2022), bind to 81 and 68 of the 433 genes, respectively. Further examples include CBF1, CBF2, and CBF3, well-documented regulators in cold tolerance, binding to 31, 35, and 27 of the 433 genes, respectively (Fig. 6e). Compared with the 433 cold-induced genes (Fig. 6c), the 30 TFs show an earlier response to cold induction (Fig. S13), likely suggesting a hierarchical regulatory network of the TFs and the downstream 433 targets.

Moreover, numerous cold-induced genes within the 433 targets, such as COR27, COR15A, KIN1, and RD29A, have been experimentally confirmed to be upregulated by LHY, CCA1, CBF1, CBF2, and/or CBF3 (Novillo et al., 2007; Wang & Hua, 2009; Dong et al., 2011; Wang et al., 2017). Importantly, akin to the R1-WGD-derived GRP7 discussed earlier (Fig. 5e), the major of these cold-induced TFs also emerged as duplicates from R1-WGD events during global cooling periods (Fig. 6f). This convergence underscores the co-retention of cold-responsive RBP duplicates and TF duplicates following R-WGDs

across different angiosperm lineages. The findings demonstrate a pivotal role for Arabidopsis TF duplicates derived from R1-WGD in upregulating the transcriptional expression of the 433 coldinduced genes, which are subsequently targeted by R1-WGDderived GRP7 for RNA processing at the posttranscriptional level. This suggests a rewiring of circadian and cold-regulatory networks at both transcriptional and posttranscriptional levels in Arabidopsis following R1-WGD during global cooling epochs.

Discussion

During periods of global cooling, notably around the K-Pg boundary and in the Late Cenozoic Icehouse, multiple independent WGD events occurred across different angiosperm lineages. Subsequent to these WGD events, a convergent retention of stress-related genes occurred (Song et al., 2020; Wu et al., 2020; Guo et al., 2022). While previous research has predominantly focused on TF gene families to elucidate the transcriptional rewiring of gene regulatory networks in response to challenging environments (Song et al., 2020; Wu et al., 2020; Guo et al., 2022; Nie et al., 2022), it is imperative to recognize the significance of posttranscriptional regulation in plants adapting to challenging environments.

In this study, a category of RBP genes was identified. It was demonstrated that, beyond stress-related TF genes, duplicates of RBPs involved in cold responses exhibited convergent retention across diverse angiosperm lineages following independent WGD events. These RBP duplicates primarily originated from one or two relatively recent rounds of WGD events occurring around the K-Pg boundary and/or in the Late Cenozoic Icehouse. Despite the increasing loss of RBP genes as time progresses (Fig. S14), their retention after WGD remained significantly higher compared with that of non-RBP genes in angiosperms (Fig. 2b). Within the pool of RBP duplicates, 29 RBP orthogroups were identified, exhibiting both high retention after R-WGD events and high conservation of cold induction across diverse angiosperm lineages.

To date, numerous components, including messenger molecules, protein kinases and phosphatases, and TFs, have been identified in the signaling pathways responding to cold stress (Kidokoro et al., 2022). Among these, CBFs are particularly well-characterized to play as master regulators in the CBF-COR signaling pathway. The expression of CBF is modulated by many factors, including the circadian clock or circadian clock-related TFs such as the core oscillator components REVEILLES (RVEs) and CCA1/LHY (Kidokoro et al., 2022; Guo et al., 2023). The circadian clock system plays a crucial role in governing daily energy expenditure in response to environmental fluctuations, affording plants heightened tolerance to abiotic stresses essential for survival, growth, and reproduction (Greenham & McClung, 2015; Markham & Greenham, 2021). Analogous to the retention of RBP duplicates from R-WGD events, specific cold- and/or circadian-related TFs (e.g. CBF2/4, CCA1/LHY, and RVE4/8) were convergently retained across different angiosperms following independent R-WGDs during periods of global cooling (Fig. 6f). Of the R-WGD-derived RBP duplicates, GRP7/8 serve as cold- and circadian-regulated slave oscillators, subject to negative control by CCA1/LHY expression (Schmal et al., 2013; Johansson & Koster, 2019). Our findings indicate that numerous GRP7 target transcripts are transcriptionally regulated by cold- and/or circadian-related factors, including, but not limited to, R1-WGD-derived CBF2/4, RVE4/8, and CCA1/LHY. Notably, some downstream targets of RBPs and TFs were also convergently retained from R-WGD (e.g. SPA3/4, LEA2/5, and ABI1/2) or duplicated through subsequent TD events (e.g. RD29A/B, COR15A/B, and KIN1/2) during similar periods of global cooling (Fig. S15). This convergence suggests a rewiring of the regulatory network through both upstream and downstream components.

Therefore, we integrated these interconnected genes retained from R-WGD, illustrating the complexity and stability of the current circadian and cold-tolerance regulatory network, which is cooperatively regulated at both transcriptional and posttranscriptional levels (Fig. 7a). By contrast, for comparison, we inferred an ancestral network preceding R-WGD. Among RWGD- derived duplicates, one was deleted, while the other was retained, resulting in the inferred ancestral network (Fig. 7b). This inferred ancestral network was a simplified form, exhibiting fragility susceptible to damage upon mutation or gene loss within the network. This analysis underscores a critical role of R-WGD in reshaping,

enhancing, and adapting the circadian and cold-tolerance regulatory network in the context of global cooling.

Notably, in this study, we focused on R-WGD events, but did not explore ancient WGD events (i.e. f/e/c/r/s-WGD) in depth. The primary challenge in delving into ancient events arises from the inherent difficulty in accurately predicting their occurrences due to an extensive evolutionary history. For example, eand f-WGD events were inferred to have occurred c. 192 and c. 319 Ma, respectively (Jiao et al., 2011), but have been contested (Ruprecht et al., 2017). The prolonged evolutionary trajectory not only led to a significantly lower number of duplicates retained from ancient WGDs but also resulted in their divergent expressions and functions. Additionally, obtaining corresponding paleoclimate data for these ancient events is challenging (Scotese et al., 2021; Shaviv et al., 2023). These factors collectively hinder our assessment of the correlation between gene retention postancient WGDs and paleoclimate changes. Furthermore, it is plausible that ancient WGDs occurred during periods of global cooling. For example, c- and r-WGD events (c. 120 Ma) might coincide with the Aptian–Albian Cold Snap (Scotese et al., 2021). This assumption is not so far-fetched, as there is support for higher ratios of polyploids in colder climates (Schinkel et al., 2016; Rice et al., 2019; David, 2022). Additionally, cold treatments increase the likelihood of producing diploid male gametes (De Storme et al., 2012).

In summary, our research provides valuable insights into the convergent retention of RBP duplicates and TF duplicates in angiosperms following independent WGDs for cooperatively rewiring gene regulatory networks at both transcriptional and posttranscriptional levels, offering novel perspectives on the role of WGDs in plant adaptation under harsh climate changes.

Acknowledgements

We thank the members of Dr Wenwu Wu's laboratory at Zhejiang A&F University for suggestions to improve the quality of the study. This work is supported by the National Natural Science Foundation of China (grant no.: 31871233). YVdP acknowledges funding by Ghent University (Methusalem funding, BOF.MET.2021.0005.01) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no.: 833522). Competing interests None declared.

Author contributions

WW and LG conceived the idea for the project. WW, YVdP and YL managed the project. WW, LG and SW developed methods. LG, XJ, SW, XY, HL and DD performed the analysis. WW and LG wrote the draft manuscript. WW, YVdP and YL reviewed and edited the manuscript. LG, SW and XJ contributed equally to this work.

ORCID

Liangyu Guo https://orcid.org/0009-0008-2379-4663 Yan Li https://orcid.org/0000-0002-1503-0515 Yves Van de Peer https://orcid.org/0000-0003-4327-3730 Wenwu Wu https://orcid.org/0000-0003-2996-8675 Xiaoxue Ye https://orcid.org/0000-0001-7818-299X

Data availability

The RNA-seq datasets of eight plants (Arabidopsis thaliana, Carya illinoinensis, Glycine max, Populus trichocarpa, Oryza sativa, Phyllostachys edulis, Setaria italica, and Zea mays) under cold stress have been deposited in the NCBI BioProject (Accession: PRJNA767196). The programming code for identifying Pfam domains, classifying duplication modes, calculating Ks, assessing retention differences of RBP genes and other gene families, and drawing relative expression heatmaps was included in Dataset S1.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. Ast G. 2004. How did alternative splicing evolve? Nature Reviews Genetics 5: 773–782.

Bach-Pages M, Homma F, Kourelis J, Kaschani F, Mohammed S, Kaiser M, van der Hoorn R, Castello A, Preston GM. 2020. Discovering the RNA-binding proteome of plant leaves with an improved RNA interactome capture method. Biomolecules 10: 661.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422: 433–438.

Calixto C, Guo W, James AB, Tzioutziou NA, Entizne JC, Panter PE, Knight H, Nimmo HG, Zhang R, Brown J. 2018. Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome. Plant Cell 30: 1424–1444.

Carretero-Paulet L, Van de Peer Y. 2020. The evolutionary conundrum of whole-genome duplication. American Journal of Botany 107: 1101–1105.

Chaudhary S, Jabre I, Reddy A, Staiger D, Syed NH. 2019. Perspective on alternative splicing and proteome complexity in plants. Trends in Plant Science 24: 496–506.

Cheng S, van den Bergh E, Zeng P, Zhong X, Xu J, Liu X, Hofberger J, de Bruijn S, Bhide AS, Kuelahoglu C et al. 2013. The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. Plant Cell 25: 2813–2830.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6: 836–846.

David KT. 2022. Global gradients in the distribution of animal polyploids. Proceedings of the National Academy of Sciences, USA 119: e2080897177.

De Smet R, Sabaghian E, Li Z, Saeys Y, Van de Peer Y. 2017. Coordinated functional divergence of genes after genome duplication in Arabidopsis thaliana. Plant Cell 29: 2786–2800.

De Storme N, Copenhaver GP, Geelen D. 2012. Production of diploid male gametes in Arabidopsis by cold-induced destabilization of postmeiotic radial microtubule arrays. Plant Physiology 160: 1808–1826.

Dong MA, Farre EM, Thomashow MF. 2011. Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in Arabidopsis. Proceedings of the National Academy of Sciences, USA 108: 7241–7246.

Du Y, Cao L, Wang S, Guo L, Tan L, Liu H, Feng Y, Wu W. 2023. Differences in alternative splicing and their potential underlying factors between animals and plants. Journal of Advanced Research, in press. doi: 10.1016/j.jare.2023.11.017.

Eddy SR. 2011. Accelerated profile HMM searches. PLoS Computational Biology 7: e1002195.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20: 238.

Fawcett JA, Maere S, Van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proceedings of the National Academy, USA 106: 5737–5742.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proceedings of the National Academy of Sciences, USA 102: 16176–16181.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual Review of Plant Biology 60: 433–453.

Friedman WE. 2009. The meaning of Darwin's 'abominable mystery'. American Journal of Botany 96: 5–21.

Greenham K, McClung CR. 2015. Integrating circadian dynamics with physiological processes in plants. Nature Reviews Genetics 16: 598–610.

Guo L, Wang S, Nie Y, Shen Y, Ye X, Wu W. 2022. Convergent evolution of AP2/ERF III and IX subfamilies through recurrent polyploidization and tandem duplication during eudicot adaptation to paleoenvironmental changes. Plant Communications 3: 100420.

Guo L, Xu Z, Wang S, Nie Y, Ye X, Jin X, Zhu J, Wu W. 2023. Integrative multi-omics analysis of three early diverged Rosid species reveals an ancient hierarchical cold-responsive regulatory network. Physiologia Plantarum 175: e13892.

Jain M, Misra G, Patel RK, Priya P, Jhanwar S, Khan AW, Shah N, Singh VK, Garg R, Jeena G et al. 2013. A draft genome sequence of the pulse crop chickpea (Cicer arietinum L.). The Plant Journal 74: 715–729.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97–100.

Johansson M, Koster T. 2019. On the move through time – a historical review of plant clock research. Plant Biology 21: 13–20.

Junium CK, Zerkle AL, Witts JD, Ivany LC, Yancey TE, Liu C, Claire MW. 2022. Massive perturbations to atmospheric sulfur in the aftermath of the Chicxulub impact. Proceedings of the National Academy of Sciences, USA 119: e2119194119.

Kamioka M, Takao S, Suzuki T, Taki K, Higashiyama T, Kinoshita T, Nakamichi N. 2016. Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED1 in the Arabidopsis circadian clock. Plant Cell 28: 696–711.

Katoh K, Standley DM. 2013. Mafft multiple sequence alignment software v.7: improvements in performance and usability. Molecular Biology and Evolution 30: 772–780.

Kidokoro S, Shinozaki K, Yamaguchi-Shinozaki K. 2022. Transcriptional regulatory network of plant cold-stress responses. Trends in Plant Science 27: 922–935.

Kim D, Langmead B, Salzberg SL. 2015. Hisat: a fast spliced aligner with low memory requirements. Nature Methods 12: 357–360.

Kojima S, Shingle DL, Green CB. 2011. Post-transcriptional control of circadian rhythms. Journal of Cell Science 124(Pt 3): 311–320.

Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. 2011. Computational methods for gene orthology inference. Briefings in Bioinformatics 12: 379–391.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Molecular Biology and Evolution 34: 1812–1819.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359.

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

Markham KK, Greenham K. 2021. Abiotic stress through time. New Phytologist 231: 40–46.
Meyer K, Koster T, Nolte C, Weinholdt C, Lewinski M, Grosse I, Staiger D. 2017. Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7. Genome Biology 18: 204.
Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E et al. 2015. The pineapple genome and the evolution of CAM photosynthesis. Nature Genetics 47: 1435–1442.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37: 1530–1534.
Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer E, Tosatto S, Paladin L, Raj S, Richardson LJ et al. 2021. Pfam: the protein families database in 2021. Nucleic Acids Research 49(D1): D412–D419.
Nagel DH, Doherty CJ, Pruneda-Paz JL, Schmitz RJ, Ecker JR, Kay SA. 2015. Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis. Proceedings of the National Academy of Sciences, USA 112: E4802–E4810.
Nakaminami K, Seki M. 2018. RNA regulation in plant cold stress response. Advances in Experimental Medicine and Biology 1081: 23–44.
Nie Y, Guo L, Cui F, Shen Y, Ye X, Deng D, Wang S, Zhu J, Wu W. 2022. Innovations and stepwise evolution of CBFs/DREB1s and their regulatory networks in angiosperms. Journal of Integrative Plant Biology 64: 2111–2125.
Novillo F, Medina J, Salinas J. 2007. Arabidopsis CBF1 and CBF3 have a different function than CBF2 in cold acclimation and define different gene classes in the CBF regulon. Proceedings of the National Academy of Sciences, USA 104: 21002–21007.
O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. 2016. Cistrome and epicistrome features shape the regulatory DNA landscape. Cell 165: 1280–1292.
Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33: 290–295.
Plomion C, Aury JM, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonne N, Labadie K, Le Provost G et al. 2018. Oak genome reveals facets of long lifespan. Nature Plants 4: 440–452.
Pope KO, Baines KH, Ocampo AC, Ivanov BA. 1997. Energy, volatile production, and climatic effects of the Chicxulub Cretaceous/Tertiary impact. Journal of Geophysical Research 102(E9): 21645–21664.
Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biology 20: 38.
Reddy AS, Marquez Y, Kalyna M, Barta A. 2013. Complexity of the alternative splicing landscape in plants. Plant Cell 25: 3657–3683.
Reichel M, Liao Y, Rettel M, Ragan C, Evers M, Alleaume AM, Horos R, Hentze MW, Preiss T, Millar AA. 2016. In planta determination of the mRNA-binding proteome of Arabidopsis etiolated seedlings. Plant Cell 28: 2435–2452.
Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia- Hernandez M, Huala E, Lander G, Montoya M et al. 2003. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Research 31: 224–228.
Rice A, Smarda P, Novosolov M, Drori M, Glick L, Sabath N, Meiri S, Belmaker J, Mayrose I. 2019. The global biogeography of polyploid plants. Nature Ecology & Evolution 3: 265–273.
Rizzon C, Ponger L, Gaut BS. 2006. Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Computational Biology 2: e115.
Ruprecht C, Lohaus R, Vanneste K, Mutwil M, Nikoloski Z, Van de Peer Y, Persson S. 2017. Revisiting ancestral polyploidy in plants. Science Advances 3: e1603195.
Schinkel CC, Kirchheimer B, Dellinger AS, Klatt S, Winkler M, Dullinger S, Horandl E. 2016. Correlations of polyploidy and apomixis with elevation and associated environmental gradients in an alpine plant. AoB Plants 8: plw064.
Schmal C, Reimann P, Staiger D. 2013. A circadian clock-regulated toggle switch explains AtGRP7 and AtGRP8 oscillations in Arabidopsis thaliana. PLoS Computational Biology 9: e1002986.
Scotese CR, Song H, Mills BJW, van der Meer DG. 2021. Phanerozoic paleotemperatures: the earth's changing climate during the last 540 million years. Earth-Science Reviews 215: 103503.
Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends in Genetics 20: 461–464.
Shaviv NJ, Svensmark H, Veizer J. 2023. The Phanerozoic climate. Annals of the New York Academy of Sciences 1519: 7–19.
Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proceedings of the National Academy of Sciences, USA 111: E5593–E5601.
Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. American Journal of Botany 96: 336–348.
Song XM, Wang JP, Sun PC, Ma X, Yang QH, Hu JJ, Sun SR, Li YX, Yu JG, Feng SY et al. 2020. Preferential gene retention increases the robustness of cold regulation in Brassicaceae and other plants after polyploidization. Horticulture Research 7: 20.
Streitner C, Koster T, Simpson CG, Shaw P, Danisman S, Brown JW, Staiger D. 2012. An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in Arabidopsis thaliana. Nucleic Acids Research 40: 11240–11255.
Van de Peer Y, Ashman TL, Soltis PS, Soltis DE. 2021. Polyploidy: an evolutionary and ecological force in stressful times. Plant Cell 33: 11–26.
Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nature Reviews Genetics 18: 411–424.
Vanneste K, Baele G, Maere S, Van de Peer Y. 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. Genome Research 24: 1334–1347.

Vekemans D, Proost S, Vanneste K, Coenen H, Viaene T, Ruelens P, Maere S, Van de Peer Y, Geuten K. 2012. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. Molecular Biology and Evolution 29: 3793–3806.
Vellekoop J, Sluijs A, Smit J, Schouten S, Weijers JW, Sinninghe DJ, Brinkhuis H. 2014. Rapid short-term cooling following the Chicxulub impact at the Cretaceous-Paleogene boundary. Proceedings of the National Academy of Sciences, USA 111: 7537–7541.
Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics, Proteomics & Bioinformatics 8: 77–80.
Wang P, Cui X, Zhao C, Shi L, Zhang G, Sun F, Cao X, Yuan L, Xie Q, Xu X. 2017. COR27 and COR28 encode nighttime repressors integrating Arabidopsis circadian clock and cold response. Journal of Integrative Plant Biology 59: 78–85.
Wang S, Shen Y, Deng D, Guo L, Zhang Y, Nie Y, Du Y, Zhao X, Ye X, Huang J et al. 2023a. Orthogroup and phylotranscriptomic analyses identify transcription factors involved in the plant cold response: a case study of Arabidopsis BBX29. Plant Communications 4: 100684.
Wang S, Zhang Y, Ye X, Shen Y, Liu H, Zhao X, Guo L, Cao L, Du Y, Wu W. 2023b. A phylotranscriptomic dataset of angiosperm species under cold stress. Scientific Data 10: 399.
Wang Y, Hua J. 2009. A moderate decrease in temperature induces COR15a expression through the CBF signaling cascade and enhances freezing tolerance. The Plant Journal 60: 340–349.
Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H et al. 2012a. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research 40: e49.
Wang Y, Wang X, Paterson AH. 2012b. Genome and gene duplications and gene expression divergence: a view from plants. Annals of the New York Academy of Sciences 1256: 1–14.
Wu S, Han B, Jiao Y. 2020. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. Molecular Plant 13: 59–71.
Yu G, Wang LG, He QY. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31: 2382– 2383.
Zhang Y, Mayba O, Pfeiffer A, Shi H, Tepperman JM, Speed TP, Quail PH. 2013. A quartet of PIF bHLH factors provides a transcriptionally centered signaling hub that regulates seedling morphogenesis through differential expression-patterning of shared target genes in Arabidopsis. PLoS Genetics 9: e1003244.
Zhang Z, Boonen K, Ferrari P, Schoofs L, Janssens E, van Noort V, Rolland F, Geuten K. 2016. UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. Plant Methods 12: 42.
Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L. 2012. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochemical and Biophysical Research Communications 419: 779–781.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

FIGURES

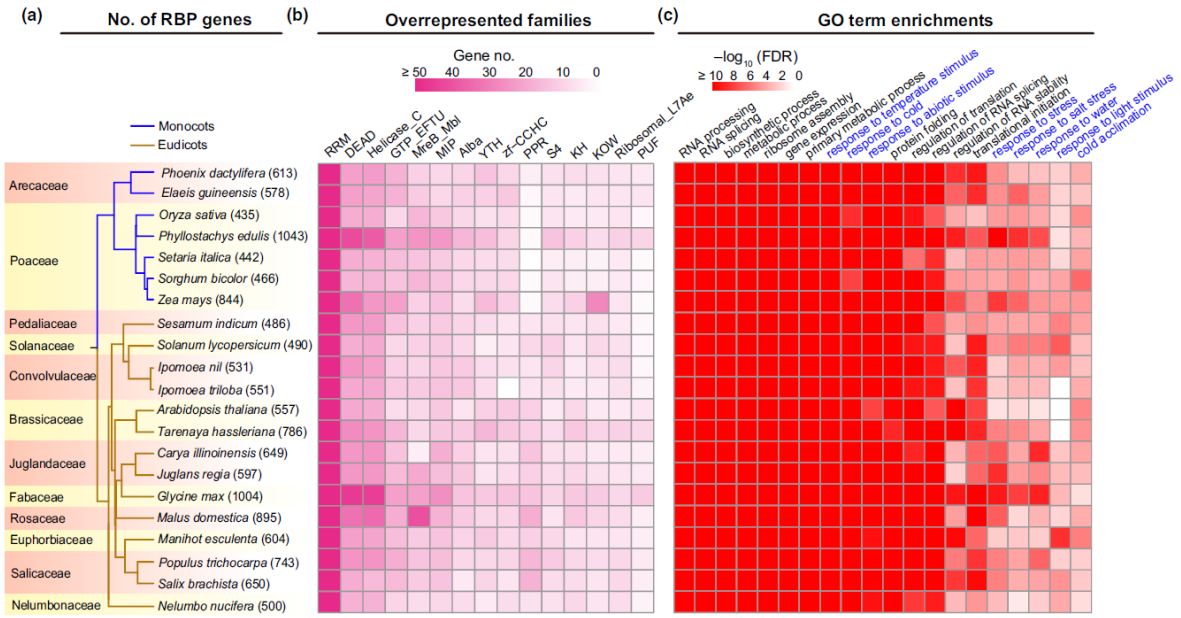

Fig. 1 Genome-wide identification of angiosperm RNA-binding protein (RBP) genes, overrepresented families, and functional roles. (a) Genome-wide identification of RBP genes in 21 selected angiosperms. Two colored branches of the tree indicate the two major lineages (eudicots and monocots) of angiosperms. The number in parentheses after the species indicates the identified number of RBP genes in the genome (Supporting Information Table S3). The workflow for RBP identification is illustrated in Fig. S1. (b) Overrepresented gene families of RBP genes. The top 15 occurrences of RBP gene families are shown and the RNA recognition motif (RRM)-containing family is the most prominent. (c) Gene Ontology (GO) term enrichments of RBP genes. The top 20 significant GO terms were selected to represent biological processes involving RBP genes. The terms in blue indicate enrichment in abiotic stress responses. The enrichment levels (*P*-value) were adjusted by the Benjamini–Hochberg false discovery rate (FDR) and scaled as enrichment score $-\log_{10}$ (FDR).

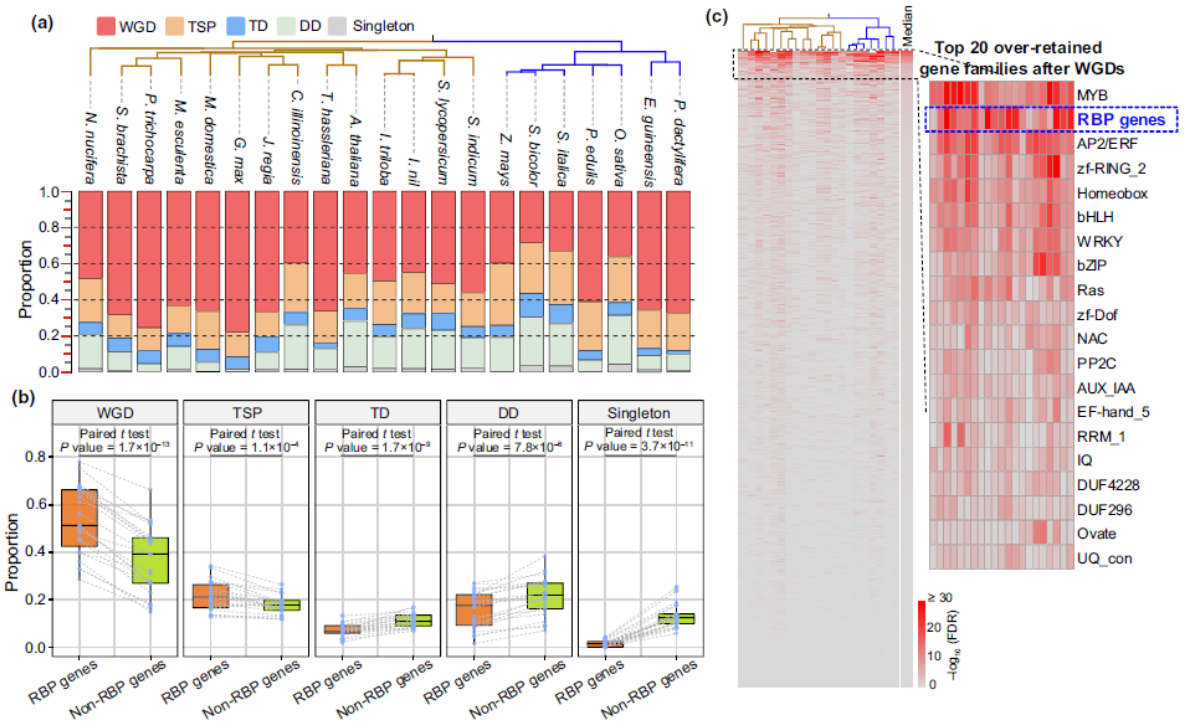

Fig. 2 Major source of RNA-binding protein (RBP) genes in angiosperms from WGDs. (a) Proportions of RBP genes from five duplication modes. DD, dispersed duplication; Singleton, single copy gene; TD, tandem duplication; TSP, transposed duplication; WGD, whole-genome duplication. The programs MCScanX (Wang *et al.*, 2012a) and DupGen_finder (Qiao *et al.*, 2019) were integrated to infer duplication modes of RBP genes (Supporting Information Table S3). This analysis was conducted based on the 21 selected angiosperms, including *Elaeis guineensis*, *Phoenix dactylifera*, *Oryza sativa*, *Phyllostachys edulis*, *Setaria italica*, *Sorghum bicolor*, *Zea mays*, *Sesamum indicum*, *Solanum lycopersicum*, *Ipomoea nil*, *Ipomoea triloba*, *Arabidopsis thaliana*, *Tarenaya hassleriana*, *Carya illinoinensis*, *Juglans regia*, *Glycine max*, *Malus domestica*, *Manihot esculenta*, *Populus trichocarpa*, *Salix brachista*, and *Nelumbo nucifera*. (b) Proportion differences between RBP and non-RBP genes resulting from the five duplication modes. The control group (non-RBP genes) comprised all coding genes, excluding RBP genes, for comparison with the test group (RBP genes) generated from each of the five duplication modes. The proportion of each duplication mode in generating the group genes was calculated relative to all genes in the group. A dashed line connects the same species between the test and control groups, and their difference was assessed by a two-sided paired *t*-test. Boxplots depict the median, interquartile range (IQR), and 1.5 × IQR (with outliers). (c) Retention differences in RBP genes and 494 gene-rich families in angiosperms after WGDs. The top 20 retained families (sorted by median retentions in 21 selected angiosperms) after WGDs are enlarged. Subsequent to the most over-retained MYB family, RBP genes are shown in a blue dashed box. The retention level of each family, compared with all coding genes in the genome after WGDs, was assessed by one-sided Fisher's exact test with Benjamini–Hochberg false discovery rate (FDR) correction, and its retention level was scaled as a retention score ($-\log_{10}$ (FDR)).

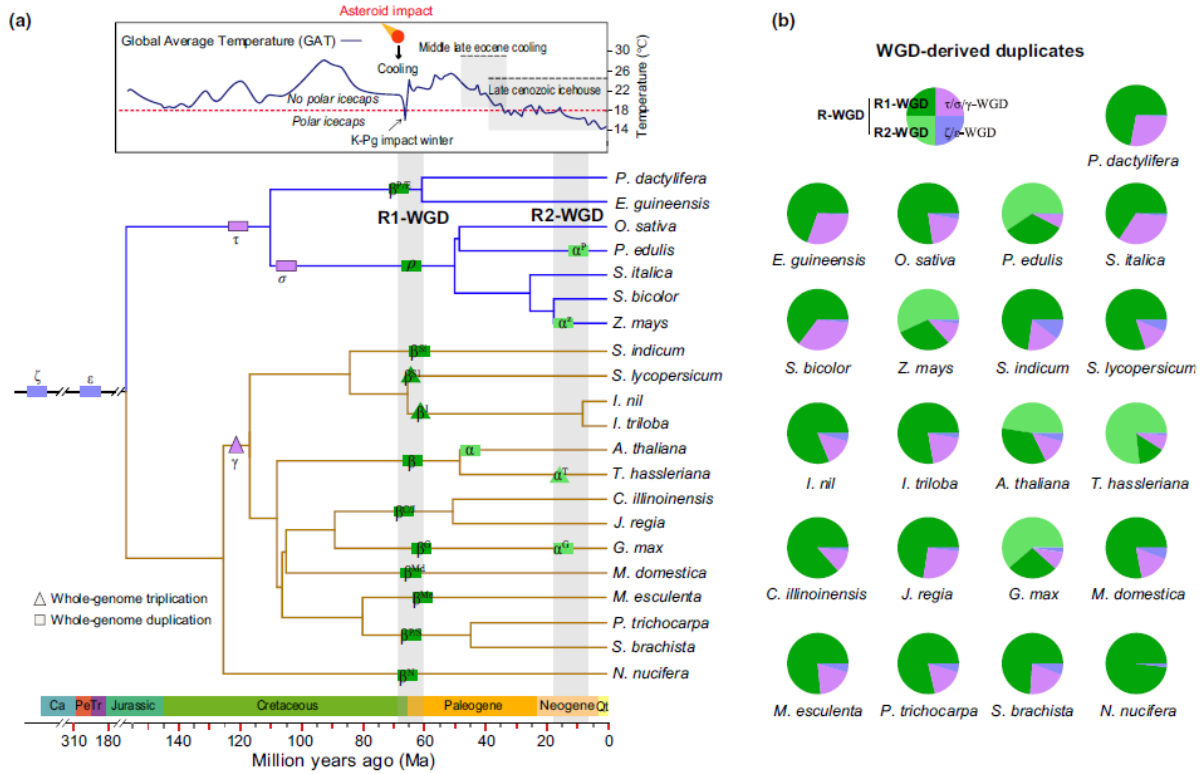

Fig. 3 Convergent retention of RNA-binding protein (RBP) duplicates in angiosperms after independent whole-genome duplication (WGD) events during periods of global cooling. (a) Historical occurrences of WGD events in 21 selected angiosperms correlated with paleotemperature changes. The upper curve graph displays inferred paleotemperature fluctuations in the past 160 Ma (Scotese *et al.*, 2021), including a sharp decline in global average temperature (GAT) around the K-Pg boundary (*c.* 66 Ma) and continuing GAT decline during the Middle–Late Eocene Cooling (*c.* 48.5–34 Ma) and Late Cenozoic Icehouse (*c.* 39.4 Ma to present). When the GAT was below 18°C (red dashed line), large polar icecaps were formed. The lower tree graph depicts species divergences and well-documented WGD events over geological times. In addition to ancient angiosperm-shared ζ/ε-WGD and monocot- or core eudicot-shared τ/σ/γ-WGD events, two more recent rounds of WGD events (R-WGD), including R1-WGD (*c.* 66 Ma around the K-Pg boundary) and R2-WGD (< 25 Ma in the Late Cenozoic Icehouse), occurred independently in different angiosperm lineages (supporting Information Table S6). (b) Predominant retention of RBP duplicates from R-WGDs, especially R1-WGD. The pie size indicates the proportion of RBP duplicates retained from different WGD periods. RBP duplicates derived from each WGD were inferred by a rectified process (Materials and Methods section). This analysis was conducted based on the 21 selected angiosperms, including *Elaeis guineensis*, *Phoenix dactylifera*, *Oryza sativa*, *Phyllostachys edulis*, *Setaria italica*, *Sorghum bicolor*, *Zea mays*, *Sesamum indicum*, *Solanum lycopersicum*, *Ipomoea nil*, *Ipomoea triloba*, *Arabidopsis thaliana*, *Tarenaya hassleriana*, *Carya illinoinensis*, *Juglans regia*, *Glycine max*, *Malus domestica*, *Manihot esculenta*, *Populus trichocarpa*, *Salix brachista*, and *Nelumbo nucifera*.

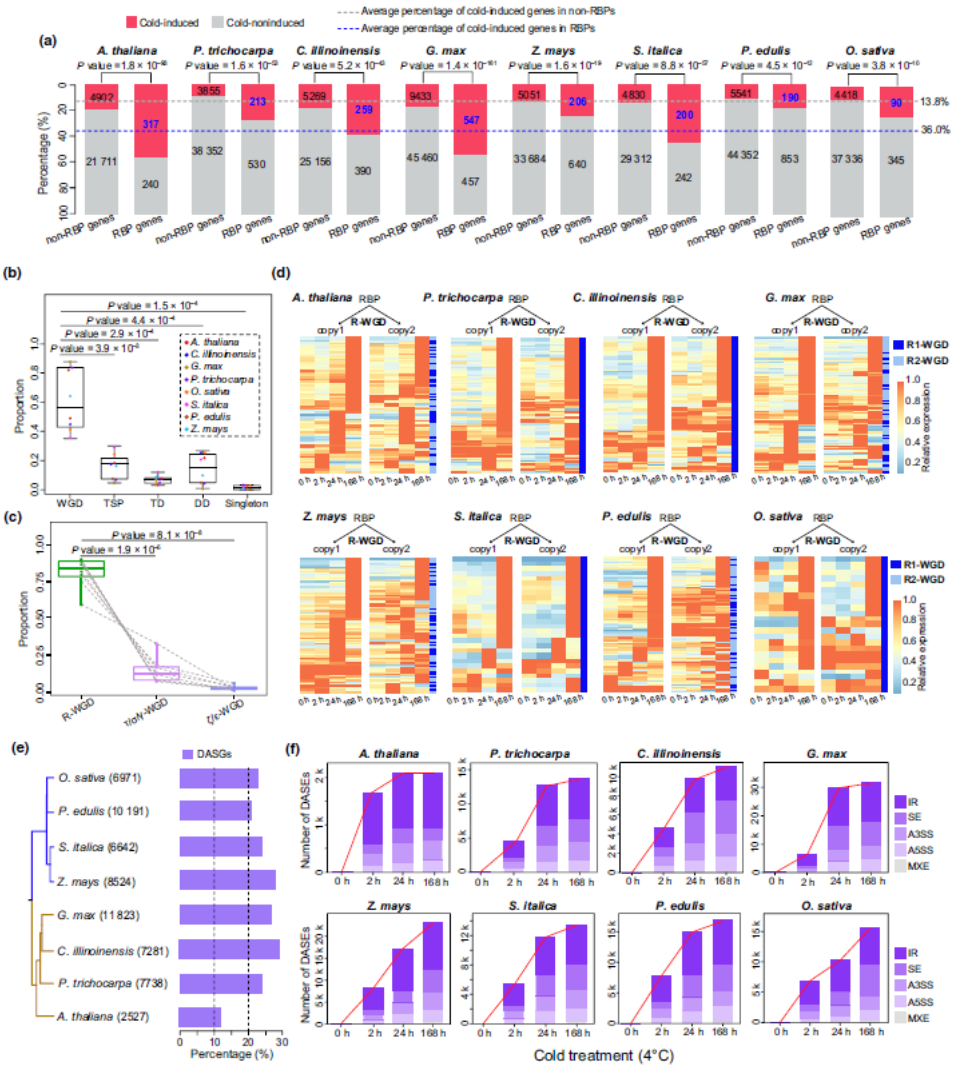

Fig. 4 Cold induction of R-WGD-derived RNA-binding protein (RBP) duplicates and modulation of thousands of alternative splicing genes. (a) A significant proportion of RBP genes display cold induction. Eight species (four eudicots: *Arabidopsis thaliana*, *Carya illinoinensis*, *Glycine max*, and *Populus trichocarpa*; and four monocots: *Oryza sativa*, *Phyllostachys edulis*, *Setaria italica*, and *Zea mays*) were simultaneously sampled at noon on the same day, at similar developmental stages, for cold-treated experiments (4°C). Cold-induced genes, identified with false discovery rate (FDR) < 0.01 and fold change > 1.5 under at least one condition with expression exceeding 20, were assessed for difference using Fisher's exact test between RBP and non-RBP genes with and without cold induction. (b) Proportion differences among the five duplication models in producing cold-induced RBP genes. Differences in cold-induced RBP genes from whole-genome duplication (WGD) and other duplication models were evaluated using a two-sided paired *t*-test. Boxplots depict the median, interquartile range (IQR), and 1.5 × IQR (with outliers). (c) Proportion differences among different periods of WGDs in producing cold-induced RBP duplicates. Differences in cold-induced RBP genes from R-WGD and other WGDs were evaluated using a two-sided paired *t*-test. Dashed lines connect the same species in the compared groups. Boxplots depict the median, interquartile range (IQR), and 1.5 × IQR (with outliers). (d) Expression heatmaps of cold-induced RBP duplicates from R-WGD events. Each pair of R-WGD-derived RBP duplicates is shown, with one randomly assigned as copy1 and the other as copy2. Relative expression indicates the deviation of gene expression at each time point from its highest level. (e) Thousands of cold-induced differentially alternatively spliced genes (DASGs). The number in parentheses indicates the count of cold-induced DASGs, with the percentage bar representing their proportion in multiple-exon genes of the species. (f) Number distribution of cold-induced differentially alternatively spliced events (DASEs). A5SS/A3SS, alternative 5′/3′ splicing site; IR, intron retention; MXE, mutually exclusive exons; SE, skipping exon. The types of DASEs for each gene under cold treatments are listed in Supporting Information Table S9.

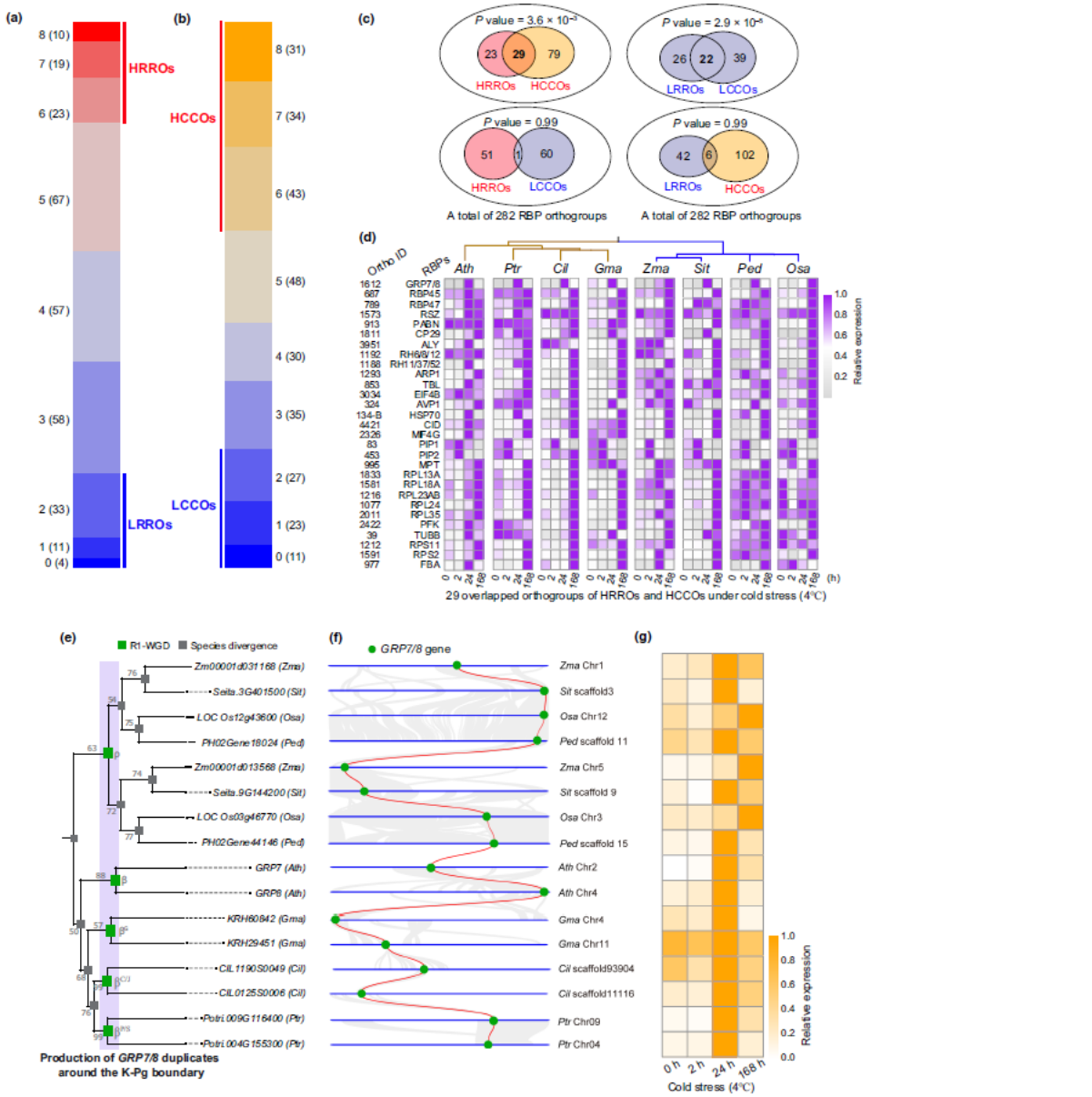

Fig. 5 Retained and cold-induced RNA-binding protein (RBP) duplicates are statistically associated within orthogroups. (a) Highly/Lowly Retained R-WGD-derived Orthogroups (HRROs/LRROs) of RBP genes. The numbers on the left indicate the count of retained R-WGD-derived orthogroups in specified species among the eight cold-treated species. For example, 6(23) means 23 orthogroups retained in six species after R-WGD. Orthogroups retained in seven or more species (or two or less species) after R-WGD were designated as HRROs (or LRROs). (b) Highly/Lowly Conserved Cold-induced Orthogroups (HCCOs/LCCOs) of RBP genes. The numbers on the right indicate the count of cold-induced orthogroups in specified species among the eight cold-treated species. For example, 5(48) means that 48 orthogroups are cold-induced in five species. Orthogroups that were cold-induced in seven or more species (or two or fewer species) were designated as HCCOs (or LCCOs). (c) Statistical analyses of the overlap between LCCOs, LRROs, HCCOs, and HRROs. The significance of overlap was assessed by Fisher's exact test. (d) Conserved cold-induced expression profiles of 29 overlapping orthogroups from HCCOs and HRROs. Orthogroup expression is represented by the average expression value of its containing genes. Relative expression indicates the deviation of orthogroup expression at each time point from its highest level. (e–g) An overlapping orthogroup example of HRROs and HCCOs in angiosperms. In the left panel (e), shading across the phylogenetic tree shows glycine-rich RNA-binding protein (*GRP*) duplicates from independent R1-WGD events around the K-Pg boundary. The number on branches indicates the bootstrap value for 1000 replicates. In the middle panel (f), genomic collinearity analysis supports whole-genome duplication (WGD)-derived *GRP* duplicates and their retention in angiosperms. Green nodes indicate corresponding collinear *GRP* genes on different chromosomal fragments, with average *Ks* values supporting WGD events around the K-Pg boundary (Supporting Information Table S20). In the right panel (g), a heatmap exhibits relative expression for *GRP* genes across different time points under cold stress. Ath, *Arabidopsis thaliana*; Ptr, *Populus trichocarpa*; Cil, *Carya illinoinensis*; Gma, *Glycine max*; Zma, *Zea mays*; Sit, *Setaria italica*; Ped, *Phyllostachys edulis*; Osa, *Oryza sativa*.

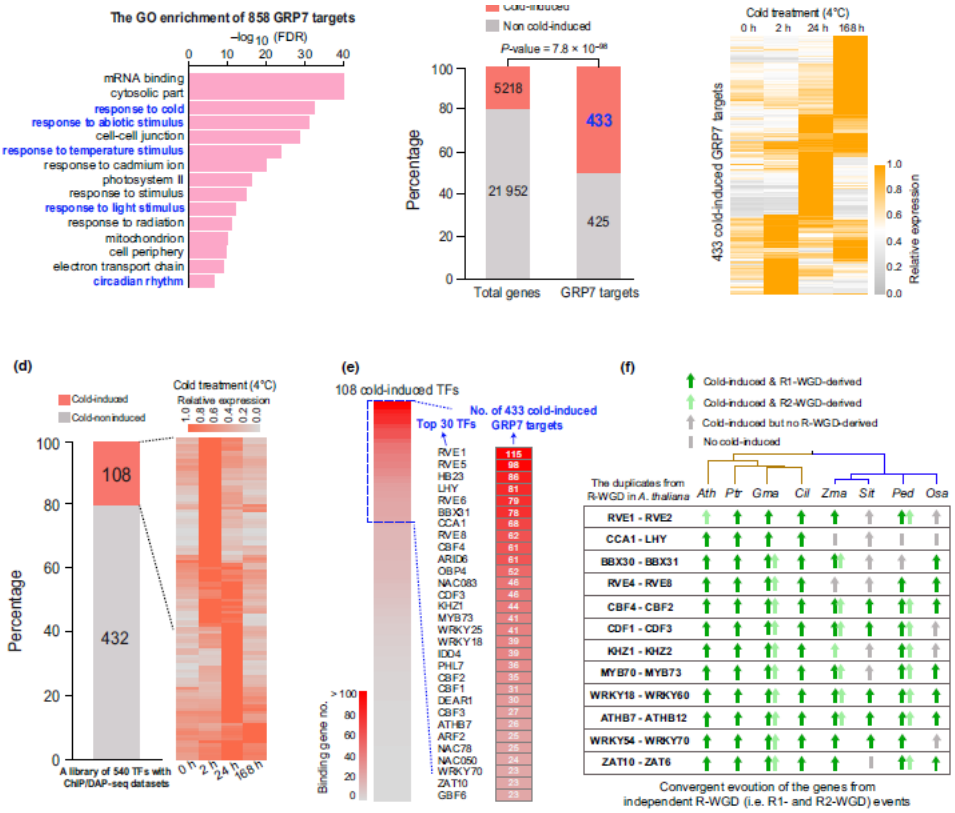

Fig. 6 Posttranscriptional and transcriptional regulation of 433 cold-induced transcripts by R1-WGD-derived glycine-rich RNA-binding protein 7 (GRP7) and transcription factor (TF) duplicates in *Arabidopsis*. (a) Gene Ontology (GO) term enrichments of *Arabidopsis* 858 GRP7-binding transcript genes. The terms of abiotic responses are highlighted in blue. The 858 transcripts were identified through an *Arabidopsis* genome-wide iCLIP-seq experiment (Meyer *et al*., 2017). (b) A high proportion of GRP7 targets are cold-induced. Difference between GRP7 binding genes and *Arabidopsis* protein-coding genes with and without cold induction was assessed by Fisher's exact test. (c) Expression profile of 433 cold-induced genes from GRP7 targets. Three main gene expression clusters were identified at different time points (2, 24, and 168 h) under cold stress. Relative expression indicates the deviation of gene expression at each time point from its highest expression. (d, e) Cold-induced TFs involved in regulating the expression of 433 cold-induced GRP7 targets. The expression heatmap exhibits 108 cold-induced TFs from a library of 540 TFs (d). The order is decreasing based on the number of 433 GRP7 targets bound by the 108 cold-induced TFs (e). For instance, 'RVE1 115' indicates that RVE1 can bind to 115 of the 433 cold-induced GRP7 targets. The zoom-in view shows the top 30 TFs binding to the 433 GRP7 targets, with at least one binding peak in their promoters (1-kb) analyzed from DAP/ChIP-seq. (f) Convergent retention of the TF duplicates in *Arabidopsis* and other angiosperms after independent R-WGDs. The TF duplicates retained from R-WGDs are highly conserved in cold induction in eight cold-treated species. Ath, *Arabidopsis thaliana*; Ptr, *Populus trichocarpa*; Cil, *Carya illinoinensis*; Gma, *Glycine max*; Zma, *Zea mays*; Sit, *Setaria italica*; Ped, *Phyllostachys edulis*; Osa, *Oryza sativa*.

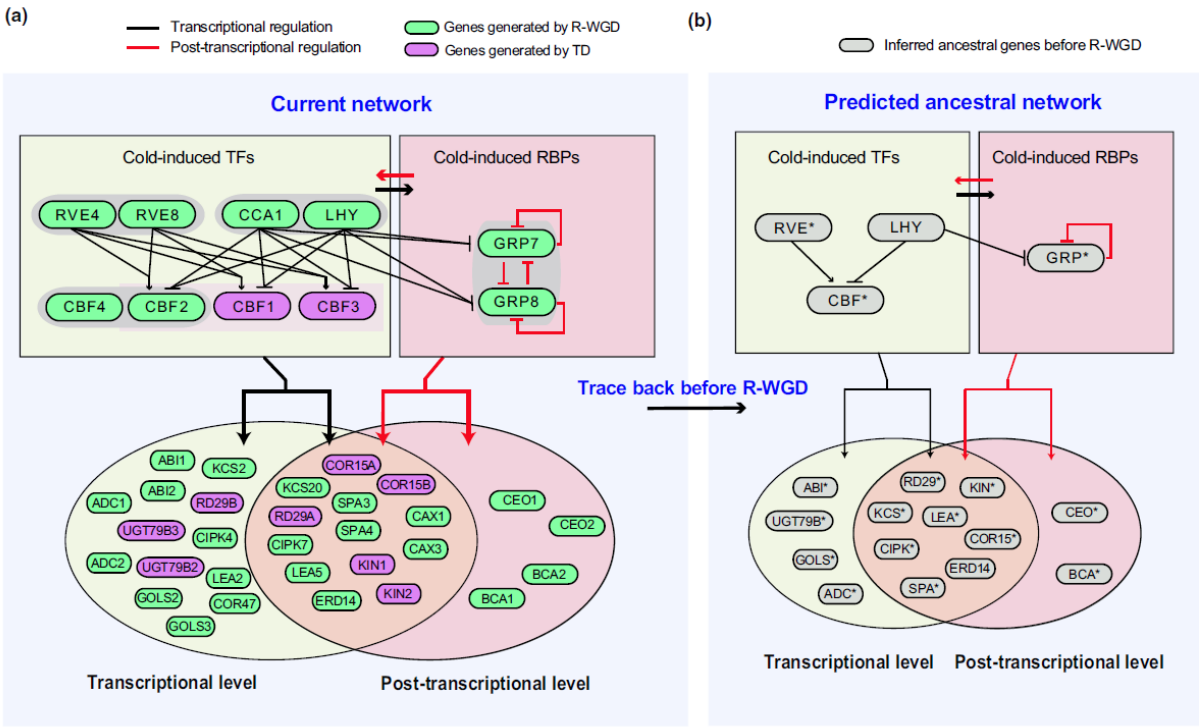

Fig. 7 Proposed model shows the impact of R-WGD on the circadian and cold-regulatory network at transcriptional and posttranscriptional levels. The current circadian and cold-regulatory network (a) and the inferred ancestral regulatory network before R-WGD (b). Green indicates R-WGD-derived duplicates, while purple denotes duplicates (e.g. *CBF1/3* and *COR15A/15B*) generated by tandem duplication events during global cooling in the Late Cenozoic Icehouse (Supporting Information Fig. S15). In the ancestral network, asterisks indicate ancestral forms of the duplicates.