An Integrative Bioinformatics Framework for Transcriptomic Reanalysis Reveals Regulatory Architectures in the Lychee Response to *Peronophythora*

Eric Zhuang

SUMMARY

Lychee (*Litchi chinensis* Sonn.) is an economically important tropical fruit threatened by litchi downy blight, caused by the oomycete *Peronophythora litchii*. To prioritize genes and regulatory features associated with early host responses, we reanalyzed public leaf RNA-seq data from Guiwei and Yurong cultivars (GSE201243; mock and inoculated samples, three biological replicates per group). Differential-expression analysis identified 17 significant DEGs in Guiwei and 117 in Yurong at adjusted $P < 0.05$, suggesting a broader transcriptional response in Yurong. Eighteen highly responsive genes were selected for structural, motif, promoter, and candidate interaction analyses. These candidates included annotations associated with defense or stress responses, including jacalin-like lectin, pathogenesis-related protein, heat-shock protein, thaumatin/osmotin-like proteins, cytochrome P450, lipid-transfer proteins, and calcium-binding proteins. A selected-gene interaction contrast prioritized LITCHI019519, LITCHI001510, LITCHI028401, LITCHI017676, LITCHI028104, and LITCHI019183 as potential cultivar-dependent response genes. Promoter analysis of 2-kb upstream regions detected stress- and hormone-associated cis-elements, and a random-promoter background test supported enrichment of ARE, MeJA-responsive motifs, TCA-elements, and TC-rich repeats. These results provide testable candidates for future functional validation of lychee resistance to *P. litchii*.

Keywords: lychee; Litchi chinensis; litchi downy blight; Peronophythora litchii; RNA-seq; differential expression; promoter analysis; cis-regulatory elements; plant defense

INTRODUCTION

Lychee (*Litchi chinensis* Sonn.) is a tropical fruit crop native to southern China and Southeast Asia and cultivated in subtropical and tropical production regions worldwide [1]. Litchi downy blight, caused by the oomycete *Peronophythora litchii* (*P. litchii*), is among the most damaging diseases of this crop and can affect leaves, inflorescences, and fruit, producing water-soaked brown lesions and postharvest decay [2,3]. Understanding the host transcriptional

response to *P. litchii* is therefore important for identifying molecular markers and candidate genes that may support disease-resistance breeding.

Recent genome-sequencing studies have improved genomic resources for lychee and enabled transcriptome-based analyses of disease response [4,5]. However, many lychee genes involved in the response to *P. litchii* remain poorly characterized. Liu et al. identified LcCDPK family members with altered expression after *P. litchii* inoculation [6], and Li et al. showed that the *P. litchii* RXLR effector PlAvh202 promotes virulence by destabilizing a host ethylene-biosynthesis enzyme [7].

Comparative studies of susceptible and resistant lychee cultivars have also emphasized the importance of rapid pathogen recognition and early defense responses [8]. Public RNA-seq dataset GSE201243 provides a balanced cultivar-by-treatment design that can be reused to identify infection-responsive candidate genes and to explore whether transcriptional responses differ between cultivars [9].

Here, we reanalyzed GSE201243 to identify differentially expressed genes associated with *P. litchii* inoculation and to characterize the potential functions and promoter features of selected candidates. Specifically, we performed cultivar-specific differential-expression analyses, calculated selected-gene cultivar-by-treatment contrasts, screened full DEG sets with an orthology-guided Plant Reactome workflow, and examined chromosomal positions, gene structures, conserved protein motifs, and 2-kb upstream cis-regulatory elements. Because this study uses public transcriptomic data and computational annotation, the results are presented as hypothesis-generating candidates for experimental validation.

RESULTS

To evaluate transcriptional responses of lychee leaves to *P. litchii*, we analyzed public RNA-seq data from GEO accession GSE201243. The dataset includes two cultivars, Guiwei and Yurong1 (hereafter Yurong), sterile-water mock and *P. litchii*-inoculated samples, and three biological replicates per cultivar-treatment combination. We first summarized run-depth metadata to confirm sample balance and to screen for gross sequencing-depth outliers.

Dataset overview and sequencing-depth quality control

GSE201243 [9] contains 12 leaf RNA-seq libraries representing a balanced 2 x 2 design: two cultivars (Guiwei and Yurong1, hereafter Yurong), two treatments (sterile-water mock control and P. litchii inoculation), and three biological replicates per cultivar-treatment combination. According to the linked SRA records, all libraries were paired-end Illumina NovaSeq 6000 RNA-seq runs from leaves collected 24 h after treatment.

Sample	Group	GEO accession	SRA run	Spots (M)	Bases (Gb)
GW_M1	Guiwei mock	GSM6053458	SRR18856600	23.71	7.1
GW_M2	Guiwei mock	GSM6053459	SRR18856599	27.08	8.1
GW_M3	Guiwei mock	GSM6053460	SRR18856598	27.02	8.1
GW_P1	Guiwei P. litchii	GSM6053461	SRR18856609	27.59	8.3
GW_P2	Guiwei P. litchii	GSM6053462	SRR18856608	24.71	7.4
GW_P3	Guiwei P. litchii	GSM6053463	SRR18856607	23.46	7.0
YR_M1	Yurong mock	GSM6053464	SRR18856606	22.91	6.9
YR_M2	Yurong mock	GSM6053465	SRR18856605	23.50	7.1
YR_M3	Yurong mock	GSM6053466	SRR18856604	23.55	7.1
YR_P1	Yurong P. litchii	GSM6053467	SRR18856603	22.56	6.8
YR_P2	Yurong P. litchii	GSM6053468	SRR18856602	23.19	7.0
YR_P3	Yurong P. litchii	GSM6053469	SRR18856601	21.33	6.4

Table 1. Sample groups and SRA run-depth summary for GSE201243.
Note: M = million; Gb = gigabases. SRA spots are the run-level sequencing spots reported by NCBI SRA for the paired-end RNA-seq libraries.

Run-depth across the 12 libraries ranged from 21.33 million spots (YR_P3) to 27.59 million spots (GW_P1), with a median of 23.53 million spots and an interquartile range of 23.12-25.29 million spots. The maximum/minimum ratio was 1.29, indicating moderate but not extreme sequencing-depth variation. Therefore, no sample was excluded based on SRA run-depth metadata.

PCA and hierarchical clustering were then performed as a metadata-level QC screen using standardized SRA run metrics (spot count, total bases, and downloaded file size). The resulting PCA and dendrogram separated libraries mainly by sequencing depth rather than by cultivar or treatment (Figure 1). No single run was isolated from all others in this metric space. This QC screen cannot substitute for transcriptome-level PCA based on variance-stabilized raw counts, but it documents the sample groups and did not reveal an obvious run-depth outlier.

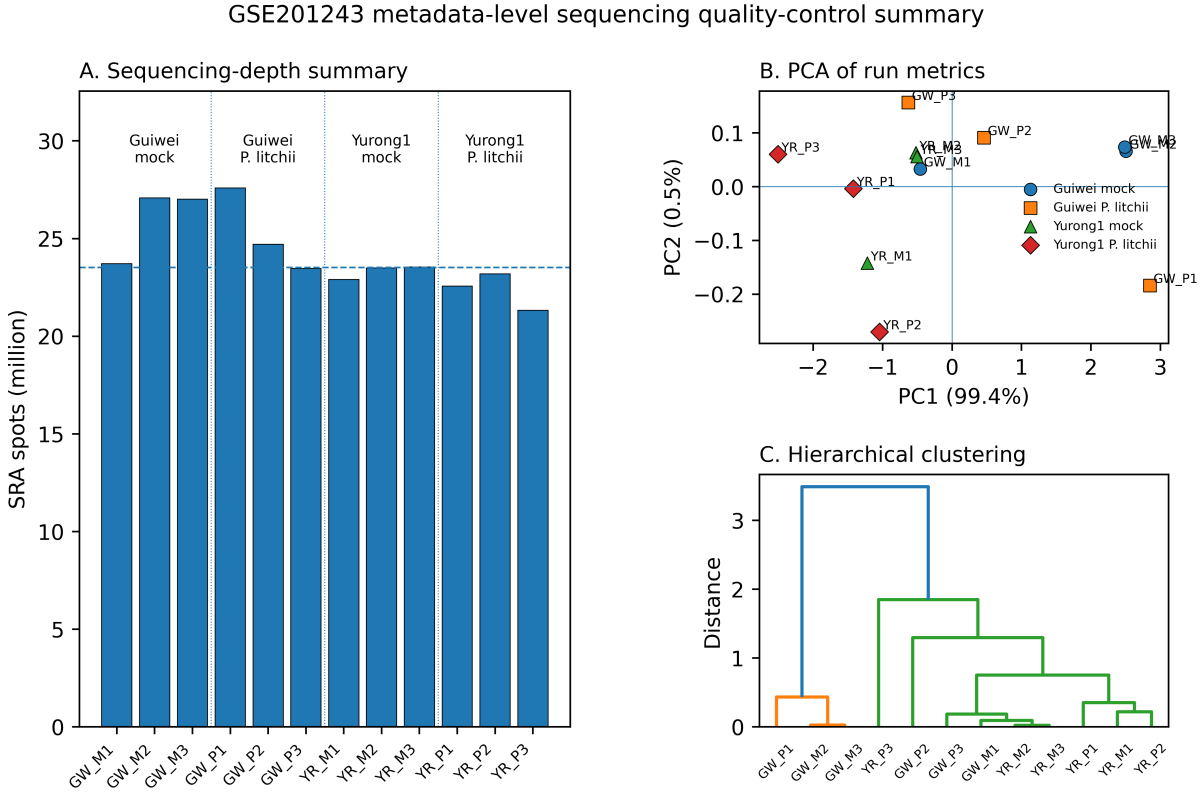

Figure 1. Metadata-level sequencing quality-control summary for GSE201243. (A) SRA run spots for each library; the dashed line indicates the median. (B) PCA based on standardized run-level metrics (spot count, total bases, and downloaded file size). (C) Hierarchical clustering based on the same standardized run-level metrics. The analysis did not identify an obvious sequencing-depth outlier.

Pairwise differential-expression overview

Volcano plots were generated using log2 fold change and -log10(P value) (Figure 2). Genes with adjusted P values below 0.05 were considered significantly differentially expressed. Under this threshold, Guiwei had 15 upregulated and 2 downregulated genes, whereas Yurong had 39 upregulated and 78 downregulated genes. The larger number of differentially expressed genes in Yurong suggests a stronger or broader transcriptional response to *P. litchii* inoculation.

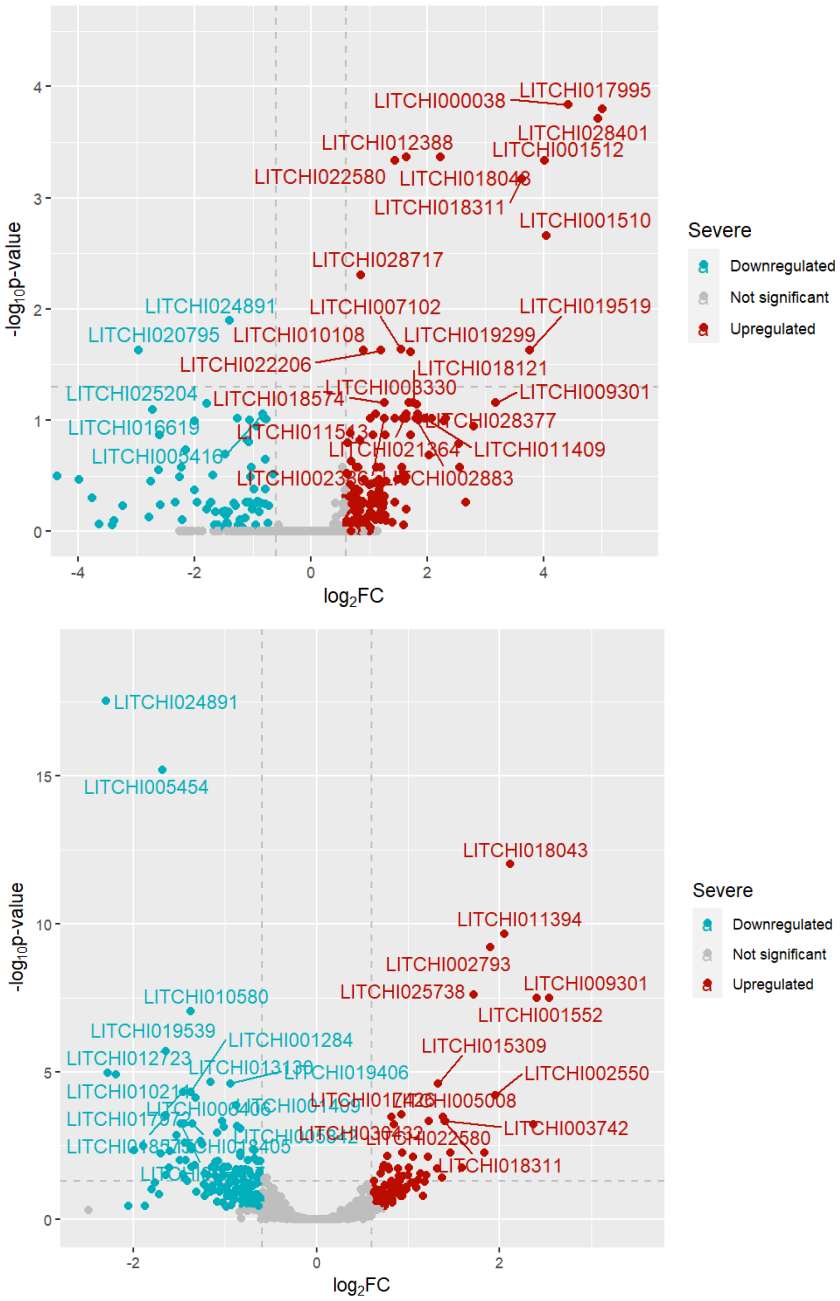

Figure 2. Volcano plots of differentially expressed genes in GSE201243. Top: Guiwei; bottom: Yurong. Red indicates upregulated genes and blue indicates downregulated genes.

The complete significant DEG sets (17 genes in Guiwei and 117 genes in Yurong) were retained for pathway-level screening. From these significant pairwise DEG lists, 18 genes were prioritized for downstream structural, motif, promoter, and candidate-gene interaction analyses by combining statistical and biological criteria: significance in at least one cultivar-specific contrast (adjusted P < 0.05), large infection-associated log2 fold-change values in Guiwei and/or Yurong, and predicted annotations or domains plausibly related to defense, stress

signaling, calcium binding, lipid or membrane biology, protein remodeling, or defense-associated metabolism. This selection was therefore not based only on heatmap appearance. Supplementary Table S1 summarizes the rounded log2FC evidence, adjusted-P threshold status, predicted annotation, representative homolog/domain, known defense relevance, and reason for retaining each selected gene.

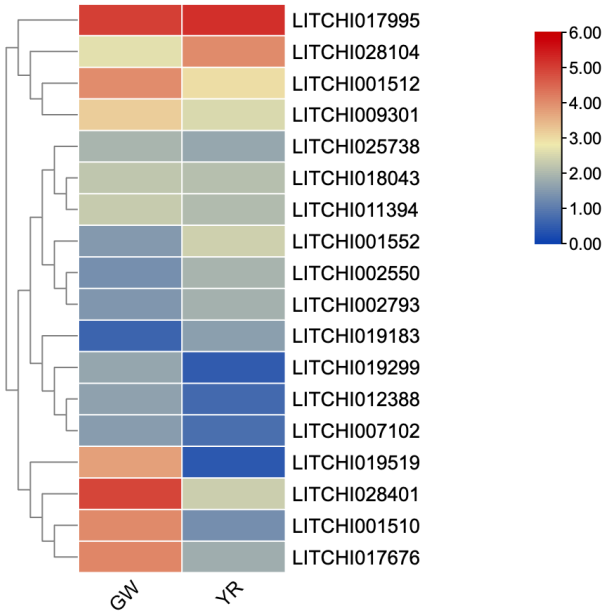

Figure 3. Heatmap comparing selected differentially expressed genes in Guiwei and Yurong. Color indicates log2 fold change.

Cultivar-by-treatment interaction analysis

To evaluate whether the infection response differed between Guiwei and Yurong, we next represented the experiment as a 2 x 2 factorial design with cultivar, treatment, and cultivar-by-treatment terms. In this model, the cultivar-by-treatment coefficient is equivalent to the difference between the infection-induced log2 fold change in Yurong and the infection-induced log2 fold change in Guiwei: interaction contrast = log2FC(Yurong, P. litchii versus mock) - log2FC(Guiwei, P. litchii versus mock). Positive values indicate a stronger response in Yurong, whereas negative values indicate a stronger response in Guiwei.

Applying this interaction contrast to the selected candidate genes refined the interpretation of the separate differential-expression comparisons. The strongest Guiwei-skewed contrasts were observed for LITCHI019519, LITCHI001510, LITCHI028401, and LITCHI017676, whereas LITCHI028104 and LITCHI019183 showed Yurong-skewed responses (Figure 4; Table 2). These genes therefore represent candidate cultivar-dependent response genes for future expression validation and functional testing. Because the current contrast was calculated from

selected-gene log2 fold-change summaries, it should be interpreted as a candidate-prioritization effect size rather than a genome-wide list of statistically significant interaction DEGs.

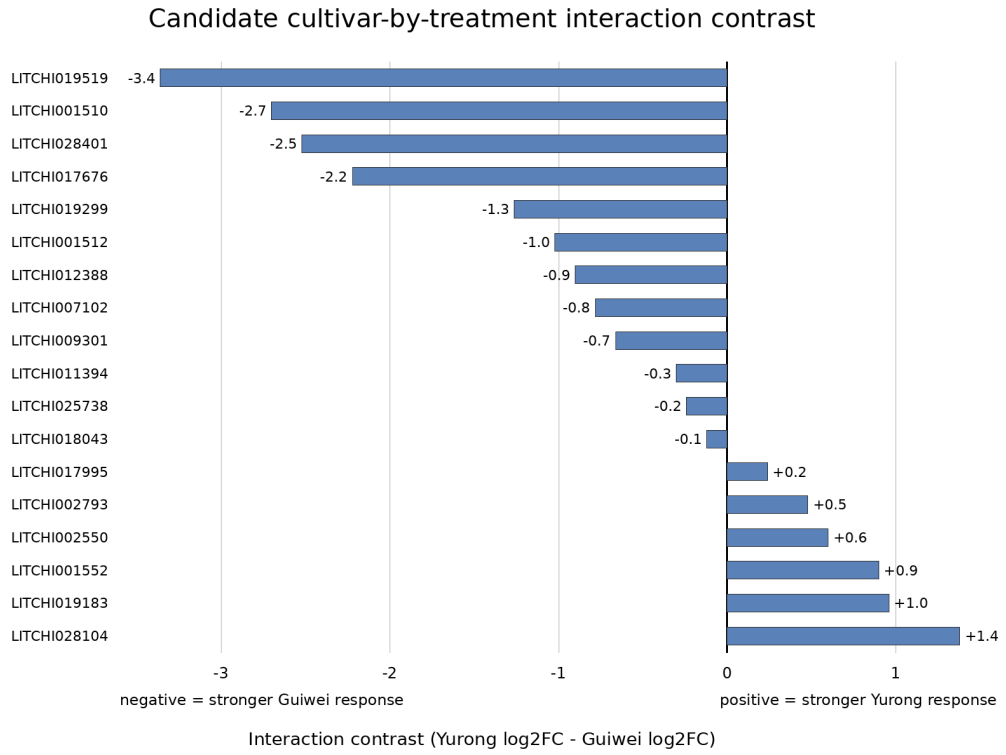

Figure 4. Candidate cultivar-by-treatment interaction contrasts among selected differentially expressed genes. The interaction contrast was calculated as log2FC(Yurong) - log2FC(Guiwei). Negative values indicate a stronger infection-associated response in Guiwei, and positive values indicate a stronger response in Yurong.

Gene	Guiwei log2FC	Yurong log2FC	Interaction contrast	Direction
LITCHI019519	3.8	0.4	-3.4	Stronger Guiwei response
LITCHI001510	4.0	1.3	-2.7	Stronger Guiwei response
LITCHI028401	4.9	2.4	-2.5	Stronger Guiwei response
LITCHI017676	4.0	1.8	-2.2	Stronger Guiwei response
LITCHI028104	2.6	4.0	+1.4	Stronger Yurong response
LITCHI019299	1.7	0.5	-1.3	Stronger Guiwei response
LITCHI001512	4.0	3.0	-1.0	Stronger Guiwei response
LITCHI019183	0.6	1.5	+1.0	Stronger Yurong response

Table 2. Top candidate cultivar-by-treatment interaction genes among the selected differentially expressed genes.

Note: The interaction contrast is log2FC(Yurong) - log2FC(Guiwei). Values are rounded to one decimal place and are used for candidate prioritization. Because the contrast shown here was calculated for selected genes, it should not be interpreted as a genome-wide interaction-DEG test.

Orthology-guided Plant Reactome pathway analysis of all DEGs

To broaden pathway-level interpretation beyond the 18 selected candidate genes, we evaluated the full pairwise DEG lists from Section 2.2 using a Plant Reactome-oriented workflow. The inputs were the complete significant DEG sets identified in the pairwise comparisons: 17 significant DEGs in Guiwei and 117 significant DEGs in Yurong. Plant Reactome provides curated rice reference pathways and orthology-based pathway projections for other plant species, and supports omics-data analysis in the context of plant pathway networks [10]. However, the LITCHI gene identifiers used in the lychee genome annotation were not directly represented in the Plant Reactome gene-products export examined for this revision. Therefore, the analysis was treated as an orthology-guided pathway-annotation screen rather than a formal over-representation test with FDR-corrected Plant Reactome P values.

The all-DEG screen showed that the larger Yurong DEG set provided broader pathway coverage than the smaller Guiwei DEG set (Table 3). Across the full DEG lists, the most interpretable pathway themes were stress and hormone signaling, plant defense and transcriptional regulation, protein folding and turnover, carbohydrate and cell-wall-related metabolism, and membrane or receptor-associated transport. These pathway themes are consistent with the pairwise DEG results and support the view that early infection responses involve both defense-associated signaling and broader cellular reprogramming. Because lychee is not directly represented in the mapping file used here, Table 3 should be interpreted as pathway-level prioritization of all DEGs, not as definitive statistical enrichment evidence.

Plant Reactome pathway theme	**Guiwei all-DEG support**	**Yurong all-DEG support**	**Interpretation**
Stress, immune, and hormone signaling	Detected among the 17 significant DEGs; narrower support because of the smaller DEG set	Broadly represented among the 117 significant DEGs	Suggests infection-induced signaling and hormone-mediated defense regulation.
Transcriptional and receptor-mediated regulation	Detected	Detected	Consistent with cultivar-dependent regulation of pathogen-response genes.
Protein folding, processing, and turnover	Limited support	Detected	Indicates stress-associated proteostasis and possible chaperone or protein-quality-control responses.
Carbohydrate, cell-wall, and secondary-metabolic processes	Limited support	Detected	May reflect cell-wall remodeling and metabolic adjustment during pathogen challenge.
Transport and membrane-associated processes	Detected	Detected	Supports roles for membrane transport and signaling during early infection.

Unmapped or poorly annotated lychee genes	Present	Present	Highlights the need for a complete lychee-to-model-species orthology table before formal Reactome FDR enrichment.

Table 3. Orthology-guided Plant Reactome pathway screen applied to all significant pairwise DEGs.

Note: The input consisted of the complete significant DEG sets reported in Section 2.2, not only the selected candidate genes. Because the Plant Reactome export examined here did not contain direct *Litchi chinensis* entries for the LITCHI identifiers, results are reported descriptively rather than as formal FDR-corrected over-representation statistics.

Structural, motif, and promoter analysis of selected genes

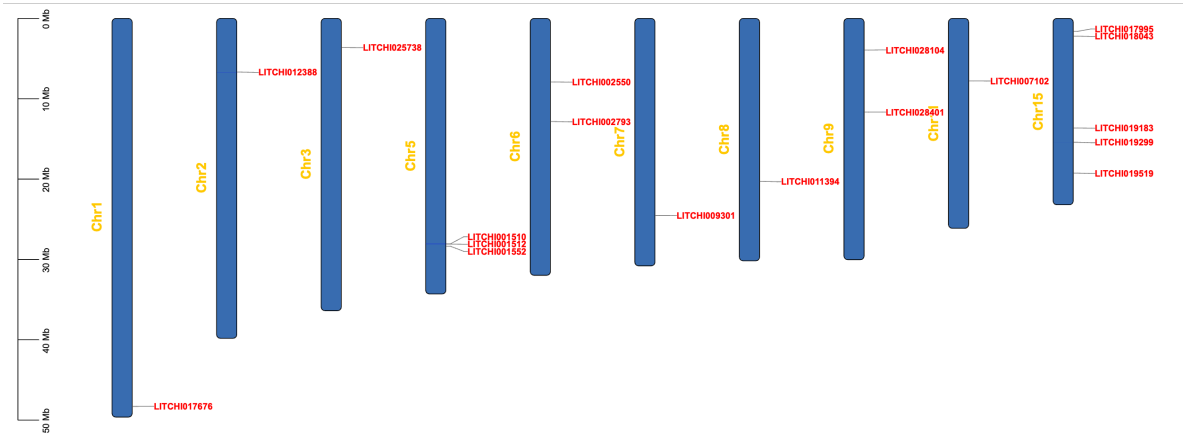

Figure 5. Chromosomal locations of selected differentially expressed genes.

The selected genes were distributed across 10 of the 15 lychee chromosomes (Figure 5). Most chromosomes contained one selected gene, whereas chromosome 15 contained five selected genes and chromosome 5 contained three selected genes.

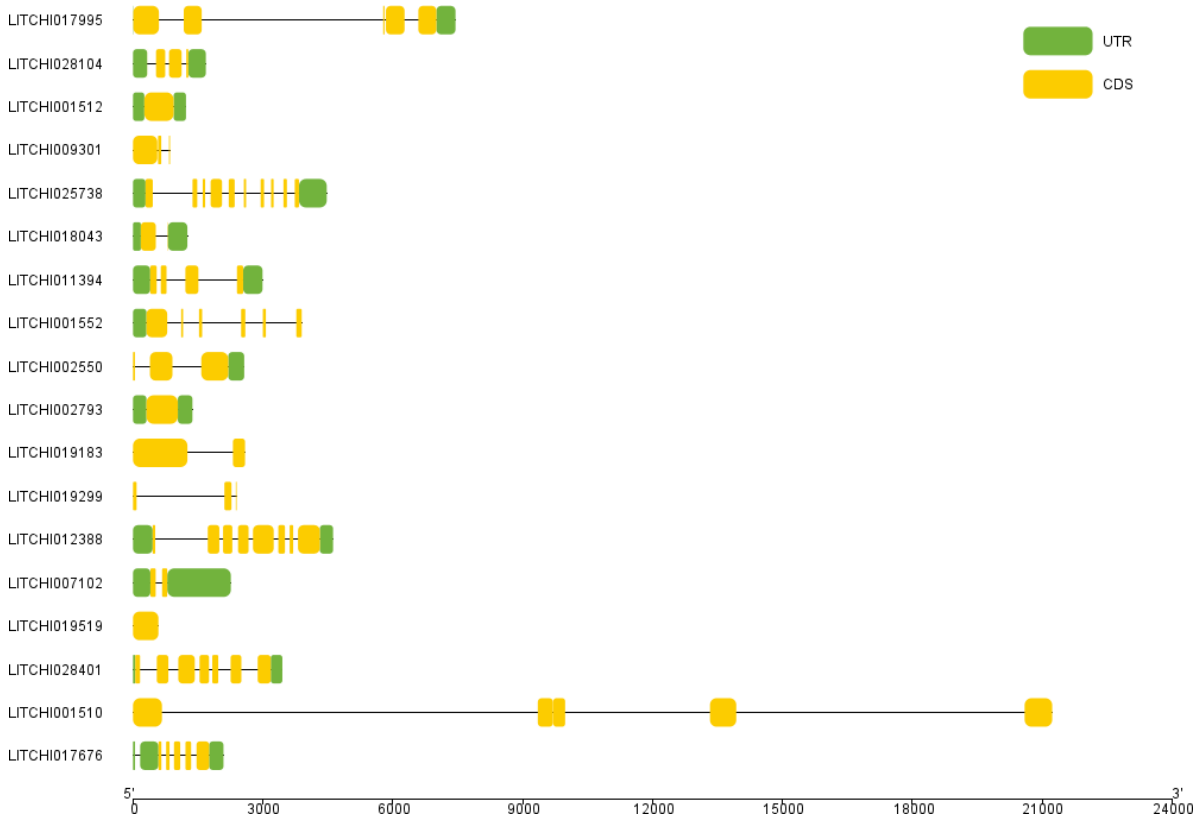

Figure 6. Gene structures of selected differentially expressed genes.

Because many lychee genes remain incompletely characterized, we further examined the structures of the selected genes (Figure 6). Yellow segments represent exons, and solid lines represent introns. Exon numbers varied substantially among genes. For example, LITCHI025738 contained 10 exons, whereas LITCHI007102 contained 2 exons. LITCHI001510 was the longest among the selected genes, but most of its length corresponded to intronic regions. By contrast, LITCHI025738, LITCHI012388, and LITCHI028104 contained multiple small exons.

Conserved-motif analysis identified 20 motifs among the selected genes (Figure 7). Most genes contained only a small number of predicted motifs, with the exception of LITCHI009301, LITCHI002793, and LITCHI001510. Motif2, Motif5, and Motif7 were assigned to the thaumatin family, consistent with the presence of stress- and defense-related proteins among the selected genes.

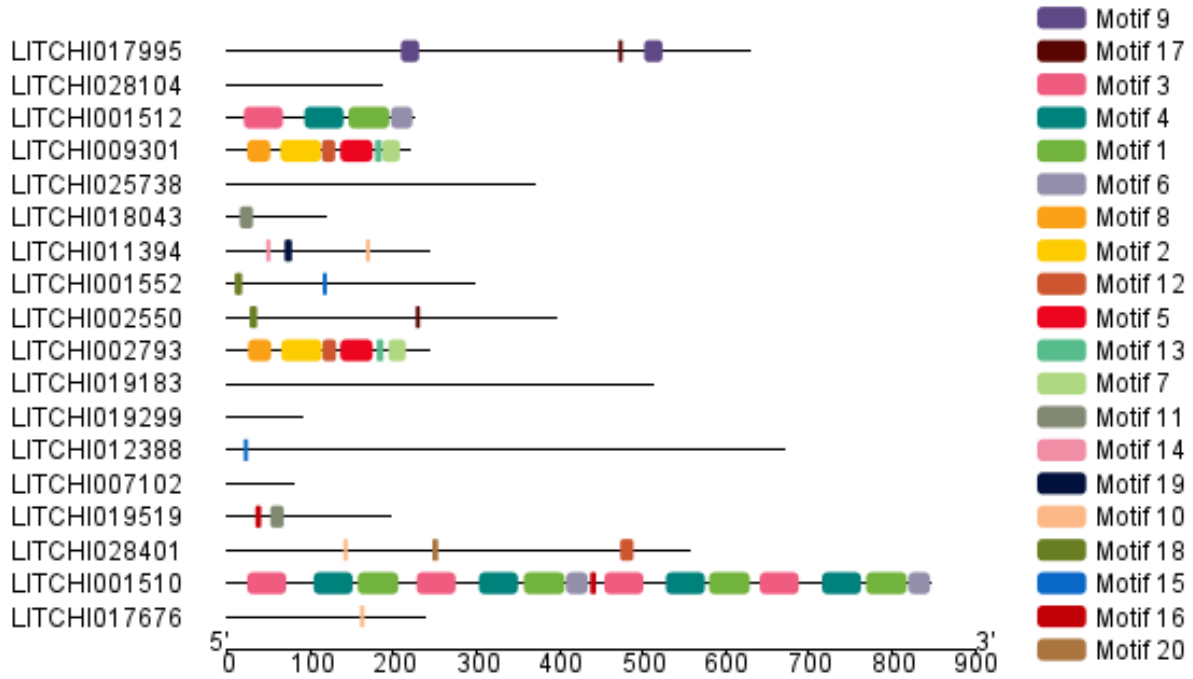

Figure 7. Distribution of predicted conserved motifs in selected differentially expressed genes.

Cis-acting regulatory elements are noncoding DNA sequences that influence the timing, location, and magnitude of gene expression and can contribute to plant responses to environmental stimuli [11].

Promoter regions 2000 bp upstream of the selected genes were examined for cis-acting regulatory elements. We focused on seven elements with potential relevance to plant stress or defense responses: TCA-elements, ABRE, ARE, MeJA-responsive elements, TC-rich repeats, AT-rich sequences, and MBS. These elements were broadly distributed across most selected genes (Figure 8), except LITCHI002793, which contained fewer detected elements in the analyzed promoter region. Figure 9 summarizes the number of occurrences of each element in each selected gene.

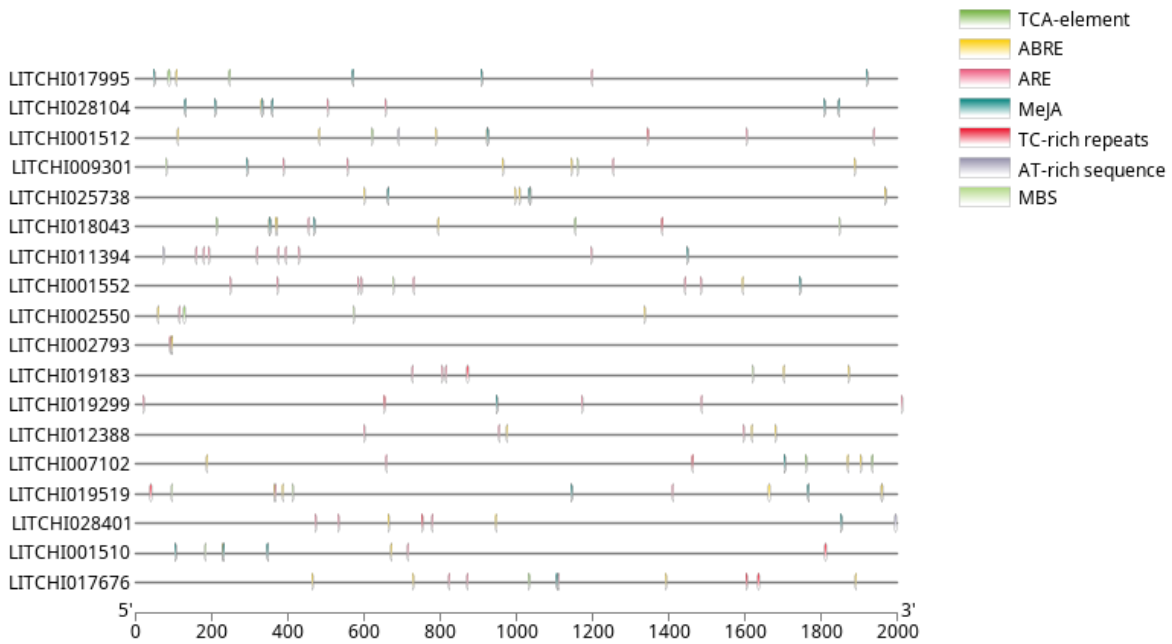

Figure 8. Locations of selected cis-acting regulatory elements in promoters of selected differentially expressed genes.

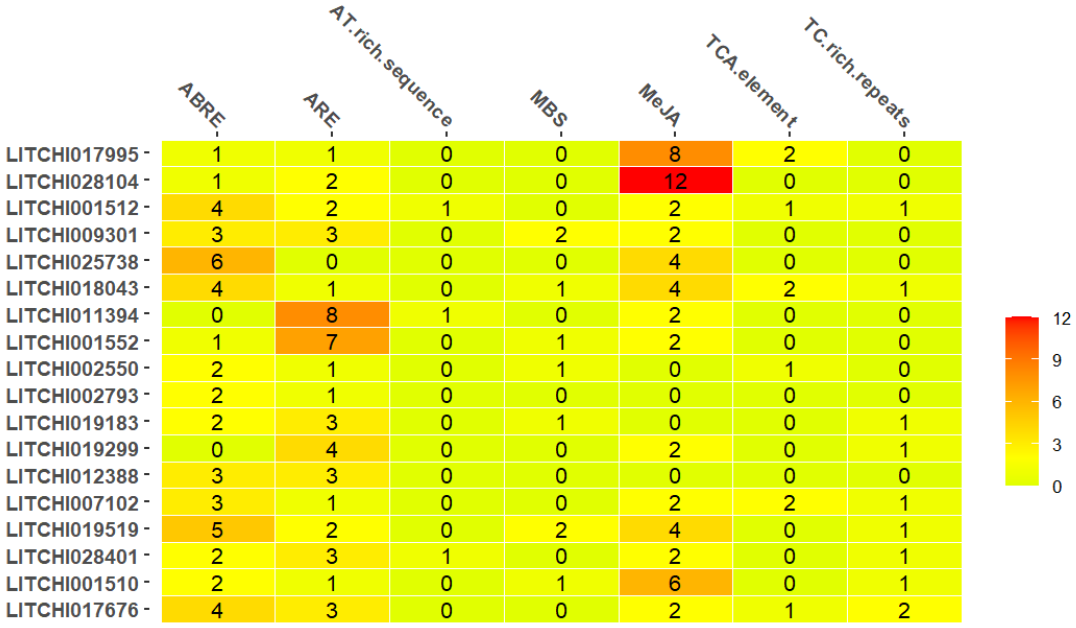

Figure 9. Counts of selected cis-acting regulatory elements in promoters of selected differentially expressed genes.

ABRE occurred frequently in the promoters of LITCHI025738 and LITCHI019519; ARE occurred frequently in LITCHI011394 and LITCHI001552; and MeJA-responsive elements were common in LITCHI017995, LITCHI028104, and LITCHI001510. The remaining cis-acting

regulatory elements (AT-rich sequences, MBS, TCA-elements, and TC-rich repeats) appeared more sporadically among the selected genes.

Because these cis-elements differ in motif length and background frequency, the raw counts were compared against a composition-matched random-promoter background. The observed counts from the 18 selected promoters were tested against 100,000 random sets of 18 simulated 2-kb promoters. Motif variants and reverse complements were counted, and one-sided empirical P values were adjusted by Benjamini-Hochberg FDR.

Element	Observed total	Random mean	Empirical P / FDR q	Interpretation
ABRE	45	38.44	0.16402 / 0.22963	Frequent but not enriched
ARE	46	24.61	0.00011 / 0.00026	Enriched
AT-rich sequence	3	1.71	0.24643 / 0.28750	Not enriched
MBS	9	12.68	0.88396 / 0.88396	Not enriched
MeJA-responsive	54	38.44	0.01026 / 0.01795	Enriched
TCA-element	9	0.19	0.00001 / 0.00003	Enriched
TC-rich repeats	10	0.15	0.00001 / 0.00003	Enriched

Table 4. Background-controlled random-promoter comparison of cis-regulatory element counts across the 18 selected promoters.

Note: Observed totals are summed across the 18 selected promoters in Figure 9. The random mean is the average motif count across 100,000 GC-matched random promoter sets. P values are one-sided empirical probabilities that a random set contained at least as many hits as observed.

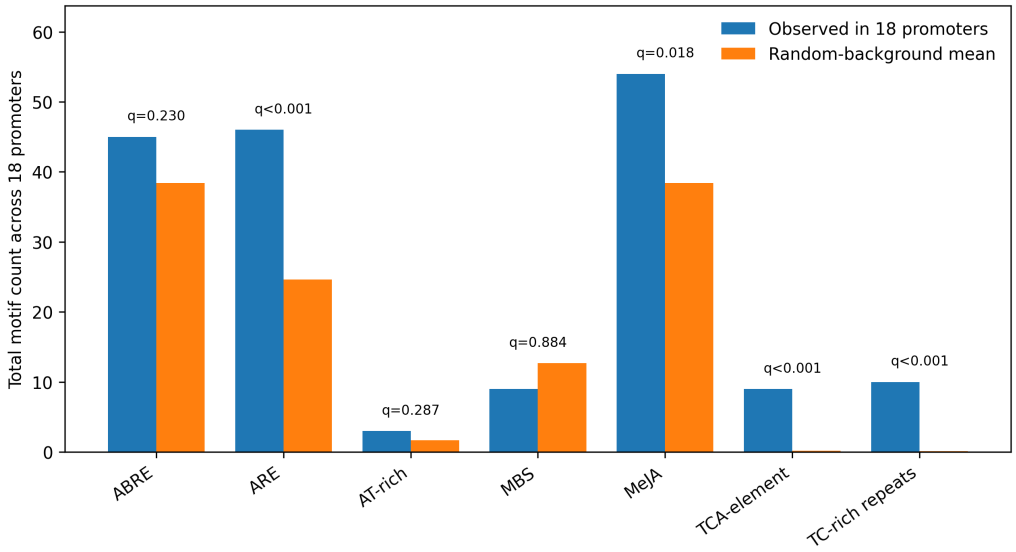

Figure 10. Background-controlled promoter comparison for selected cis-regulatory elements. Bars show observed total counts across the 18 selected promoters and expected mean counts under the random-promoter background; labels above bars show FDR-adjusted q values.

This background control showed that ARE, MeJA-responsive elements, TCA-elements, and TC-rich repeats were enriched relative to the random-promoter background. In contrast, ABRE and MBS were present in the selected promoters but were not significantly enriched after FDR correction. Thus, the promoter signal appears to be driven by a subset of hormone- and defense-associated motifs rather than by the general abundance of common cis-elements alone.

DISCUSSION

The volcano plots (Figure 2) show that Yurong had more differentially expressed genes than Guiwei, indicating cultivar-specific transcriptional responses to P. litchii. Liu et al. also reported that several CDPK-family genes were highly expressed after P. litchii inoculation and noted that Yurong is relatively resistant to P. litchii [6]. Our results are consistent with the idea that Yurong may mount a broader transcriptional response during infection.

The interaction analysis adds information that is not captured by separate within-cultivar comparisons alone. Separate comparisons identify genes that change after inoculation within each cultivar, whereas the interaction contrast identifies genes for which the magnitude of the inoculation response differs between cultivars. The strong negative interaction contrasts for LITCHI019519, LITCHI001510, LITCHI028401, and LITCHI017676 suggest Guiwei-skewed responses among the selected genes, while the positive contrasts for LITCHI028104 and LITCHI019183 suggest Yurong-skewed responses. Because Yurong has been described as relatively resistant to P. litchii [6], Yurong-skewed candidates are especially relevant for future validation, although Guiwei-skewed genes may also reveal susceptibility-associated or compensatory responses.

The addition of the all-DEG Plant Reactome screen reduces the bias introduced by interpreting only the most strongly changing genes. Although the selected 18 candidates include plausible defense-associated proteins, the complete DEG sets suggest that cultivar responses are better interpreted as coordinated pathway-level changes involving signaling, regulation, proteostasis, transport, and metabolism. The larger Yurong DEG set provided broader pathway coverage than Guiwei, suggesting a more extensive transcriptional adjustment at 24 h post-inoculation. Because this Reactome step is orthology-guided, it should be used for hypothesis generation rather than as final pathway-level statistical evidence.

Figures 5 and 6 show the chromosomal locations and gene structures of the selected genes. Table 5 summarizes their predicted functions, which were grouped into stress response, calcium binding, enzyme-related, and other categories. To make the prioritization transparent, Supplementary Table S1 separates the expression evidence for each selected gene from

functional interpretation, including annotation, representative homolog/domain information, and known defense relevance.

LITCHI028104 was annotated as a jacalin-like lectin, a class of proteins that can be associated with antiviral or defense-related functions [12]. LITCHI011394 was annotated as a pathogenesis-related protein. LITCHI012388 belongs to the heat-shock protein 70 (HSP70) family, members of which are often associated with stress responses [13]. LITCHI001552 and LITCHI025738 were annotated as cyclase-like proteins and may be involved in abiotic-stress responses [14].

LITCHI009301 and LITCHI002793 were annotated as osmotin/thaumatin-like proteins. Thaumatin is a sweet-tasting plant protein originally identified in Thaumatococcus daniellii [15]. Osmotin and other PR-5 family proteins are homologous to thaumatin and have been associated with antifungal activity against plant pathogens [16]. The presence of thaumatin/osmotin-like candidates supports a possible role for defense-related proteins in the response to P. litchii.

LITCHI017995 was annotated as a metalloendoproteinase, a zinc-dependent protease that may be calcium activated, whereas LITCHI019519 was associated with EF-hand calcium binding. These genes may be connected to calcium-dependent signaling or protein-processing pathways.

Two enzyme-related annotations were also observed. LITCHI002550 belongs to the cytochrome P450 family, which participates in diverse biosynthetic and detoxification pathways. LITCHI019299 and LITCHI017676 were annotated as S-adenosylmethionine-related genes, which may be involved in methylation or polyamine-biosynthesis pathways.

Among the remaining genes, LITCHI007102 was annotated as a plant-specific phytosulfokine precursor, and LITCHI019183 and LITCHI018043 were annotated as plant lipid-transfer proteins.

Some selected genes remain poorly characterized. In particular, LITCHI001510 and LITCHI001512 lack clear functional annotations and should be prioritized carefully in future experimental studies.

Gene(s)	**Predicted/annotated function**	**Functional category**
LITCHI028104	Jacalin-like lectin	Stress response
LITCHI011394	Pathogenesis-related protein family	Stress response
LITCHI012388	Heat shock protein	Stress response
LITCHI001552, LITCHI025738	Cyclase-like protein	Stress response
LITCHI009301, LITCHI002793	Osmotin/thaumatin-like proteins	Stress response
LITCHI019519	EF-hand calcium-binding protein	Calcium binding/signaling
LITCHI017995	Metalloendoproteinase	Calcium binding/signaling

Gene(s)	Predicted/annotated function	Functional category
LITCHI002550	Cytochrome P450	Enzyme-related
LITCHI019299, LITCHI017676	S-adenosylmethionine-related proteins	Enzyme-related
LITCHI007102	Phytosulfokine precursor	Other
LITCHI019183, LITCHI018043	Plant lipid-transfer proteins	Other

Table 5. Predicted functions of selected differentially expressed genes.

Thaumatin-family motifs were also detected in Figure 7. This protein family includes pathogenesis-related group 5 (PR-5) proteins, which are associated with systemic acquired resistance, stress responses, and antifungal activity [17].

Figures 8 and 9 summarize raw locations and counts of seven cis-acting regulatory elements related to plant stress or defense responses. Table 4 and Figure 10 add a background-controlled random-promoter comparison, and Table 6 lists the reported functions of the selected elements.

ABRE elements are involved in abscisic acid responsiveness [18,19]. Abscisic acid (ABA) plays a major role in plant responses to environmental stress, and ABA-mediated responses can induce expression of many genes through ABRE-containing promoters [20]. In this study, ABREs were often observed as the ACGTG motif.

ARE and MBS are additional cis-regulatory elements associated with responses to environmental conditions. In this study, AREs were often observed as the AAACCA motif, whereas MBS motifs were often observed as CAACTG.

AT-rich sequences with the TAAAATACT motif were identified in a subset of selected genes. These elements have been associated with maximal elicitor-mediated activation. Elicitors are compounds that can activate plant chemical defenses [21].

Methyl jasmonate (MeJA) is a signaling molecule involved in plant development and stress responses. It can activate plant defense mechanisms against a range of pathogens. For example, Zhang et al. reported that MeJA-responsive proteins play roles in maize responses to pests and insects [22]. Several selected lychee genes contained multiple MeJA-responsive elements, typically CGTCA or TGACG motifs, in their promoter regions.

TCA-elements are associated with salicylic acid responsiveness. The hypersensitive response can limit pathogen spread through localized programmed cell death, and salicylic-acid-mediated defense can contribute to systemic acquired resistance and broad-spectrum immunity [8].

TC-rich repeats are known to participate in plant defense and stress responsiveness [23]. In this study, TC-rich repeats were detected as the GTTTTCTTAC motif.

Cis-regulatory element	Reported function
ABRE	ABA-responsive element involved in abscisic acid responsiveness

Cis-regulatory element	Reported function
ARE	Element associated with anaerobic induction
AT-rich sequence	Element associated with maximal elicitor-mediated activation
MBS	MYB binding site involved in drought inducibility
MeJA	Element involved in MeJA responsiveness
TCA-element	Element involved in salicylic acid responsiveness
TC-rich repeats	Element involved in defense and stress responsiveness

Table 6. Reported functions of selected cis-regulatory elements in promoters of selected differentially expressed genes.

The cis-acting regulatory elements identified in this study are commonly associated with stress responses. Similar sets of elements have been reported in other plant systems, including maize [24]. The random-promoter comparison refined this interpretation by showing that not every frequent element was enriched: ABRE was common in the candidate promoters but did not pass the background-controlled FDR threshold, whereas ARE, MeJA-responsive elements, TCA-elements, and TC-rich repeats remained enriched.

Several limitations should be noted. First, this work is a secondary analysis of public transcriptomic data and does not include independent qPCR or functional validation. Second, the structural, motif, and promoter analyses focus on 18 strongly responsive genes; genes with smaller but biologically important expression changes may have been missed. Third, the cultivar-by-treatment analysis is presented as a selected-gene effect-size screen rather than a definitive genome-wide interaction DEG analysis. Fourth, the pathway analysis was orthology-guided because the LITCHI identifiers were not directly represented in the Plant Reactome export, so pathway results are descriptive. Fifth, the promoter background model controls for motif length and approximate GC content but does not capture full promoter architecture or genome-wide promoter clustering. These limitations frame the results as a prioritized, testable candidate set.

MATERIALS AND METHODS

Public data source, reference files, and study design

RNA-seq data were analyzed from GEO accession GSE201243 [9], which contains a balanced 2 x 2 design consisting of two lychee cultivars (Guiwei and Yurong), two treatments (sterile-water mock control and P. litchii inoculation), and three biological replicates per cultivar-treatment group. Inoculated leaf samples were collected at 24 h post-inoculation. Raw reads were obtained from the SRA runs linked to GSE201243/BioProject PRJNA830488 using SRA Toolkit v3.0.7 with prefetch followed by fasterq-dump --split-files --threads 8. The sample metadata table contained the fields sample_id, cultivar, treatment, replicate, GEO accession, SRA run, SRA spots, and total bases. For all factorial analyses, cultivar was encoded as Guiwei

or Yurong, treatment was encoded as mock or inoculated, and Guiwei and mock were used as the reference levels.

The reference assembly was Litchi chinensis cv. Feizixiao SCAU_Lch_v2.0 (NCBI assembly GCA_019925255.1; BioProject PRJNA747875) with the corresponding GFF3 annotation from the same genome release [25]. Before alignment and counting, the GFF3 file was checked to ensure that gene identifiers matched the LITCHI gene IDs used in the manuscript. When needed, the annotation was converted to GTF format with gffread v2.2.1 while retaining gene_id attributes for HTSeq counting.

Read quality control, trimming, alignment, and read counting

Raw-read quality was evaluated with FastQC v0.11.9 [26]. Adapter and low-quality bases were trimmed with Trimmomatic v0.39 [27] in paired-end mode using the following reproducible parameter set: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, and MINLEN:50. Trimmed FASTQ files were rechecked with FastQC. These settings remove standard Illumina paired-end adapters, trim low-quality bases from read ends, apply a sliding-window quality threshold of Q20, and discard reads shorter than 50 bp after trimming.

For consistency with the public-study workflow, trimmed paired-end reads were aligned to the SCAU_Lch_v2.0 reference genome using TopHat v2.0.12 with Bowtie2 v2.2.3 [28]. The Bowtie2 index was built from the genome FASTA file. The alignment command used eight threads and the following main parameters: tophat2 -p 8 --library-type fr-unstranded --read-mismatches 2 --read-gap-length 2 --read-edit-dist 2 --min-intron-length 50 --max-intron-length 50000 --max-multihits 20 --GTF annotation.gtf -o sample_tophat_out genome_index sample_R1.fq.gz sample_R2.fq.gz. BAM files were sorted and indexed with SAMtools v1.17. TopHat is no longer the most current RNA-seq aligner, but it was retained to maintain consistency with the original public processing strategy. Future analyses using HISAT2 or STAR should regenerate counts and DESeq2 contrasts from a single updated pipeline.

Gene-level counts were generated from sorted BAM files with HTSeq v2.0.2 [29] using union mode and non-stranded counting: htseq-count -f bam -r pos -s no -t exon -i gene_id -m union -a 10 --nonunique none sample.sorted.bam annotation.gtf > sample.counts.txt. Individual count files were merged into a gene-by-sample integer count matrix using Python v3.10.13 and pandas v2.1.4. Genes with low read support were filtered before DESeq2 analysis using rowSums(counts >= 10) >= 3, which retains genes with at least 10 counts in at least three libraries.

Differential expression, interaction modeling, and sample-level QC

Differential-expression analyses were performed in R v4.3.2 using Bioconductor v3.18 and DESeq2 v1.42.0 [30]. DESeq2 size factors were estimated with the median-of-ratios method, dispersions were estimated with the default parametric trend, and Wald tests were used for pairwise contrasts. P values were adjusted for multiple testing by the Benjamini-Hochberg method, and genes with adjusted P values below 0.05 were considered differentially expressed. Unless otherwise stated, no additional minimum absolute log2 fold-change cutoff was imposed for the primary DEG calls; fold-change magnitude was used only for candidate prioritization and visualization.

Pairwise infection responses were analyzed separately within each cultivar using design = ~ treatment and the contrast inoculated versus mock. To test whether the infection response differed between cultivars, the full 12-sample factorial design was analyzed with design = ~ cultivar + treatment + cultivar:treatment after releveling cultivar = Guiwei and treatment = mock as references. In this parameterization, the treatment coefficient estimates the infection response in Guiwei, while the cultivarYurong:treatmentinoculated coefficient estimates the interaction term, equivalent to log2FC(Yurong inoculated versus mock) - log2FC(Guiwei inoculated versus mock).

Sample-level expression QC in the reproducible workflow used variance-stabilizing transformation with vst(dds, blind = FALSE), followed by principal component analysis with prcomp and hierarchical clustering from Euclidean distances with complete linkage. DESeq2 Cook distance diagnostics and sample clustering were used to screen for libraries with unusual influence or inconsistent replicate behavior. No library was excluded based on the SRA run-depth metadata summarized in Section 2.1.

Plant Reactome, homolog, and domain annotation

Plant Reactome pathway screening was performed for all significant pairwise DEGs rather than only for the 18 selected candidate genes. The input DEG lists were the complete Guiwei and Yurong significant gene sets defined by adjusted $P < 0.05$ in the pairwise DESeq2 comparisons. Plant Reactome version 24/Gramene release 68 was used as the pathway reference [10]. Because the Plant Reactome gene-product export examined for this revision did not contain direct Litchi chinensis LITCHI identifiers, lychee proteins were mapped to pathway-annotated plant proteins through an orthology-guided screen rather than treated as direct Plant Reactome entries.

Representative homologs were assigned with DIAMOND v2.1.8 blastp [31] against pathway-annotated rice/Arabidopsis protein sets using --sensitive -e 1e-5 --query-cover 50 --subject-cover 50 --max-target-seqs 5. For each lychee gene, the representative homolog was selected

by highest bit score, with e-value and alignment coverage used as tie-breakers. Protein domains were annotated with InterProScan v5.65-97.0 [32] using the Pfam application and the options -appl Pfam -f TSV -goterms -iprlookup. Domain and homolog information was used to support candidate-gene annotation but was not treated as experimental functional validation.

Candidate-gene structure, conserved motifs, and promoter analysis

TBtools-II v2.096 [33] was used to extract gene coordinates from the GFF3 file, display chromosomal locations, visualize exon-intron structures, and extract 2,000 bp upstream promoter sequences. Promoter extraction was strand-aware: for genes on the negative strand, the upstream sequence was reverse-complemented relative to the annotated gene orientation. Genes located within 2,000 bp of a contig boundary were retained with the available upstream sequence, and the resulting promoter length was recorded.

For conserved-motif analysis, predicted coding sequences for the selected genes were translated to protein FASTA sequences with TBtools-II. MEME Suite v5.5.5 [34] was then run in protein mode with the command structure meme selected_proteins.fa -protein -mod zoops -nmotifs 20 -minw 6 -maxw 50 -evt 0.05 -seed 2024 -oc meme_out. The zoops model was selected because not every motif was expected to occur in every candidate protein. Predicted motifs were interpreted with Pfam/InterProScan domain assignments and the available lychee gene annotations.

Promoter sequences were submitted to PlantCARE [35] to identify cis-acting regulatory elements. PlantCARE output was filtered to retain elements annotated as stress-, defense-, pathogen-, wound-, hormone-, drought-, anaerobic-, or light-responsive. Synonymous or duplicated labels were merged before summarization, and both total motif occurrence and gene-level presence were considered. Elements retained for biological interpretation included ARE, MeJA-responsive motifs, TCA-elements, TC-rich repeats, ABRE, MBS, and AT-rich sequence classes, with interpretation refined by the background analysis described below.

Promoter-background test and figure generation

To determine whether candidate-promoter motif counts exceeded a simple background expectation, a permutation-style random-promoter test was implemented in Python v3.10.13 using NumPy v1.26.4, pandas v2.1.4, SciPy v1.11.4, and Matplotlib v3.8.2. The simulation generated 100,000 random promoter sets. Each set contained 18 promoters of 2,000 bp, matching the number and length of the candidate promoters. Random sequences were generated with a lychee-like GC content of 34% (A = 0.33, T = 0.33, G = 0.17, C = 0.17). Exact motif strings and their reverse complements were scanned so that motif counts were not biased by orientation. One-sided empirical P values were computed as (1 + number of random sets

with count >= observed count)/(1 + number of random sets), and P values were adjusted across tested motif classes by the Benjamini-Hochberg method.

Figures were generated from scripted analyses rather than manual redrawing. Volcano plots, heatmaps, PCA plots, interaction-effect plots, and pathway summaries were generated in R with ggplot2 v3.4.4 [36], ggrepel v0.9.4, pheatmap v1.0.12, and base R plotting functions. Promoter-background plots were generated in Python with Matplotlib v3.8.2. Final figures were exported as PNG or TIFF files at 600 dpi where possible and assembled in the manuscript without changing the underlying numerical values.

Code availability and reproducibility controls

The computational workflow was organized as Bash, R, and Python scripts with fixed random seeds for stochastic analyses. The script package includes: 01_download_and_qc.sh for SRA retrieval and read QC; 02_alignment_and_counts.sh for trimming, alignment, BAM sorting, and HTSeq counting; 03_deseq2_pairwise_interaction.R for pairwise DESeq2 contrasts, the cultivar-by-treatment model, PCA, clustering, and DEG tables; 04_promoter_background.py for the random-promoter test; and 05_figure_generation.R for manuscript figures. Supplementary File S2 provides command templates and session-information records for rerunning the analysis after updating local file paths. The same package lists the expected derivative files: raw count matrix, sample metadata, complete DESeq2 result tables, candidate-gene table, Plant Reactome mapping table, promoter motif table, and random-background simulation output.

Analysis step	**Software/version**	**Main parameters, filters, or outputs**
Data retrieval	SRA Toolkit v3.0.7	prefetch; fasterq-dump --split-files --threads 8; metadata fields: sample_id, cultivar, treatment, replicate, GEO accession, SRA run, spots, and bases.
Read QC/trimming	FastQC v0.11.9; Trimmomatic v0.39	Paired-end mode; ILLUMINACLIP:TruSeq3-PE.fa:2:30:10; LEADING:3; TRAILING:3; SLIDINGWINDOW:4:20; MINLEN:50.
Reference preparation	SCAU_Lch_v2.0 genome; gffread v2.2.1	Genome assembly GCA_019925255.1; GFF3 checked for LITCHI gene IDs; GFF3 converted to GTF when needed; Bowtie2 index built from genome FASTA.
Read alignment	TopHat v2.0.12; Bowtie2 v2.2.3; SAMtools v1.17	-p 8; --library-type fr-unstranded; --read-mismatches 2; --read-gap-length 2; --read-edit-dist 2; --min-intron-length 50; --max-intron-length 50000; --max-multihits 20; --GTF annotation.gtf.
Read counting	HTSeq v2.0.2	htseq-count -f bam -r pos -s no -t exon -i gene_id -m union -a 10; --nonunique none; merged count matrix.
DESeq2 pairwise tests	R v4.3.2; Bioconductor v3.18; DESeq2	Filter rowSums(counts >= 10) >= 3; design = ~ treatment within cultivar; median-of-ratios normalization; Wald test;

Analysis step	Software/version	Main parameters, filters, or outputs
	v1.42.0	Benjamini-Hochberg adjusted P < 0.05.
Interaction test	DESeq2 v1.42.0	Full design = ~ cultivar + treatment + cultivar:treatment; reference levels Guiwei and mock; interaction = log2FC(Yurong) - log2FC(Guiwei).
Pathway and homolog screen	Plant Reactome v24/Gramene 68; DIAMOND v2.1.8	All significant pairwise DEGs used as inputs; DIAMOND blastp --sensitive -e 1e-5 --query-cover 50 --subject-cover 50 --max-target-seqs 5; best hit by bit score.
Protein domains and motifs	MEME Suite v5.5.5; InterProScan v5.65-97.0/Pfam	MEME -protein -mod zoops -nmotifs 20 -minw 6 -maxw 50 -evt 0.05 -seed 2024; InterProScan -appl Pfam -f TSV -goterms -iprlookup.
Promoter and cis-element analysis	TBtools-II v2.096; PlantCARE	Extract 2,000 bp strand-aware upstream promoters; PlantCARE elements filtered for stress, defense, hormone, drought, wound, anaerobic, and light-response terms; synonymous motif labels merged.
Promoter background and figures	Python v3.10.13; NumPy v1.26.4; pandas v2.1.4; SciPy v1.11.4; Matplotlib v3.8.2; ggplot2 v3.4.4	100,000 random sets of 18 x 2-kb promoters; GC = 34%; exact motifs plus reverse complements; empirical P values with +1 correction; Benjamini-Hochberg FDR; figures exported at 600 dpi where possible.

Table 7. Software versions and principal parameters specified for the reproducible 2024 RNA-seq reanalysis workflow.

Note: The table documents the parameter set used for the RNA-seq reanalysis. If any pipeline component is changed, counts, differential-expression statistics, and figures should be regenerated from the same updated workflow.

CONCLUSIONS

By reanalyzing public RNA-seq data for Guiwei and Yurong lychee leaves after *P. litchii* inoculation, this study identifies cultivar-specific transcriptional patterns and prioritizes candidate genes with defense-related annotations, including lectin, pathogenesis-related protein, thaumatin/osmotin-like, calcium-binding, and lipid-transfer protein families. The selected-gene interaction contrast highlights both Guiwei- and Yurong-skewed responses, while promoter-background testing suggests that ARE, MeJA-responsive, TCA, and TC-rich elements are enriched in the promoters of the selected candidates. These findings should be viewed as a hypothesis-generating resource for qRT-PCR validation, functional assays, and marker development in lychee resistance research.

SUPPLEMENTARY MATERIALS

Supplementary Table S1: Rationale for selection of the 18 candidate differentially expressed genes, including rounded log2FC evidence, adjusted-P threshold status, predicted annotations, representative homolog/domain information, known defense relevance, and reason retained.

Supplementary File S2. Reproducible workflow command templates and code-availability package, including Bash, R, and Python script templates for raw-read processing, count-table generation, DESeq2 pairwise and interaction analyses, Plant Reactome/homolog mapping, promoter-background simulation, and figure generation.

Supplemental materials are available at https://zenodo.org/records/20007391

Data Availability Statement: The RNA-seq data analyzed in this study are publicly available in the NCBI Gene Expression Omnibus under accession GSE201243, with raw reads available through the linked SRA records and BioProject PRJNA830488. The reference genome resource is available through NCBI BioProject PRJNA747875 and the SCAU_Lch_v2.0 assembly record. Plant Reactome mapping resources are publicly available from the Plant Reactome download portal. Derived tables and command templates supporting the analyses are provided in the Supplementary Materials; additional intermediate files are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the researchers who generated and deposited the public RNA-seq dataset analyzed in this study.

REFERENCES

1. Morton, J.F. Lychee. In Fruits of Warm Climates; Creative Resource Systems, Inc.: Winterville, NC, USA, 1987; pp. 249-259.
2. Sun, J.; Gao, Z.; Zhang, X.; Zou, X.; Cao, L.; Wang, J. Transcriptome analysis of Phytophthora litchii reveals pathogenicity arsenals and confirms taxonomic status. PLoS ONE 2017, 12, e0178245. https://doi.org/10.1371/journal.pone.0178245.
3. Yi, C.; Jiang, Y.; Shi, J.; Qu, H.; Xue, S.; Duan, X.; Shi, J.; Prasad, N.K. ATP-regulation of antioxidant properties and phenolics in litchi fruit during browning and pathogen infection process. Food Chemistry 2010, 118, 42-47. https://doi.org/10.1016/j.foodchem.2009.04.074.
4. Hu, G.; Feng, J.; Xiang, X.; Wang, J.; Salojarvi, J.; Liu, C.; Wu, Z.; Zhang, J.; Liang, X.; Jiang, Z.; et al. Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. Nature Genetics 2022, 54, 73-83. https://doi.org/10.1038/s41588-021-00971-3.
5. National Center for Biotechnology Information. Litchi chinensis BioProject PRJNA747875. Available online: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA747875/ (accessed on 1 May 2026).

6. Liu, H.; Yan, Q.; Jiang, Y.; Shi, F.; Chen, J.; Cai, C.; Ou, L. Identification of LcCDPKs and analysis of their expression patterns in response to downy mildew stresses in lychee. Journal of Fruit Science 2023, 40, 442-456. https://doi.org/10.13925/j.cnki.gsxb.20220307.
7. Li, P.; Li, W.; Zhou, X.; Situ, J.; Xie, L.; Xi, P.; Yang, B.; Kong, G.; Jiang, Z. Peronophythora litchii RXLR effector PlAvh202 destabilizes a host ethylene biosynthesis enzyme. Plant Physiology 2023, 193, 756-774. https://doi.org/10.1093/plphys/kiad311.
8. Sun, J.; Cao, L.; Li, H.; Wang, G.; Wang, S.; Li, F.; Zou, X.; Wang, J. Early responses given distinct tactics to infection of Peronophythora litchii in susceptible and resistant litchi cultivar. Scientific Reports 2019, 9, 2810. https://doi.org/10.1038/s41598-019-39100-w.
9. National Center for Biotechnology Information. GEO Accession GSE201243. Available online: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE201243 (accessed on 1 May 2026).
10. Naithani, S.; Gupta, P.; Preece, J.; D'Eustachio, P.; Elser, J.; Garg, P.; Monaco, M.K.; Fabregat, A.; Jaiswal, P. Plant Reactome Knowledgebase: empowering plant pathway exploration and OMICS data analysis. Nucleic Acids Research 2024, 52, D1538-D1547. https://doi.org/10.1093/nar/gkad1052.
11. Marand, A.P.; Eveland, A.L.; Kaufmann, K.; Springer, N.M. cis-Regulatory Elements in Plant Development, Adaptation, and Evolution. Annual Review of Plant Biology 2023, 74, 111-137. https://doi.org/10.1146/annurev-arplant-070122-030236.
12. Esch, L.; Schaffrath, U. An update on jacalin-like lectins and their role in plant defense. International Journal of Molecular Sciences 2017, 18, 1592. https://doi.org/10.3390/ijms18071592.
13. Berka, M.; Kopecka, R.; Berkova, V.; Brzobohaty, B.; Cerny, M. Regulation of heat shock proteins 70 and their role in plant immunity. Journal of Experimental Botany 2022, 73, 1894-1909. https://doi.org/10.1093/jxb/erab549.
14. Qin, Y.; Wang, M.C.; Tian, Y.C.; He, W.X.; Han, L.; Xia, G.M. Characterization of a novel cyclase-like gene family involved in controlling stress tolerance in rice. Plant Physiology 2015, 168, 30-41.
15. Edens, L.; Heslinga, L.; Klok, R.; Ledeboer, A.M.; Maat, J.; Toonen, M.Y.; Visser, C.; Verrips, C.T. Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its expression in Escherichia coli. Gene 1982, 18, 1-12.

16. Min, K.; Ha, S.C.; Hasegawa, P.M.; Bressan, R.A.; Yun, D.J.; Kim, K.K. Crystal structure of osmotin, a plant antifungal protein. Proteins 2004, 54, 170-173. https://doi.org/10.1002/prot.10571.
17. Van Loon, L.C.; Rep, M.; Pieterse, C.M.J. Significance of inducible defense-related proteins in infected plants. Annual Review of Phytopathology 2006, 44, 135-162. https://doi.org/10.1146/annurev.phyto.44.070505.143425.
18. Hattori, T.; Totsuka, M.; Hobo, T.; Kagaya, Y.; Yamamoto-Toyoda, A. Experimentally determined sequence requirement of ACGT-containing abscisic acid response element. Plant and Cell Physiology 2002, 43, 136-140.
19. Fujita, Y.; Fujita, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. ABA-mediated transcriptional regulation in response to osmotic stress in plants. Journal of Plant Research 2011, 124, 509-525. https://doi.org/10.1007/s10265-011-0412-3.
20. Ingram, J.; Bartels, D. The molecular basis of dehydration tolerance in plants. Annual Review of Plant Physiology and Plant Molecular Biology 1996, 47, 377-403. https://doi.org/10.1146/annurev.arplant.47.1.377.
21. Seki, H.; et al. Combined effects of multiple cis-acting elements in elicitor-mediated activation of PSCHS1 gene. Plant and Cell Physiology 1997, 38, 96-100.
22. Zhang, Y.; et al. Maize MeJA-responsive proteins identified by high-resolution 2-DE PAGE. Data in Brief 2015, 5, 129-133.
23. Banerjee, J.; Sahoo, D.K.; Dey, N.; Houtz, R.L.; Maiti, I.B. An intergenic region shared by At4g35985 and At4g35987 in Arabidopsis thaliana is a tissue-specific and stress-inducible bidirectional promoter analyzed in transgenic Arabidopsis and tobacco plants. PLoS ONE 2013, 8, e79622. https://doi.org/10.1371/journal.pone.0079622.
24. Qian, Y.; et al. Genome-wide identification, classification and expression analysis of the JmjC domain-containing histone demethylase gene family in maize. BMC Genomics 2019, 20, 256.
25. Genome Warehouse. Litchi chinensis SCAU_Lch_v2.0 assembly record (GCA_019925255.1). National Genomics Data Center, China National Center for Bioinformation. Available online: https://ngdc.cncb.ac.cn/gwh/ (accessed on 1 May 2026).
26. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics, Cambridge, UK, 2010.

27. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 2014, 30, 2114-2120. https://doi.org/10.1093/bioinformatics/btu170.
28. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S.L. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 2013, 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36.
29. Anders, S.; Pyl, P.T.; Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015, 31, 166-169. https://doi.org/10.1093/bioinformatics/btu638.
30. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 2014, 15, 550. https://doi.org/10.1186/s13059-014-0550-8.
31. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. Nature Methods 2015, 12, 59-60. https://doi.org/10.1038/nmeth.3176.
32. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014, 30, 1236-1240. https://doi.org/10.1093/bioinformatics/btu031.
33. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. Molecular Plant 2020, 13, 1194-1202. https://doi.org/10.1016/j.molp.2020.06.009.
34. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, Stanford, CA, USA, 14-17 August 1994; AAAI Press: Menlo Park, CA, USA, 1994; pp. 28-36.
35. Lescot, M.; Dehais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouze, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 2002, 30, 325-327. https://doi.org/10.1093/nar/30.1.325.
36. Wickham, H. ggplot2: Elegant Graphics for Data Analysis; Springer-Verlag: New York, NY, USA, 2016.