Alternative Promoters Expand Transcriptional Complexity of Temperature Stress Responses in Cassava

Akihiro Ezoe[1,2*], Yoshinori Utsumi[3], Tetsuya Sakurai[4], Satoshi Takahashi[1], Tomoyuki Takeda[1], Daisuke Todaka[1], Yutaka Ogawa[1], June-Sik Kim[2,5], Junko Ishida[1], Maho Tanaka[1], Chikako Utsumi[1], Motoaki Seki[1,6,7,8*], and Keiichi Mochida[2*]

1: Plant Genomic Network Research Team, RIKEN Center for Sustainable Resource Science
2: Bioproductivity Informatics Research Team, RIKEN Center for Sustainable Resource Science
3: Faculty of Life Science and Biotechnology Department of Nutrition and Life Science, Fukuyama University
4: Multidisciplinary Science Cluster, Interdisciplinary Science Unit, Kochi University
5: Institute of Plant Science and Resources, Okayama University
6: Plant Epigenome Regulation Laboratory, RIKEN Cluster for Pioneering Research
7: Kihara Institute for Biological Research, Yokohama City University
8: Graduate School of Science and Engineering, Saitama University

Contact: akihiro.ezoe@riken.jp

Plant abiotic stress responses involve two major gene expression regulatory mechanisms: alternative promoter usage and differential expression. Although differentially expressed genes (DEGs) have been extensively studied, alternative promoter genes (APGs) remain poorly characterized despite their potential importance. We systematically compared APGs and DEGs in cassava (*Manihot esculenta*), a major crop tolerant to heat, but sensitive to cold. On the basis of a high-resolution transcription start site analysis (nAnT-iCAGE-seq), we identified 1,705 heat-responsive and 491 cold-responsive APGs as well as 4,172 heat-responsive and 582 cold-responsive DEGs. We identified three key distinctions. First, APGs contain fewer upstream open reading frames (uORFs) than DEGs and result in proteins with a truncated N-terminal, leading to altered subcellular localization, especially from plastids to the cytoplasm and from the cytoplasm to the nucleus. Second, while DEGs are highly conserved between cassava and Arabidopsis thaliana, the number of APGs increased in a lineage-specific manner through recent gene duplication events. Third, these duplications enabled the selective modification of cis-regulatory elements, contributing to increased APG expression (relative to DEG expression). Our findings suggest alternative promoters are essential components of the regulatory mechanism underlying the diversity in protein localization during plant stress responses, while also complementing well-characterized DEG-related stress responses.

Keywords: Transcription start site (TSS) shift, Temperature stress, Abiotic stress, Crop improvement, Alternative TSS, Alternative transcriptional initiation (ATI), CAGE

1. Introduction

Many genes possess multiple transcription start sites (TSSs), and the selective use of these TSSs enhances the diversity of mRNAs and proteins derived from a single genomic locus. This phenomenon is due to alternative promoters, although it does not necessarily imply the existence of multiple distinct promoter sequences [1]. Instead, it involves TSS shifts within the same gene, resulting in transcript isoforms with different 5′ ends [2,3]. This regulatory mechanism is particularly important for cellular responses to environmental stimuli because it enables rapid changes to gene expression patterns. In animals, many important findings regarding TSSs have been reported over the past three decades, including the well-known case of the tumor suppressor gene *p53* being transcribed from multiple TSSs [4,5].

The biological significance of alternative promoters in plant stress responses has gradually been revealed. Environmental conditions trigger TSS shifts in numerous genes; for example, light-induced photoinhibition in *Arabidopsis thaliana* (Arabidopsis) activates 2,104 alternative promoter genes (APGs) [1]. This implies that the importance of APGs in plants is similar to the well-established significance of differentially expressed genes (DEGs). In fact, on average, Arabidopsis genes have four active TSSs [6]. Moreover, current annotations in Arabidopsis and cassava reveal that approximately 33% and 20% of genes, respectively, possess multiple TSSs, reflecting the widespread potential for alternative promoter regulatio. These proportions are comparable to those in animals, in which 30% of genes are regulated by alternative promoters. Considering numerous studies have demonstrated the critical functions of alternative promoters in animals, plant alternative promoters likely have crucial roles in abiotic stress responses [7].

Multiple TSS isoforms driven by alternative promoters enhance transcriptional plasticity, thereby influencing translational efficiency, tissue or developmental specificity, and protein isoform diversity due to N-terminal variability. For example, Arabidopsis *GSTF8*, a glutathione S-transferase family member, has 10 TSS isoforms that respond to abiotic stress in distinct tissues [8]. These isoforms yield proteins with differing transit peptides, leading to altered subcellular localization after N-terminal truncation [8,9]. Such examples highlight alternative promoters as versatile regulators of conditional expression, organ specificity, protein diversity, and subcellular targeting. Given their regulatory effects, alternative promoters may play a crucial role in plant stress responses.

Despite the potential importance of alternative promoters, plant abiotic stress research remains mainly focused on DEGs because they are easily detected and abundant under abiotic stress conditions [10-13]. However, APGs may be as important, if not more so, as critical regulators of abiotic stress responses because mRNA isoform diversity may modulate expression levels, tissue specificity, translational efficiency, and protein functionality [1,14,15]. A key challenge for identifying APGs is that conventional RNA-seq cannot distinguish TSS-specific isoforms. Advances in cap analysis gene expression (CAGE)-based sequencing technologies have enabled researchers to precisely map TSSs at single-nucleotide resolution in plants; the identified APGs may be useful for addressing knowledge gaps in stress biology research that DEG-centric approaches have failed to fill [16,17]. This study was conducted to identify candidate APGs with critical functions in abiotic stress responses and differentiate their roles from those of DEGs.

To investigate APGs important for abiotic stress response, we focused on cassava (*Manihot esculenta*), a tropical crop exhibiting extreme stress tolerance traits. More specifically, it is highly resilient to heat stress, but is sensitive to cold (i.e., ceases to grow below 16 °C) [18-20]. By using a method for highly quantitative and positionally precise sequencing to comprehensively map APGs and DEGs identified under heat and cold stress conditions, we addressed the following three key questions: (1) What are the quantitative contributions and distinct functional properties of APGs and DEGs in

stress responses? (2) How does APG-mediated protein diversity contribute to post-transcriptional regulation, particularly through altered subcellular localization? (3) Are stress-responsive APGs evolutionarily conserved or expanded lineage-specific manner? The results of our comprehensive analysis challenges the current DEG-centric paradigm and suggest that alternative promoters are indispensable regulators that have expanded the molecular toolkit for plant stress adaptations.

2. Results

2.1 Transcriptional responses of APGs and DEGs to heat and cold stress conditions

To investigate the extent of the effects of heat and cold stresses on alternative promoters and differential expression, we constructed no-amplification non-tagging improved CAGE sequencing (nAnT-iCAGE-seq) libraries that were sequenced using Illumina platforms [21]. For the nAnT-iCAGE-seq analysis, we used RNA samples extracted from 1-month-old cassava plants incubated under normal, heat (42 °C), and cold (4 °C) conditions.

We mapped nAnT-iCAGE-seq reads to the cassava reference genome and confirmed that they were enriched at the annotated representative TSSs of individual gene. We detected 40,520 non-redundant TSS peaks, including peaks for 33,408 TSSs that were assigned to 6,012 genes with multiple TSSs. To identify heat- and cold-responsive APGs, we compared the ratios of nAnT-iCAGE sequencing reads between TSS pairs (Fig. 1a). Using this approach, we identified 1,705 heat- and 491 cold-responsive APGs (Fig. 1b). To identify heat- and cold-responsive DEGs, we compared the ratios of the total number of reads at TSSs assigned to each locus between normal and heat or cold conditions (Fig. 1a). The total number of reads at each locus was significantly correlated with gene expression levels determined in an earlier study using RNA-seq[22]. We identified 4,172 heat- and 582 cold-responsive DEGs (Fig. 1b). Notably, 397 genes were identified as both heat-responsive APGs and DEGs (23% of 1,705 heat-responsive APGs), while 29 genes were identified as both cold-responsive APGs and DEGs (6% of 491 cold-responsive APGs). This limited overlap implies APGs and DEGs have distinct regulatory roles during stress responses.

For a reference, we assessed the similarities in the transcriptional responses to heat and cold stresses. DEGs and TSS peaks in APGs were then classified as up- or down-regulated. Heat-responsive DEGs comprised 2,010 up- and 2,162 down-regulated genes, whereas cold-responsive DEGs included 456 up- and 128 down-regulated genes. Each APG contained both up- and down-regulated TSSs. Moreover, 146 APGs and 201 DEGs were identified as genes that were non-directionally responsive to both heat and cold stress conditions, of which 100 APGs and 164 DEGs were similarly affected by both stresses (i.e., up- or down-regulated). Assuming a 50% probability of similar responses to heat and cold stresses (e.g., up–up or down–down *vs* up–down or down–up), the observed overlaps significantly exceeded expectations [APG: 100/146 and DEG: 164/201 (expected values: 73 and 100.5); APG: $P = 1.00 \times 10^{-3}$, DEG: $P = 1.10 \times 10^{-11}$, chi-squared test].

As a contrasting reference, we further compared these abiotic stress-responsive DEGs with biotic stress-responsive DEGs identified in a previous study [23]. The previous study detected 3,017 and 3,889 DEGs in response to two biotic stress conditions in cassava. The proportion of DEGs with similar responses to heat/cold stresses (164/583 = 28%) was significantly higher than the proportion of DEGs with similar responses to abiotic/biotic stresses (435/3,017 = 14.4% and 490/3,889 = 12.6%) ($P = 3.41 \times 10^{-11}$ and 3.62×10^{-16}, chi-squared test), confirming shared regulatory mechanisms unique to temperature stress responses. This indicates that temperature stress responses in cassava were regulated by specific network distinct from the biotic stress responses.

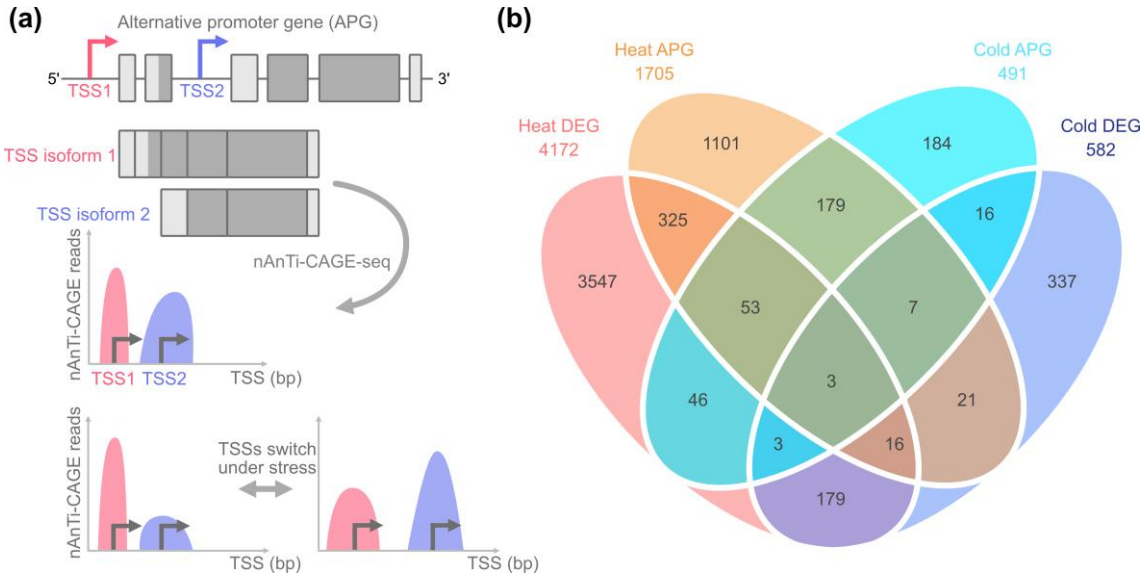

Fig. 1: Detection of alternative promoter genes (APGs) and identification of APGs and differentially expressed genes (DEGs) under heat or cold conditions.

(a) Schematic representation of how a single gene locus can switch between multiple TSSs, resulting in TSS isoforms (pre-mRNA variants) with distinct 5′ ends. Light gray boxes represent untranslated regions (UTRs), whereas dark gray boxes represent coding sequence (CDS) regions. The first exon transcribed from TSS2 can serve as either a UTR or part of the CDS, depending on the isoform. The lower panels show nAnT-iCAGE-seq read distributions, demonstrating TSS switching under stress conditions (with increased TSS2 usage indicating an up-regulated TSS).

(b) Venn diagram showing the number of APGs and DEGs identified under heat or cold conditions. Overlapping regions indicate genes that are shared across multiple categories.

2.2 TSS switching in stress-responsive APGs favors uORF-free isoforms

Upstream open reading frames (uORFs) in UTRs of transcripts affect translational efficiency through ribosome stalling [3], and a previous study showed that this phenomenon can be induced by alternative promoters [24]. Thus, heat- and cold-responsive alternative promoters can also switch TSS isoforms with and without uORFs. Therefore, we compared the proportions of transcripts with and without uORFs, which revealed that heat- and cold-responsive APGs had higher proportions of transcripts without uORFs [heat-responsive APGs: 29.9% (373/1,246), cold-responsive APGs: 27.4% (75/274)] than all cassava genes and expressed genes [all genes: 20.4% (6,676/26,129), expressed genes: 23.5% (5,027/21,404)] [Fig. 2; heat APG *vs* all: false discovery rate (FDR) = 7.42×10^{-16}, cold APG *vs* all: 8.15×10^{-3}, heat APG *vs* expressed: 2.57×10^{-16}, cold APG *vs* expressed: 4.12×10^{-2}, chi-squared test]. Moreover, heat- and cold-responsive APGs had higher proportions of transcripts without uORFs than heat- and cold-responsive up-regulated DEGs [heat: 27.2% (547/2,010), cold: 21.7% (99/455)] (Fig. 2; heat APG *vs* up-regulated DEG: FDR = 4.12×10^{-2}, cold APG *vs* up-regulated DEG: 4.12×10^{-2}, chi-squared test). These results suggest that heat- and cold-responsive TSS isoforms tend to lack uORFs. Because of the lack of uORF, these TSS isoforms are likely translated efficiently, which may lead to the substantial abundance of proteins associated with heat and cold stress response mechanisms.

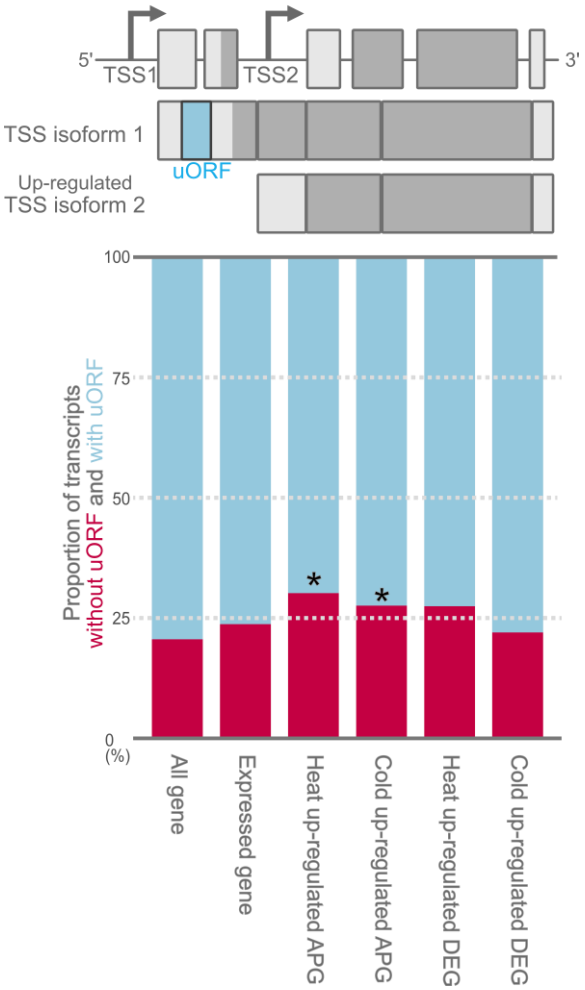

Fig. 2: Proportion of transcripts with and without uORFs for all genes, expressed genes, APGs, and DEGs in cassava under heat and cold stress conditions.

Schematic representation of alternative transcription start site (TSS) usage affecting the presence of an upstream open reading frame (uORF). Gray arrows indicate TSS1 and TSS2 positions. Dark and light gray boxes represent coding regions and untranslated regions, respectively. The blue box represents the uORF present in TSS isoform 1, but absent in up-regulated TSS isoform 2, demonstrating how alternative TSS usage can eliminate uORFs from transcripts. Bars represent the proportion of transcripts without uORFs (red) and with uORFs (blue) for all cassava genes, expressed genes, APGs up-regulated by heat and cold stresses, and DEGs up-regulated by heat and cold stresses. For all cassava genes and expressed genes, we identified 150 nt downstream from the representative TSS. For heat- and cold-responsive APGs, we identified 150 nt downstream from non-representative TSSs up-regulated under heat and cold stress conditions, respectively. For heat- and cold-responsive DEGs, we focused on up-regulated DEGs. Asterisks indicate significant differences (FDR <0.05). APGs tended to have TSS isoforms lacking uORFs.

2.3 Heat- and cold-responsive alternative promoter usage promotes the truncation of N-terminal transit peptides and the formation of nuclear localization signals

Since transit peptides often reside in the N-terminal region, where alternative promoter usage can induce truncation, we assessed whether such truncation alters subcellular localization of APGs in cassava under temperature stress [1,25]. Assuming that the annotated reading frames were unaffected, we identified 499 and 146 genes predicted to encode multiple protein isoforms with different N-terminals from among the 1,705 heat- and 491 cold-responsive APGs, respectively (Fig. 3a). To assess the localization tendencies of the proteins encoded by these genes, we compared the subcellular localization of representative proteins and protein isoforms (Fig. 3a) using DeepLoc 2 [26]. We focused on the plastid, cell membrane, lysosome/vacuole, cytoplasm, and nucleus, whereas endoplasmic reticulum, mitochondrion, Golgi apparatus, extracellular, and peroxisome were combined (i.e., "others") because relatively few proteins were localized in these cell organelles/regions. Representative proteins encoded by heat-responsive APGs tended to localize in the plastid, while representative proteins encoded by cold-responsive APGs tended to localize in the cell membrane or lysosome/vacuole [Fig. 3b; plastid: (heat) FDR = 2.03×10^{-15}, (cold) FDR = 8.33×10^{-8}, cell membrane: (heat) FDR = 2.09×10^{-2}, (cold) FDR = 4.10×10^{-3}, lysosome/vacuole: (heat) FDR = 4.02×10^{-6}, (cold) FDR = 8.74×10^{-4}, chi-squared test]. By contrast, protein isoforms encoded by heat- and cold-responsive APGs were more likely to be localized in the cytoplasm [Fig. 3b; cytoplasm: (heat) FDR = 4.55×10^{-13}, (cold) FDR = 2.50×10^{-9}, chi-squared test]. These results suggest that representative proteins tend to have transit peptides that result in the localization in the plastid, cell membrane, or lysosome/vacuole, while many protein isoforms have an N-terminal that lacks these transit peptides (i.e., truncated N-terminal).

On the basis of these findings, we hypothesized that many temperature-responsive APGs would produce representative proteins containing transit peptides, while their isoforms lack these sequences due to N-terminal truncation. To analyze these proteins, we examined the predicted subcellular localization changes between pairs of representative proteins and protein isoforms encoded by 293 heat- and 69 cold-responsive APGs. Because of the diverse combinations of predicted subcellular localizations and changes, we focused on the five most frequent changes (Fig. 3c). As expected, under heat stress conditions, the most common subcellular localization change was from the plastid to the cytoplasm. In response to cold stress, although subcellular localization changes involving the cell membrane and lysosome/vacuole were not among the five most common changes, we still detected many cytoplasm-localized isoforms. The most frequent change under cold conditions was from the plastid to the cytoplasm, confirming that the truncation of transit peptides in representative proteins due to alternative promoter usage leads to the localization of protein isoforms in the cytoplasm in both heat and cold stress responses.

Notably, the proportion of genes encoding nuclear proteins was higher for APGs than for all

cassava genes encoding representative proteins [Fig. 3a; 36.6% (APG) *vs* 31.0% (all), $P = 6.02 \times 10^{-5}$, chi-squared test]. This indicates that stress-responsive APGs may be regulated in terms of the localization of the encoded protein between the nucleus and other locations. This trend was observed for both heat- and cold-responsive APGs. Additionally, because there was no significant difference between representative proteins and protein isoforms, the proportion of nuclear localization signals (NLSs) exclusive to representative proteins or protein isoforms was almost the same. Therefore, we also speculated that there are many instances of representative proteins localized in the nucleus, while the corresponding protein isoforms are localized in the cytoplasm. As expected, subcellular localization changes between representative proteins and protein isoforms were in both directions between the nucleus and cytoplasm in response to heat and cold stresses. Considered together, these observations suggest that subcellular localization changes induced by temperature (i.e., heat and cold) stress-responsive alternative promoter usage are often associated with the truncation of transit peptides or the formation of NLSs.

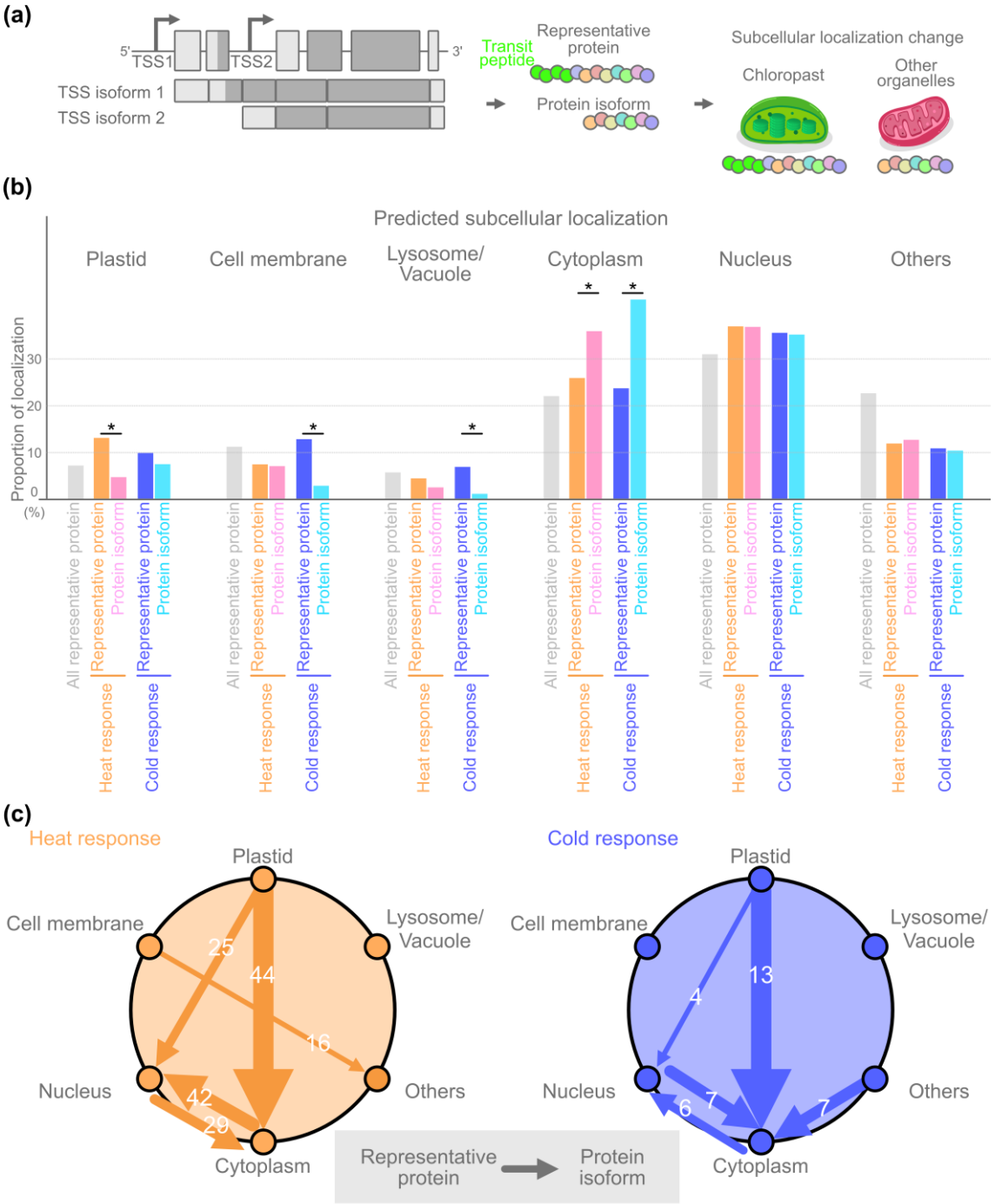

Fig. 3: Proportions of subcellular localizations and the directions of subcellular localization changes between representative proteins and protein isoforms.

(a) Schematic representation of how alternative transcription start sites (TSS1 and TSS2) can generate different transcript isoforms; TSS isoform 1 encodes a protein that contains a transit peptide leading to the localization to chloroplasts, while TSS isoform 2 encodes a protein lacking this transit peptide, resulting in the localization to other cellular compartments. Light gray boxes: untranslated regions, dark gray boxes: coding sequences/exons; green circles: transit peptide sequence, other colored circles:

amino acids.

(b) Subcellular localizations of all cassava representative proteins as well as heat- or cold-responsive representative proteins and protein isoforms are shown as bar plots. Gray bars indicate the subcellular localization of all representative proteins as a benchmark. Asterisks indicate a significant difference between representative proteins and protein isoforms (FDR <0.05, chi-squared test). (c) Directions of the subcellular localization changes between representative proteins and protein isoforms are shown. Arrows indicate the change in subcellular localization (i.e., from representative proteins to protein isoforms). The top five subcellular localization changes are presented; arrow thickness and the number on arrows reflect the frequency of the subcellular localization change. Endoplasmic reticulum, mitochondrion, Golgi apparatus, extracellular, and peroxisome are included in "Others." Plastid to cytoplasm was the most common subcellular localization change.

2.4 Distinct functions of APGs and DEGs

We performed a Gene Ontology (GO) enrichment analysis to functionally characterize heat- and cold-responsive APGs and DEGs, with an emphasis on APG-specific terms. Of the 3,913 GO terms, 27, 16, 54, and 43 were enriched among heat-responsive APGs, cold-responsive APGs, heat-responsive DEGs, and cold-responsive DEGs, respectively (Fig. 4). According to the significantly enriched GO terms, the heat- and cold-responsive APGs were related to similar processes, including transcription and gene expression, which were represented by the following GO terms: regulation of DNA-templated transcription (GO:0006355) and regulation of RNA metabolic process (GO:0051252). These results imply that APG transcription may be triggered in response to environmental stimuli. These terms were absent in DEGs, highlighting that a distinct role of APGs in transcriptional regulation during temperature stresses.

Interestingly, response to auxin (GO:0009733) was an enriched GO term among 31 heat-responsive APGs. Examples include the auxin-responsive genes Manes.11G033100 and Manes.04G134400, which encode proteins that negatively regulate auxin signaling; the degradation of these proteins activates ARF transcription factors that increase the expression of stress-responsive genes [27] (IAA26; https://pantherdb.org/). These genes, which were detected as heat-responsive APGs, are reportedly up-regulated genes in other plant species during responses to heat stress [28]. Intriguingly, heat stress activates an alternative promoter of *IAA26*, leading to the N-terminal truncation of the encoded protein. This protein isoform was predicted to be localized in the cytoplasm (the corresponding representative protein was localized in the nucleus). We speculated that cytosol-localized IAA26 cannot suppress downstream ARFs. This was supported by the fact earlier research showed that the transition of IAA26 from a nuclear protein to a cytosolic protein activates downstream ARFs in *Nicotiana benthamiana* [29,30]. Therefore, heat stress may induce auxin signaling because the nuclear localization of IAA26 is inhibited. Several studies demonstrated that the application of auxin or activation of auxin signals ameliorates heat stress-induced grain loss or developmental inhibition in other crops [31]. During the cassava response to heat stress, an alternative promoter may be critical for regulating auxin signals.

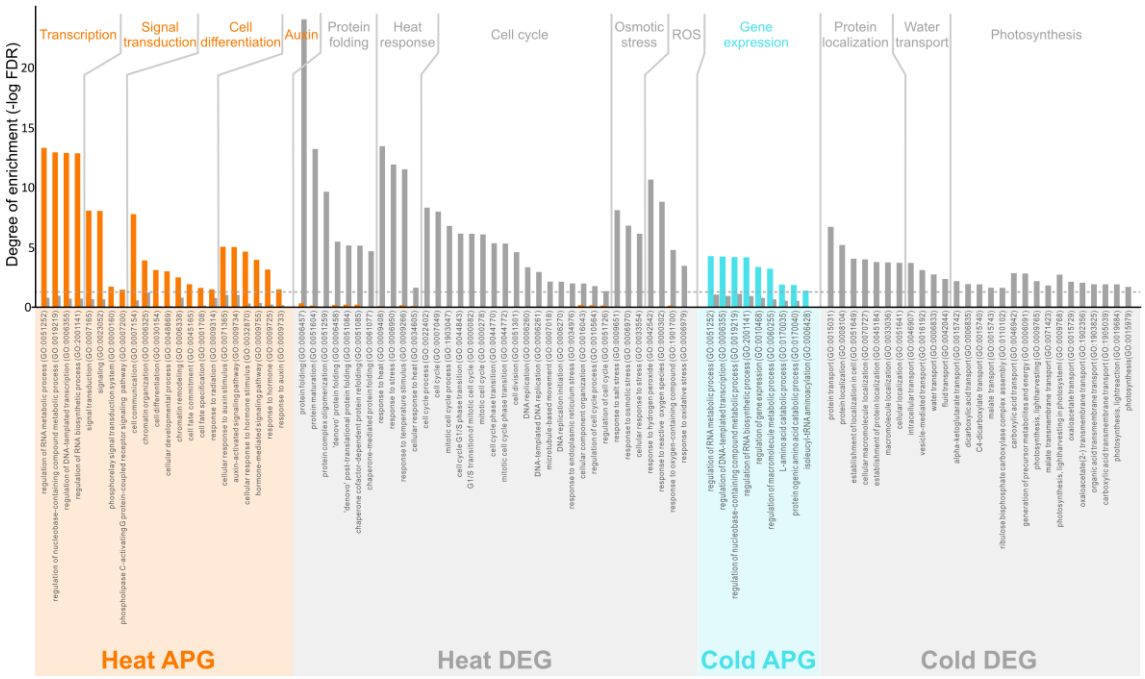

Fig. 4: Enriched functional categories among heat- and cold-responsive APGs and DEGs.

The degree of Gene Ontology (GO) term enrichment is represented by the −log(false discovery rate; FDR) value inferred by a chi-squared test. The horizontal dotted line indicates FDR <0.05. The x-axis presents enriched GO terms, with gray lines separating broad classifications of functional categories. Orange/light blue bars represent heat- and cold-responsive APGs, whereas gray bars represent heat- and cold-responsive DEGs.

2.5 Conservation of heat- and cold-responsive APGs and DEGs between cassava and Arabidopsis

To compare the evolutionary conservation of APGs and DEGs, we obtained TSS data for Arabidopsis under heat and cold conditions for an analysis of the conserved APGs and DEGs between Arabidopsis and cassava. First, we identified 10,236 orthologous gene pairs between Arabidopsis and cassava. Heat- and cold-responsive APGs and DEGs among these orthologous gene pairs are listed in Table 1. Of the heat-responsive genes, 8 APGs and 152 DEGs were conserved between Arabidopsis and cassava; the number of conserved DEGs was significantly higher than expected (Table 1; $P < 0.05$). The expected number (95% confidence interval) was determined by randomly selecting Arabidopsis and cassava genes to control the biased detection of Arabidopsis and cassava heat-responsive genes (Table 1). Among the cold-responsive genes, 6 APGs and 44 DEGs were conserved; the number of conserved cold-responsive DEGs was significantly higher than expected ($P < 0.05$). Accordingly, heat- and cold-responsive DEGs tended to be conserved between Arabidopsis and cassava. However, most heat- and cold-responsive genes emerged in a lineage-specific manner, as reflected by the fact the proportion of conserved orthologous gene pairs did not exceed 29% (152 of 533 Arabidopsis DEGs; Table 1). Therefore, both APGs and DEGs responsive to heat and cold stresses may have largely emerged in a lineage-specific manner or were lost in either Arabidopsis or cassava lineages.

Table 1: Number of genes in orthologous gene pairs between Arabidopsis (At) and cassava (Me)

	Heat responsive				Cold responsive			
	APG		DEG		APG		DEG	
	At	Me	At	Me	At	Me	At	Me
The number of genes	77	752	533	1633	90	209	754	238
Expected number of conserved genes (95% confidence interval)	4-9		74-97		1-6		10-26	
The number of conserved genes	$8^{n.s.}$		152*		$6^{n.s.}$		44*	

2.6 Transcriptional activity of APGs increased after gene duplication events

To assess whether heat- and cold-responsive genes emerged or were lost in a lineage-specific manner, we focused on their lineage-specific paralogs. We subsequently investigated the proportions of lineage-specific paralogs among the heat- and cold-responsive APGs and DEGs. Specifically, we identified Arabidopsis and cassava paralogs that emerged in a lineage-specific manner using orthoFinder (v2.5.4) [32]. The proportions of paralogs among all expressed genes were compared with the proportions of paralogs among heat- or cold-responsive genes. The ratios of genes with and without paralogs were significantly higher for heat- and cold-responsive APGs and DEGs than for all expressed genes [Fig. 5a; (Arabidopsis) heat APG: FDR = 3.84×10^{-7}, heat DEG: 7.55×10^{-10}, cold APG: 8.25×10^{-3}, cold DEG: 1.20×10^{-18}, (Cassava) heat APG: 3.12×10^{-7}, heat DEG: 1.63×10^{-6}, cold APG: 1.79×10^{-4}, cold DEG: 9.36×10^{-2}, chi-squared test]. These results suggest that the number of heat- and cold-responsive genes tended to increase in a lineage-specific manner via gene duplication events (with the exception of cold-responsive DEGs in cassava). This is consistent with the conservation of APGs and DEGs and the findings of previous studies that showed that genes related to abiotic stress responses tend to be duplicated [33,34]. Moreover, it is reasonable that the number of cold-responsive DEGs in cassava did not increase.

We further investigated the types of genes (i.e., heat APGs, cold APGs, heat DEGs, cold DEGs, expressed genes, or non-expressed genes) that tended to acquire alternative promoters through gene duplication events. Specifically, we analyzed the closest paralogs, which can reveal the origins of heat- and cold-responsive APGs and DEGs. The closest paralogs were identified on the basis of the lowest E-value for the similarity between paralog pairs. According to the results, most paralogs were derived from expressed genes. Notably, the proportion of non-expressed paralogs was lower for Arabidopsis heat-responsive APGs and cassava heat- and cold-responsive APGs than for all expressed genes and DEGs [Fig. 5b; (Arabidopsis) heat APG: FDR = 4.71×10^{-6}, heat DEG: 0.83, cold APG:

0.15, cold DEG: 0.30, (Cassava) heat APG: 7.32×10^{-17}, heat DEG: 0.20, cold APG: 2.37×10^{-3}, cold DEG: 0.072, chi-squared test]. These results indicate that APGs were likely not derived from non-expressed genes. Therefore, when non-expressed genes are duplicated, they are less likely to become APGs (compared with DEGs and all genes). This suggests that stress-responsive APGs are often derived from genes that are already responsive to the same conditions (e.g., heat and cold stresses) and acquire additional transcriptional activity from a new TSS after the duplication event.

To examine the possibility that increased transcriptional activity of APGs was acquired after duplication, we compared APG expression levels with those of the closest paralogs. Previous studies did not clarify whether genome-wide APGs form primarily through the dispersion of highly expressed genes or through the simple addition of new TSSs mediating increased transcriptional activity [35]. We hypothesized that if an APG is more highly expressed than its paralog, additional transcriptional activity was acquired after duplication. Our analysis revealed that Arabidopsis heat- and cold-responsive APGs as well as cassava heat-responsive APGs had transcriptional activities that were significantly higher than the transcriptional activities of their paralogs [Fig. 5c; (Arabidopsis) heat APG: FDR = 3.34×10^{-2}, heat DEG: 0.95, cold APG: 3.37×10^{-4}, cold DEG: 0.95, (Cassava) heat APG: 3.33×10^{-2}, heat DEG: 0.95, cold APG: 0.12, cold DEG: 0.83, paired *t*-test]. Thus, the transcriptional activities of these APGs increased (relative to the transcriptional activities of their paralogs) following duplication. This finding supports the model in which transcriptional activity increases for APGs derived from the duplication of expressed genes. Interestingly, this enhanced transcriptional activity was observed consistently across all analyzed Arabidopsis and cassava heat- and cold-responsive APGs. In contrast, among DEGs, it was observed only for cassava cold-responsive DEGs. This difference suggests that during evolution, APGs and DEGs emerged via different mechanisms, with APGs more consistently acquiring new transcriptional regulatory elements after duplication events. The increased expression levels may be attributed to the acquisition of enhancers or promoters; previous research showed that alternative promoters may form from enhancers [35].

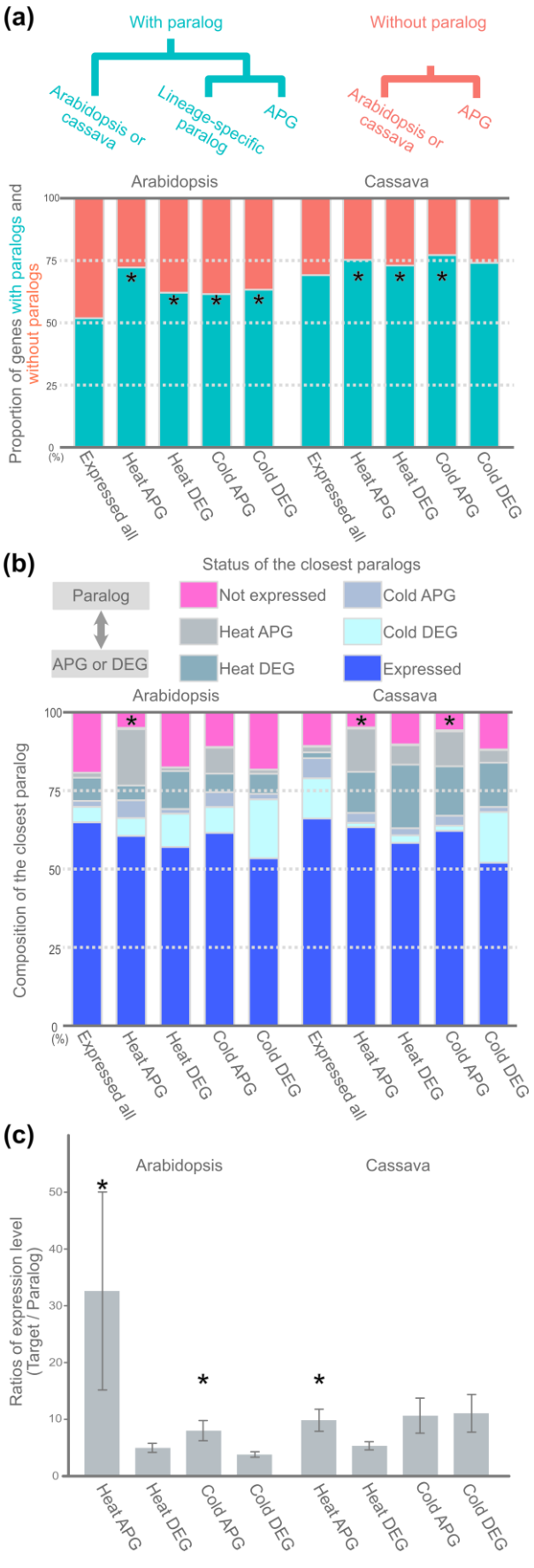
(a)
With paralog
Without paralog
Arabidopsis or cassava
Lineage-specific paralog
APG
Arabidopsis or cassava
APG
Arabidopsis
Cassava
Proportion of genes with paralogs and without paralogs
100
75
50
25
0
(%)
Expressed all
Heat APG
Heat DEG
Cold APG
Cold DEG
(b)
Status of the closest paralogs
Paralog
APG or DEG
Not expressed
Heat APG
Heat DEG
Cold APG
Cold DEG
Expressed
Arabidopsis
Cassava
Composition of the closest paralog
(c)
Arabidopsis
Cassava
Ratios of expression level (Target / Paralog)
50
40
30
20
10
0
Heat APG
Heat DEG
Cold APG
Cold DEG

Fig. 5: Evolutionary relationships of heat- and cold-responsive genes and their paralogs in Arabidopsis and cassava.

(a) Schematic representation of paralog classification. Evolutionary relationships between genes with lineage-specific paralogs (teal branch) and genes without paralogs (pink branch) in Arabidopsis or cassava lineages are presented. Additionally, the proportion of genes with and without lineage-specific paralogs among heat- and cold-responsive APGs, DEGs, and all expressed genes in Arabidopsis and cassava is indicated. Genes with paralogs (light blue) are significantly enriched among temperature-responsive genes (compared with the genomic background), suggesting that gene duplication events played a crucial role in the evolution of stress response mechanisms. Genes without paralogs are shown in pink. Asterisks indicate significant differences (FDR <0.05).

(b) Classification of the closest paralog of each APG on the basis of the expression status (non-expressed, heat/cold-responsive APG, heat/cold-responsive DEG, or expressed) to elucidate the evolutionary origins of alternative promoter usage. The closest paralogs were differentiated in terms of expression (expressed, non-expressed, APG, and DEG) for each gene category. Note the relatively low proportion of non-expressed paralogs among heat-responsive APGs in both species and cold-responsive APGs in cassava, indicating that APGs tended to evolve from expressed genes rather than from silent genomic regions.

(c) Ratios of expression levels between target genes and their closest paralogs among heat-responsive APGs, heat-responsive DEGs, cold-responsive APGs, and cold-responsive DEGs. Considering APGs were expressed at significantly higher levels than their paralogs (particularly Arabidopsis heat/cold-responsive and cassava heat-responsive APGs), these genes likely acquired enhanced transcriptional activity following duplication events, probably through the acquisition of novel enhancer elements rather than promoters alone. This distinguishes the evolutionary trajectory of APGs from that of DEGs, with the former consistently gaining new transcriptional regulatory elements after duplication events.

2.7 Difference in heat or cold response-related motifs between APGs or DEGs and their paralogs

Because APGs were expressed at higher levels than their paralogs, we compared the factors potentially contributing to the differences in transcriptional activity between APGs or DEGs and their paralogs. As a determinant of transcriptional activity, we focused on the motifs in promoter regions. APGs and DEGs were compared with their paralogs in terms of the number of promoter motifs. SEA (v5.5.7) was used to identify heat or cold response-related motifs in 100 bp regions (50 bp upstream and downstream of TSSs) in the most highly expressed genes (top 10%) because these genes are most likely to contain stress-responsive regulatory elements [36,37]. We found 260 and 108 significantly enriched motifs as heat and cold response-related motifs.

The differences in the number of motifs between APGs and their paralogs were compared with the corresponding differences for DEGs and other genes; positive and negative differences indicated that APGs, DEGs, or other genes had more and fewer motifs than their paralogs, respectively. The differences for heat- and cold-responsive APGs were larger than those for DEGs and other genes in the positive range (Fig. 6; heat APG *vs* DEG: FDR = 7.31×10^{-6}, heat APG *vs* other: 1.06×10^{-21}, cold APG *vs* DEG: 4.73×10^{-1}, cold APG *vs* other: 1.94×10^{-2}, Wilcoxon rank sum test). Accordingly, the number of heat or cold response-related motifs tended to increase in APG promoters after duplication events.

Notably, when the analysis was repeated for motifs unrelated to temperature stress responses (gray boxplots in Fig. 6), we found no significant differences among APGs, DEGs, and other genes (heat APG *vs* DEG: FDR = 0.483, heat APG *vs* other: 0.275, cold APG *vs* DEG: 0.473, cold APG *vs* other: 0.0610, Wilcoxon rank sum test). This suggests that the enrichment of stress-responsive motifs in APGs is not a general phenomenon involving all types of regulatory elements. Instead, it represents a selective accumulation of functionally relevant motifs. Collectively, these findings imply that gene duplication events followed by the acquisition of stress-responsive cis-regulatory elements contributed to the emergence of APGs with enhanced expression levels under temperature stress conditions. This represents a molecular mechanism for the functional diversification after the duplication of genes associated with stress response pathways.

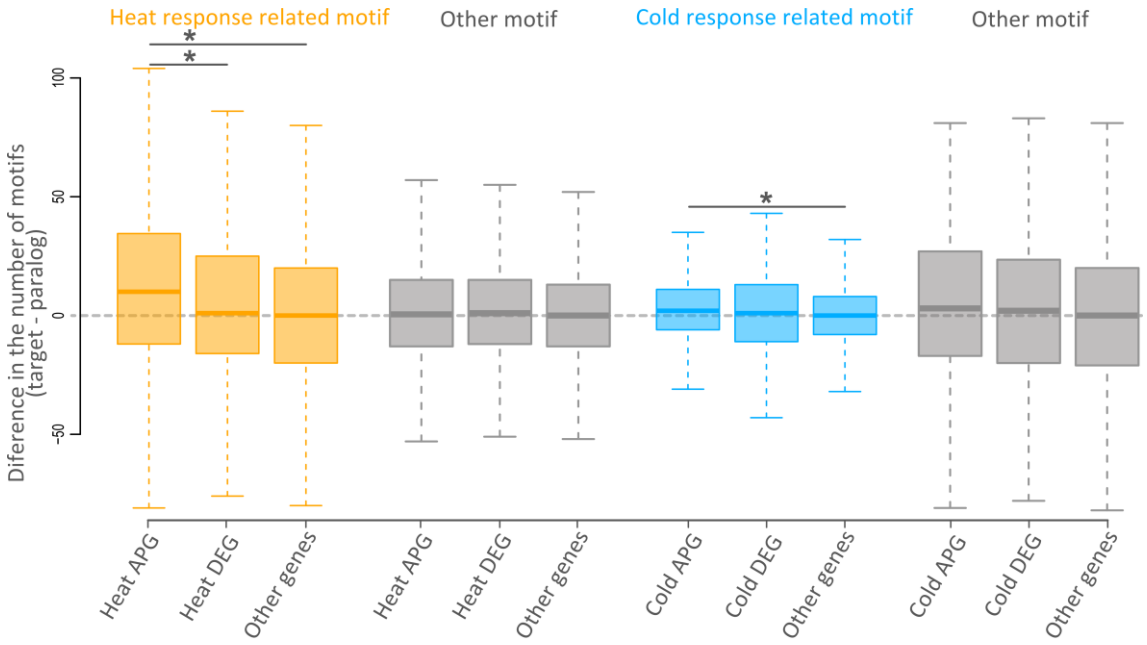

Fig. 6: Comparison of the number of motifs related to heat and cold responses in APGs and DEGs. Boxplots show the difference in the number of motifs (target minus paralog) for heat- and cold-responsive APGs, DEGs, and other genes. The horizontal dotted line indicates no difference in the number of motifs between a target gene and its paralog. Heat response-related motifs (orange) and cold response-related motifs (blue) that differed significantly between APG and DEG categories are indicated by asterisks. Other motifs (gray) did not differ significantly among categories.

3. Discussion

Our comprehensive analysis revealed that APGs contribute to a distinct and essential regulatory mechanism mediating plant responses to temperature stress. This mechanism complements the well-characterized DEG mechanism. We identified many stress-responsive APGs (1,705 and 491 responsive to heat and cold conditions, respectively), demonstrating that alternative promoter usage is a widespread phenomenon comparable in scale to differential expression. On the basis of the study findings, we provide answers to the following questions raised in the Introduction section: (1) What are the quantitative contributions and distinct functional properties of APGs and DEGs in stress responses? Although not as prominent as DEGs, there are many APGs with distinct roles (i.e., regulating gene expression and auxin signaling); (2) How does APG-mediated protein diversity contribute to post-transcriptional regulation, particularly through altered subcellular localization? Truncating transit peptides results in cytoplasmic localization or NLS formation; (3) Are stress-responsive APGs evolutionarily conserved or expanded lineage-specific manner? They were not highly conserved, but temperature stress-responsive APGs emerged following the duplication of genes that were already expressed under the same temperature stress conditions. Building on previous reports describing the involvement of APGs in light stress responses [1,38,39], our results indicate that temperature-responsive APGs are involved in a general stress response mechanism that should be investigated as extensively as DEG-related stress response mechanisms [40,41].

3.1 Number of uORFs in APG transcripts

APG transcripts had fewer uORFs than the transcripts of all genes, expressed genes, and DEGs. This suggests that improved translational efficiency contributes to rapid protein synthesis under stress conditions. By contrast, earlier research showed that the transcripts of blue light-responsive APGs tend to have many uORFs [24]. This discrepancy indicates that plant responses to changes in the light environment and temperature are mediated by different mechanisms. Considering that rapid protein-coding gene expression at high levels is crucial for plant responses to temperature stress, it is reasonable that the expression of heat- and cold-responsive APGs commonly results in transcripts with relatively few uORFs [42].

3.2 Temperature-responsive APGs with inverted isoform expression patterns as breeding targets

APGs represent promising candidates for breeding to improve heat or cold stress tolerance in plants. These genes encode protein isoforms with substantial changes in expression and subcellular localization during responses to heat and cold conditions. Manes.01G075000, Manes.05G077600, Manes.05G197600, Manes.07G024700, Manes.11G055100, Manes.15G142950, and Manes.17G066700 are heat-responsive genes, whereas Manes.05G097100 is a cold-responsive gene. These genes had high read counts (minimum read count of >50 for a specific stress condition), with

inverted protein isoform expression patterns under normal and stress conditions. The length of the protein isoforms was more than 50% of that of the corresponding representative proteins, suggesting the likelihood of their transcript being targeted for nonsense-mediated decay is relatively low [43,44]. Although these genes have not been thoroughly annotated, their substantial response to heat or cold stress indicates they may confer temperature stress tolerance. Thus, they may be relevant to breeding temperature stress-tolerant cassava. This crop species reproduces clonally and exhibits flowering that is not easily induced; however, it is relatively easy to transform [45-48].

3.3 Selection pressure and alternative promoter acquisition

Selection pressure for high expression generally maintains gene functional redundancy [49,50]. Moreover, although TSS changes have been frequently discussed in previous reports, their relevance to gene duplications, abiotic stress responses, or high expression levels has not been demonstrated [51,52]. We determined that APGs tend to have more promoter motifs than their paralogs; these motifs were related to responses to heat and cold conditions. This suggests that APG promoter regions comprise specific motifs, reflecting the effects of selection pressure on expression patterns. In fact, APGs were expressed at higher levels than their paralogs, implying that they may be subject to selection pressure favoring increased expression. Additionally, the relatively high proportion of duplicate genes among APGs might be attributable to the high total expression level due to this selection pressure.

3.4 Functional specialization of APGs revealed by a GO analysis

The enriched GO terms among heat- and cold-responsive APGs were related to transcription and expression. In addition, heat-responsive APGs were associated with signal transduction. This is consistent with the observation that many APGs encode protein isoforms that are localized in the nucleus. Consequently, these genes may be involved in plant responses to heat and cold stresses. If this involvement is experimentally verified, these genes may be useful for elucidating the temperature stress responses of cassava. For example, the heat-responsive APG Manes.08G049500, which encodes a calcium-dependent protein kinase (CDPK), was annotated with the GO term signaling (GO:0023052). Although the representative protein consists of 531 amino acids and is localized in the cell membrane, the production of its protein isoform, which contains 425 amino acids and is localized in the cytoplasm, is up-regulated under heat stress conditions. Because heat stress alters membrane fluidity to trigger temperature stress-related signaling, calcium ion channels respond accordingly, with CDPK playing a crucial role in the transmission of signals to the cytoplasm [53,54]. Thus, an increase in the abundance of the protein isoform encoded by Manes.08G049500 may contribute to a positive feedback loop in response to calcium ion signals. Accordingly, APGs are important for cassava temperature stress responses. Alternative promoters should be thoroughly investigated to comprehensively characterize abiotic stress responses.

Auxin signaling is involved in the induction of tuber formation, a key trait in cassava [55]. In fact, processes related to tuber formation, such as cell differentiation and signal transduction, were associated with APGs, but not DEGs (Fig. 4). Furthermore, a previous study on potato showed that heat stress decreases tuber yield and that heat-responsive APGs may regulate upstream signaling pathways affecting this process [56]. Therefore, it is plausible that key genes involved in regulating auxin signaling and related processes, including signal transduction and cell differentiation, are regulated by alternative promoters [57]. Accordingly, APGs annotated with the GO terms cell differentiation and signal transduction deserve attention as candidate genes involved in cassava tuber formation.

3.5 Studies on alternative promoter usage in tropical plants: methodological and data implications

Research on alternative promoters in tropical plants has been scarce. The nAnT-iCAGE-seq method used in this study, which is characterized by a lack of PCR amplification, restriction enzyme digestion, or hybridization, generates minimally biased expression data and selectively collects capped RNAs, making it specific to TSSs [21]. Therefore, it is ideal for the quantitative detection and identification of TSSs, which are necessary for studying alternative promoters. Although there are methods for determining full-length cDNA sequences using long-read sequencers to detect TSSs, previous studies on cassava that were conducted to elucidate cold response mechanisms on the basis of long-read sequencing had specific limitations (e.g., low quantitativeness). Thus, the methods used in these earlier studies are inappropriate for the quantitative detection of alternative promoters [44,58]. By contrast, the method used in the current study is conducive to the simultaneous detection of DEGs and APGs. Furthermore, it can improve the accuracy of inference when combined with other analyses. Therefore, the data generated in this study may form the basis of future diverse analyses.

4. Materials and Methods

4.1 Plant materials and growth conditions

Cassava cultivars KU50 and ECU72 were used in this study [59,60]. Heat and cold stress tests were performed using both cultivars. The data from both cultivars were analyzed together to increase reliability. Cassava seedling stems were cut into 2.0 cm segments with one bud and placed on solid MS medium (0.1% agar) in plates. All plates were incubated in a greenhouse set at 28 °C with a 12-h photoperiod. One-month-old plants were transferred to a chamber for 5-h heat and cold treatments at 42 °C and 4 °C, respectively. The youngest leaves and shoot apices were then collected and immediately frozen in liquid nitrogen prior to storage at −80 °C until used for RNA extractions.

4.2 nAnT-iCAGE-seq and bioinformatics analysis

A previously described method was used to construct nAnT-iCAGE libraries [21]. The nAnT-iCAGE-

seq analysis of ECU72 and KU50 using Illumina instruments generated 212 and 210 mega-reads, respectively. Adapter and low-quality sequences were eliminated using Trimmomatic (v0.39) (parameters: TRAILING: 34; MINLEN: 20) [61]. According to their low-quality scores, approximately 0.002% of the reads were eliminated. The remaining reads were mapped to the cassava reference genome sequence from Phytozome (v13) (https://phytozome-next.jgi.doe.gov/) using Bowtie2 (v2.4.4) software [62]. A total of 47, 39, and 52 mega-reads for ECU72 and 65, 39, and 37 mega-reads for KU50 were mapped to the cassava reference genome under normal, heat stress, and cold stress conditions, respectively. Mapped sites were clustered for each experimental condition using RECLU (v1.0) [63] to identify TSSs (TPM score >0.1; IDR score >0.1). If TSSs overlapped by ≥1 bp under different conditions, they were re-defined as one TSS peak [1]. TSS peaks overlapping other TSS peaks were merged for all KU50 and ECU72 samples under normal, heat stress, and cold stress conditions, resulting in the identification of 40,520 non-redundant TSS peaks.

4.3 Identification of APGs and DEGs

TSS peaks in the region from 500 bp upstream of the representative TSS to the midpoint between the representative TSS and the transcription termination site were assigned to the corresponding gene [64]. We identified 92 peaks whose closest loci differed from the assigned genes. Thus, 33,408 peaks were assigned to 21,404 annotated cassava genes. These genes were defined as expressed genes. We focused on 6,012 genes with multiple TSS peaks. Specifically, we counted the number of nAnT-iCAGE reads in each TSS peak. In every pair of multiple TSS peaks, the ratios of nAnT-iCAGE reads were compared between the normal and heat/cold stress conditions via a chi-squared test [1]. The FDR value was calculated from P values [65]. The criterion used to determine heat/cold-responsive alternative promoter selection was FDR <0.05 (chi-squared test; normal *vs* heat and normal *vs* cold). Using this approach, we identified 2,933 heat- and 1,835 cold-responsive APGs in ECU72 and 3,062 heat- and 1,538 cold-responsive APGs in KU50. For rigorous analyses, we used 1,705 heat- and 491 cold-responsive APGs whose heat- and cold-responsive peaks and response directions (i.e., up- or down-regulated) were the same for ECU72 and KU50.

To identify DEGs, the expression level of each gene was calculated as the number of nAnT-iCAGE reads that were mapped to the region from 500 bp upstream of the representative TSS to the midpoint between the representative TSS and the transcription termination site. By comparing the expression levels of each gene between the normal and heat/cold stress conditions using DESeq2 (FDR <0.05), we detected 4,172 heat- and 582 cold-responsive DEGs. To examine the accuracy of expression levels estimated on the basis of nAnT-iCAGE-seq data, we compared the total number of reads at each locus with previously reported expression levels under control and cold stress conditions [22]. To exclude the influence of genes unrelated to temperature stress responses or non-expressed genes, the comparison was completed using only 1,521 DEGs identified according to the data generated by

a previous study. The NCBI SRA database lacks transcriptome data for heat stress conditions.

Additionally, Arabidopsis sequencing data capturing TSS information under heat and cold stress conditions were obtained from NCBI SRA and GEO databases (DRX026748 and GSE129523) [6,66]. Applying the same procedure to Arabidopsis data, 15,476 expressed genes, 183 heat- and 194 cold-responsive APGs, and 1,075 heat- and 1,756 cold-responsive DEGs were identified. Biotic stress-responsive DEGs were obtained from a previous study [23], in which DEGs were identified using DESeq2 [67], similar to the current study. We focused on the 3,887 and 4,911 DEGs identified by comparing cassava samples at 2 and 5 days post-infection with their respective controls, with annotations based on cassava v4. The conversion to v8 yielded 3,017 and 3,889 DEGs, respectively.

4.4 Identification of uORFs

Because the average 5′ UTR length in Arabidopsis is 155 nt [68], we focused on uORFs located 150 nt downstream of up-regulated TSS in APGs. The uORFs were identified using uorf4u (v0.9.6) [69]. The number of uORFs was compared for all cassava genes, all expressed genes, and heat- and cold-responsive APGs and DEGs. For all cassava genes and all expressed genes, we identified 150 nt downstream from the representative TSS. For heat- and cold-responsive APGs, we identified 150 nt downstream from non-representative TSSs up-regulated under heat and cold stress conditions, respectively. In terms of DEGs, we focused on the 2,010 and 455 up-regulated DEGs under heat and cold conditions, respectively.

4.5 Identification of N-terminally truncated protein isoforms

We assumed that the coding sequences in TSS isoforms were in the same reading frame as the cassava genes encoding annotated proteins because of nonsense-mediated decay. Accordingly, when the midpoint of a TSS peak coincided with the 5′ upstream region of the representative start codon, the first ATG within that region was designated as the start codon for the corresponding TSS isoform. If TSS peaks were located downstream of the representative start codon, the first ATG following the 3′ TSS peak was designated as the start codon for the TSS isoform. We used diamond (v2.1.6.160) to screen for the same reading frame and candidate genes for protein isoforms [70].

4.6 Analysis of promoter region motifs and identification of heat and cold response-related motifs

Promoter motifs in a 100 bp windows (50 bp upstream and downstream of a TSS) were detected using MEME Suite (v5.5.7) [71]. Heat and cold response-related motifs were identified on the basis of the enriched motifs among highly expressed genes (top 10% expression levels under heat and cold stress conditions). Using SEA (v5.5.7), the proportions of motifs in the highly expressed genes were compared with the proportions of motifs in all cassava genes [36,37]. For an APG, TSS pairs were

classified as up- and down-regulated TSSs according to expression level fold-changes. Thus, we used up-regulated TSSs in APGs and representative TSSs of up-regulated DEGs.

4.7 Calculation of 95% confidence intervals for a conservation analysis

To calculate 95% confidence intervals, a random sampling approach was used to evaluate whether the number of conserved genes between Arabidopsis and cassava was statistically significant. We randomly selected the same number of genes from Arabidopsis and cassava and used them as random orthologous gene pairs. The number of conserved APGs or DEGs in these random orthologous gene pairs was determined. This process was repeated 10,000 times. To determine 95% confidence intervals, the 2.5th and 97.5th percentiles were extracted from the distribution of random sampling results.

5 Acknowledgments

We thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

6 References

1 Ushijima, T. *et al.* in *Cell* Vol. 171 1316-1325.e1312 (Elsevier, 2017).

2 Tokizawa, M. *et al.* Identification of Arabidopsis genic and non-genic promoters by paired-end sequencing of TSS tags. *Plant J* **90**, 587-605 (2017). https://doi.org/10.1111/tpj.13511

3 Hetzel, J., Duttke, S. H., Benner, C. & Chory, J. in *Proceedings of the National Academy of Sciences* Vol. 113 12316-12321 (National Academy of Sciences, 2016).

4 Crepieux, P. *et al.* The two functionally distinct amino termini of chicken c-ets-1 products arise from alternative promoter usage. *Gene Expr* **3**, 215-225 (1993).

5 Bourdon, J. C. p53 and its isoforms in cancer. *Br J Cancer* **97**, 277-282 (2007). https://doi.org/10.1038/sj.bjc.6603886

6 Thomas, Q. A. *et al.* in *Nature Communications* Vol. 11 2589 (2020).

7 Dutertre, M. *et al.* Estrogen regulation and physiopathologic significance of alternative promoters in breast cancer. *Cancer Res* **70**, 3760-3770 (2010). https://doi.org/10.1158/0008-5472.CAN-09-3988

8 Thatcher, L. F. *et al.* Differential gene expression and subcellular targeting of Arabidopsis glutathione S-transferase F8 is achieved through alternative transcription start sites. *J Biol Chem* **282**, 28915-28928 (2007). https://doi.org/10.1074/jbc.M702207200

9 Chen, Y. *et al.* Alternative localization of HEME OXYGENASE 1 in plant cells regulates cytosolic heme catabolism. *Plant Physiol* **195**, 2937-2951 (2024). https://doi.org/10.1093/plphys/kiae288

10 Do, T. N. Q. *et al.* 1-Butanol treatment enhances drought stress tolerance in Arabidopsis

thaliana. *Plant Mol Biol* **114**, 86 (2024). https://doi.org/10.1007/s11103-024-01479-0

11 Chang, P., Hsieh, H. Y. & Tu, S. L. The U1 snRNP component RBP45d regulates temperature-responsive flowering in Arabidopsis. *Plant Cell* **34**, 834-851 (2022). https://doi.org/10.1093/plcell/koab273

12 Todaka, D. *et al.* Application of ethanol alleviates heat damage to leaf growth and yield in tomato. *Front Plant Sci* **15**, 1325365 (2024). https://doi.org/10.3389/fpls.2024.1325365

13 Kim, J. S. *et al.* Arabidopsis TBP-ASSOCIATED FACTOR 12 ortholog NOBIRO6 controls root elongation with unfolded protein response cofactor activity. *Proc Natl Acad Sci U S A* **119** (2022). https://doi.org/10.1073/pnas.2120219119

14 Pumplin, N. *et al.* in *The Plant Cell* Vol. 28 2786-2804 (2016).

15 Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C. & Huang, T. H.-M. in *Trends in Genetics* Vol. 24 167-177 (2008).

16 Wang, X. *et al.* Deep learning approaches for non-coding genetic variant effect prediction: current progress and future prospects. *Brief Bioinform* **25** (2024). https://doi.org/10.1093/bib/bbae446

17 Zhu, S. *et al.* RNAirport: a deep neural network-based database characterizing representative gene models in plants. *J Genet Genomics* **51**, 652-664 (2024). https://doi.org/10.1016/j.jgg.2024.03.004

18 Muiruri, S. K., Ntui, V. O., Tripathi, L. & Tripathi, J. N. Mechanisms and approaches towards enhanced drought tolerance in cassava (Manihot esculenta). *Current Plant Biology* **28** (2021). https://doi.org/10.1016/j.cpb.2021.100227

19 Malik, A. I. *et al.* in *Breeding Science* Vol. 70 145-166 (2020).

20 Sukko, N., Kalapanulak, S. & Saithong, T. Trehalose metabolism coordinates transcriptional regulatory control and metabolic requirements to trigger the onset of cassava storage root initiation. *Sci Rep* **13**, 19973 (2023). https://doi.org/10.1038/s41598-023-47095-8

21 Murata, M. *et al.* Detecting expressed genes using CAGE. *Methods Mol Biol* **1164**, 67-85 (2014). https://doi.org/10.1007/978-1-4939-0805-9_7

22 Li, S. *et al.* in *Journal of Experimental Botany* (ed Ramanjulu Sunkar) (2019).

23 Amuge, T. *et al.* A time series transcriptome analysis of cassava (Manihot esculenta Crantz) varieties challenged with Ugandan cassava brown streak virus. *Sci Rep* **7**, 9747 (2017). https://doi.org/10.1038/s41598-017-09617-z

24 Kurihara, Y. *et al.* in *Proceedings of the National Academy of Sciences of the United States of America* Vol. 115 7831-7836 (National Academy of Sciences, 2018).

25 Chang, Z., Zhao, M., Qin, B. & Hong, L. Polar-localized OsLTPG22 regulates rice leaf cuticle deposition and drought response. *Plant Stress* **14** (2024).

https://doi.org/10.1016/j.stress.2024.100586

26 Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H. & Winther, O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res* **50**, W228-W234 (2022). https://doi.org/10.1093/nar/gkac278

27 Lakehal, A. *et al.* A Molecular Framework for the Control of Adventitious Rooting by TIR1/AFB2-Aux/IAA-Dependent Auxin Signaling in Arabidopsis. *Mol Plant* **12**, 1499-1514 (2019). https://doi.org/10.1016/j.molp.2019.09.001

28 Zhao, Y. *et al.* Physiological and transcriptomic analysis revealed the involvement of crucial factors in heat stress response of Rhododendron hainanense. *Gene* **660**, 109-119 (2018). https://doi.org/10.1016/j.gene.2018.03.082

29 Padmanabhan, M. S., Shiferaw, H. & Culver, J. N. The Tobacco mosaic virus replicase protein disrupts the localization and function of interacting Aux/IAA proteins. *Mol Plant Microbe Interact* **19**, 864-873 (2006). https://doi.org/10.1094/MPMI-19-0864

30 Mullender, M., Varrelmann, M., Savenkov, E. I. & Liebe, S. Manipulation of auxin signalling by plant viruses. *Mol Plant Pathol* **22**, 1449-1458 (2021). https://doi.org/10.1111/mpp.13122

31 Abeysingha, D. N. *et al.* The effect of auxins on amelioration of heat stress‐induced wheat (Triticum aestivum L.) grain loss. *Journal of Agronomy and Crop Science* **207**, 970-983 (2021). https://doi.org/10.1111/jac.12555

32 Emms, D. M. & Kelly, S. in *Genome Biology* Vol. 16 157 (2015).

33 Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K. & Shiu, S.-H. in *PLANT PHYSIOLOGY* Vol. 148 993-1003 (2008).

34 Khan, W. U., Khan, L. U., Khan, N. M., Wenquan, W. & Chen, F. Evolutionary insights and expression dynamics of the glutathione peroxidase (GPX) gene family in water lily (Nymphaea colorata) in response to multiple abiotic stresses. *Plant Stress* **14** (2024). https://doi.org/10.1016/j.stress.2024.100699

35 Kim, J., Perera, B. P. U. & Ghimire, S. Enhancer-driven alternative promoters of imprinted genes. *PLoS One* **13**, e0208421 (2018). https://doi.org/10.1371/journal.pone.0208421

36 Bailey, T. L. & Grant, C. E. SEA: Simple Enrichment Analysis of motifs. *bioRxiv*, 2021.2008. 2023.457422 (2021). https://doi.org/10.1101/2021.08.23.457422

37 Yasmeen, E., Wang, J., Riaz, M., Zhang, L. & Zuo, K. Designing artificial synthetic promoters for accurate, smart, and versatile gene expression in plants. *Plant Commun* **4**, 100558 (2023). https://doi.org/10.1016/j.xplc.2023.100558

38 Li, C., Lenhard, B. & Luscombe, N. M. Integrated analysis sheds light on evolutionary

trajectories of young transcription start sites in the human genome. *Genome Res* **28**, 676-688 (2018). https://doi.org/10.1101/gr.231449.117

39 Thodberg, M. *et al.* Comprehensive profiling of the fission yeast transcription start site activity during stress and media response. *Nucleic Acids Res* **47**, 1671-1691 (2019). https://doi.org/10.1093/nar/gky1227

40 Biniaz, Y., Tahmasebi, A., Afsharifar, A., Tahmasebi, A. & Poczai, P. Meta-Analysis of Common and Differential Transcriptomic Responses to Biotic and Abiotic Stresses in Arabidopsis thaliana. *Plants (Basel)* **11** (2022). https://doi.org/10.3390/plants11040502

41 Kumar, M. *et al.* In silico analysis and expression profiling reveal the presence of abiotic stress and developmental stage specific Aconitase genes in rice (Oryza sativa L.). *Plant Stress* **11** (2024). https://doi.org/10.1016/j.stress.2024.100416

42 Chen, H., Yu, J. & Li, W.-X. Non-conventional peptides in plants: From gene regulation to crop improvement. *The Crop Journal* **11**, 323-331 (2023). https://doi.org/10.1016/j.cj.2022.10.006

43 Alyahya, N. & Taybi, T. Transcriptome-wide characterization of alternative splicing regulation in Najran wheat (Triticum aestivum) under salt stress. *Current Plant Biology* **38** (2024). https://doi.org/10.1016/j.cpb.2024.100334

44 Ezoe, A. *et al.* Fully sequencing the cassava full-length cDNA library reveals unannotated transcript structures and alternative splicing events in regions with a high density of single nucleotide variations, insertions-deletions, and heterozygous sequences. *Plant Mol Biol* **112**, 33-45 (2023). https://doi.org/https://doi.org/10.1007/s11103-023-01346-4

45 Anjanappa, R. B. & Gruissem, W. Current progress and challenges in crop genetic transformation. *J Plant Physiol* **261**, 153411 (2021). https://doi.org/10.1016/j.jplph.2021.153411

46 Gaut, B. S., Seymour, D. K., Liu, Q. & Zhou, Y. Demography and its effects on genomic variation in crop domestication. *Nat Plants* **4**, 512-520 (2018). https://doi.org/10.1038/s41477-018-0210-1

47 Taylor, N. *et al.* A High-throughput Platform for the Production and Analysis of Transgenic Cassava (Manihot esculenta) Plants. *Tropical Plant Biology* **5**, 127-139 (2012). https://doi.org/10.1007/s12042-012-9099-4

48 Liu, L., Zhan, J. & Yan, J. Engineering the future cereal crops with big biological data: toward intelligence-driven breeding by design. *Journal of Genetics and Genomics* **51**, 781-789 (2024). https://doi.org/10.1016/j.jgg.2024.03.005

49 Ezoe, A. *et al.* Decrease in purifying selection pressures on wheat homoeologous genes: tetraploidization versus hexaploidization. *Plant J* **120**, 1190-1205 (2024). https://doi.org/10.1111/tpj.17047

50 Ezoe, A., Shirai, K. & Hanada, K. in *Molecular Biology and Evolution* Vol. 38 (ed Michael Purugganan) 1447-1459 (2021).

51 Summat, P. *et al.* Investigating cold tolerance mechanisms in rice seedlings: Alternative splicing, promoter analysis, and their applications for marker development. *Plant Stress* **13** (2024). https://doi.org/10.1016/j.stress.2024.100530

52 Ezoe, A. & Seki, M. Exploring the complexity of genome size reduction in angiosperms. *Plant Mol Biol* **114**, 121 (2024). https://doi.org/10.1007/s11103-024-01518-w

53 Raza, A. *et al.* Trehalose: A sugar molecule involved in temperature stress management in plants. *The Crop Journal* **12**, 1-16 (2024). https://doi.org/10.1016/j.cj.2023.09.010

54 Nievola, C. C., Carvalho, C. P., Carvalho, V. & Rodrigues, E. Rapid responses of plants to temperature changes. *Temperature (Austin)* **4**, 371-405 (2017). https://doi.org/10.1080/23328940.2017.1377812

55 Utsumi, Y. *et al.* Integrative omics approaches revealed a crosstalk among phytohormones during tuberous root development in cassava. *Plant Mol Biol* **109**, 249-269 (2022). https://doi.org/10.1007/s11103-020-01033-8

56 Harvey, B. M. R., Crothers, S. H., Watson, S. & Lee, H. C. Heat inhibition of tuber development in potato (Solanum tuberosum L.): effects on microtuber formation in vitro. *Potato Research* **35**, 183-190 (1992). https://doi.org/10.1007/bf02357612

57 Khoso, M. A. *et al.* Impact of plant growth-promoting rhizobacteria (PGPR) on plant nutrition and root characteristics: Current perspective. *Plant Stress* **11** (2024). https://doi.org/10.1016/j.stress.2023.100341

58 Zhao, L. *et al.* Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing. *Front Genet* **10**, 253 (2019). https://doi.org/10.3389/fgene.2019.00253

59 Perez-Fons, L., Bohorquez-Chaux, A., Gomez-Jimenez, M. I., Becerra Lopez-Lavalle, L. A. & Fraser, P. D. Untargeted and targeted metabolomics approaches to characterise, select and advance cassava pre-breeding populations with enhanced whitefly tolerance. *Plant J* **122**, e70233 (2025). https://doi.org/10.1111/tpj.70233

60 Gu, J. *et al.* RNA splicing modulates the postharvest physiological deterioration of cassava storage root. *Plant Physiol* **196**, 461-478 (2024). https://doi.org/10.1093/plphys/kiae206

61 Bolger, A. M., Lohse, M. & Usadel, B. in *Bioinformatics* Vol. 30 2114-2120 (Oxford University Press, 2014).

62 Langmead, B. & Salzberg, S. L. in *Nature Methods* Vol. 9 357-359 (2012).

63 Ohmiya, H. *et al.* in *BMC Genomics* Vol. 15 1-15 (2014).

64 Morton, T. *et al.* in *Plant Cell* Vol. 26 2746-2760 (American Society of Plant Biologists, 2014).

65 Benjamini, Y. & Hochberg, Y. in *Journal of the Royal Statistical Society: Series B (Methodological)* Vol. 57 289-300 (1995).

66 Ivanov, M., Sandelin, A. & Marquardt, S. TrancriptomeReconstructoR: data-driven annotation of complex transcriptomes. *BMC Bioinformatics* **22**, 290 (2021). https://doi.org/10.1186/s12859-021-04208-2

67 Love, M. I., Huber, W. & Anders, S. in *Genome Biology* Vol. 15 550 (2014).

68 Liu, B. *et al.* Analysis of the 5' Untranslated Region Length-Dependent Control of Gene Expression in Maize: A Case Study with the ZmLAZ1 Gene Family. *Genes (Basel)* **15** (2024). https://doi.org/10.3390/genes15080994

69 Egorov, A. A. & Atkinson, G. C. uORF4u: a tool for annotation of conserved upstream open reading frames. *Bioinformatics* **39** (2023). https://doi.org/10.1093/bioinformatics/btad323

70 Buchfink, B., Xie, C. & Huson, D. H. in *Nature Methods* Vol. 12 59-60 (Life Science Alliance, LLC, 2015).

71 Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res* **43**, W39-49 (2015). https://doi.org/10.1093/nar/gkv416