Article

Transcriptome Analysis Provides Valuable Insights into Leaf Size Variation in *Rhamnus heterophylla*

Huiying Shang [1] , Lulu Xun [1], Tao Miao [2], Chen Chen [1] , Yuan Lu [1] and Bin Li [1,*]

[1] Xi'an Botanical Garden of Shaanxi Province, Institute of Botany of Shaanxi Province, Xi'an 710061, China; shanghuiying@xab.ac.cn (H.S.)

[2] Shaanxi Qingmuchuan National Nature Reserve, Hanzhong 723400, China

[*] Correspondence: lb@xab.ac.cn

Abstract: The size of leaves is a vital factor in the development and overall biomass of a plant, serving as a key indicator of how a plant adapts to its environment. *Rhamnus heterophylla*, a species known for its heteromorphic leaves of varying sizes, presents an intriguing case for studying leaf development at the molecular level. To gain insights for further studies on the underlying mechanisms, we constructed a comprehensive reference transcriptome database using both SMART sequencing and Illumina RNA-seq technologies. Our analysis of the transcriptome data identified 88,546 isoforms, featuring an N50 size of 2386 base pairs. Furthermore, we identified 2932 transcription factors from 55 gene families, along with 14,947 unigenes that underwent alternative splicing. By comparing the gene expression patterns between large and small leaves, we pinpointed 982 differentially expressed genes (DEGs). Among these DEGs, 116 genes exhibit significantly greater activity in small leaves, while 866 genes display significantly greater activity in large leaves. Functional enrichment analyses revealed the significant involvement of these DEGs in various hormone signaling pathways. Notably, we detected a significant decrease in the expression of several genes associated with auxin synthesis, such as *ARFs*, *GRF8*, and *IAA27*, in small leaves. This finding sheds light on their potential role in leaf size regulation in *R. heterophylla*, providing valuable insights into the genes underlying this mechanism.

Keywords: plant growth; functional enrichment analysis; differentially expressed genes; heteromorphic leaves

1. Introduction

Leaves, an essential plant organ, are the primary energy factories driving the growth and development of the organism, through the intricate mechanisms of photosynthesis and respiration. They are also reservoirs of organic matter and mineral nutrients, underscoring their crucial role in plant survival and reproduction. Consequently, the study of the underlying molecular mechanisms that regulate leaf size becomes crucial for gaining insights into leaf development and offers a theoretical backing for plant breeding efforts.

Leaf development has been extensively studied, and it is commonly categorized into three well-defined phases: the inception of leaf primordia, the establishment of dorsal–ventral polarity, and the expansion of the leaf blade [1–4]. At the outset of the process, leaf primordia are initiated at the outermost region of the shoot apical meristem (SAM) [5]. Following the differentiation of the leaf primordia, the leaf undergoes gradual changes in its morphology and structure. Initially, it possesses a rod-like structure, which gradually transforms into a flattened shape [6,7]. This transformation is crucial for the leaf to assume its functional form. Once the polarity of the leaf is established, the development process advances towards leaf extension [2,8]. Leaf extension is a pivotal late event in leaf development and morphogenesis, as it determines the ultimate morphological characteristics of the mature leaf. Throughout the progression of leaf development, cells initially undergo

continuous divisions, while maintaining a constant cell size. This early stage of cell division contributes to the increase in cell number. However, as the leaf development progresses, cell division gradually ceases, and cell expansion becomes the predominant process. The final size of the mature leaves is influenced significantly by both cell expansion, marked by an augmentation in cell size, and cell proliferation, which collectively contribute to this determination [8].

In the development process of plants, plant hormones and the genes responsible for hormone production play a vital role. For instance, auxins are known to promote cell elongation, which is a critical process during leaf development [9–11]. Gibberellins, on the other hand, participate in a range of processes, including seed germination, elongation of stems, and the development of fruits [12–14]. Cytokinins contribute to cell division, differentiation, the formation of lateral buds, and shoot growth [15–18]. Brassinosteroids regulate cell elongation and division, stress responses, and the formation of vascular tissues [19]. During leaf development, the differential expression of these hormones regulates cell division and growth, resulting in alterations to both the size and morphology of the leaf. Gaining insights into how these hormones function in leaf development can offer valuable perspectives on the molecular mechanisms that underpin plant growth and maturation, potentially guiding the creation of innovative approaches to enhance crop productivity and resilience against environmental stress. Recent studies have identified a multitude of genes and transcription factors involved in hormone biosynthesis that regulate leaf growth in plants. These factors play a crucial role in the control of leaf development, affecting both cell division and expansion in a fundamental manner. Auxin response factors (*ARFs*) constitute a group of proteins that interact with auxin signaling pathways to regulate cell division and differentiation, resulting in changes in leaf size, shape, venation patterning, and polarity [20–22]. Similarly, growth-regulating factors (*GRFs*) are involved in the regulation leaf growth, with different members of the *GRFs* family contributing to cell expansion, proliferation, and modulating the development of leaf primordia by interacting with *GIF1/AN3* [23]. The teosinte branched 1/cycloidea/proliferating cell factor (*TCP*) genes, which encode transcription factors, form an additional category crucial for orchestrating plant growth and developmental processes [24]. For example, in turnips, researchers have pinpointed 39 TCP genes. Elevated levels of *BrrTCP2* are linked to reduced leaf development, due to a decrease in the rate of cell multiplication. *BrpTCP4* is involved in shaping the dimensions and contours of Chinese cabbage heads, whereas *BrTCP1* enhances the process of cell splitting and enlargement in Chinese cabbage. [25]. Cyclin-dependent kinases, known as *CDKs*, which are a group of protein kinases involved in cell cycle progression, also have significant functions in the regulation of leaf growth. *CDKA;1* is associated with the control of leaf growth and responses to non-living environmental stresses, whereas *CDKB1;1* is involved in facilitating cell division and differentiation throughout the formation of organs [26].

The control of leaf growth and development is a multifaceted process that encompasses a wide range of factors, including hormones, multiple genes, and transcription factors [22–25]. Recent advances in integrating genome-wide association and transcriptome analysis have provided valuable molecular insights into heterophylly and eco-adaptability in woody plants [27]. Additionally, comparative transcriptomics studies have enhanced our understanding of the molecular basis of heterophylly in aquatic plants, revealing the significance of leaf diversity and its ecological importance [28]. Investigations into how plants and phytohormones achieve leaf phenotypic plasticity in response to environmental signals have further deepened our understanding of the intricate interplay between environmental cues, phytohormones, and regulatory pathways that govern leaf development [29]. These advancements serve as a solid foundation for future research in this field, driving further exploration into the complexities of leaf growth and development.

R. heterophylla is a unique shrub that is native to China and typically found in temperate biomes. For centuries, the leaves of this species have been used to treat health issues, such as bleeding, irregular menstruation, and dysentery, owing to its remarkable medicinal properties [30]. Interestingly, *R. heterophylla* exhibits heteromorphic leaves that alternate

between large and small sizes, making it a fascinating subject for research. Therefore, additional research is essential to thoroughly grasp the molecular processes that control leaf size regulation in this species. In our study, the utilization of SMART sequencing and Illumina RNA-seq technologies provides a comprehensive and accurate representation of the transcriptome of *R. heterophylla*, enabling a more profound level of insight into the molecular processes that contribute to the differences in leaf size and their development within this distinctive shrub.

2. Materials and Methods

2.1. PacBio Iso-Seq Library Construction, Data Processing, and Isoform Identification

In the process of constructing a reference transcriptome database, equal amounts of RNA were extracted from root, stem, and leaf tissues of *R. heterophylla*, collected from Bashan town, Shaanxi province, China (108°9′19″ E, 32°16′48″ N), and pooled to create a single sample for SMART sequencing. The voucher specimen (accession number XZY19-3098) is deposited in the Herbarium at Xi'an Botanical Garden, Shaanxi province. The cDNA was synthesized from the mRNA using the Clontech SMARTer PCR cDNA synthesis kit, and a subsequent large-scale PCR was conducted to assemble the SMRTbell library. Finally, the SMRTbell template libraries were constructed, and the SMART cells were sequenced on the PacBio Sequel platform.

Initially, the raw sequencing data was processed using the SMRT Link v6.0 software [31]. This involved categorizing the reads from the subreads BAM file into various types: full-length non-chimeric (FLNC), non-full-length (nFL), chimeras, or short reads. This categorization was based on the detection of cDNA primer sequences and poly((A)) tail markers. Short reads were excluded from further analysis. The FLNC reads were then subjected to a clustering process utilizing the iterative clustering for error correction (ICE) method, which resulted in the generation of cluster consensus isoforms [31]. To improve the accuracy of the PacBio reads, the nFL reads were used to refine these isoforms. This refinement was archived using the Quiver algorithm, which produced polished high-quality consensus sequences, with a minimum accuracy threshold of 99%. In addition to the Pacific Biosciences data, RNA-seq short reads from the leaves were employed to further correct the sequencing errors in the consensus transcripts. This correction was performed using LoRDEC [32]. Following error correction, the CD-HIT v4.6.7 software was used to remove redundant sequences, setting a similarity threshold of 0.99 to determine the final set of unique transcriptome isoforms [33]. Finally, the completeness and quality of the conserved elements within the final transcripts were assessed. This assessment was conducted using 2326 core eudicot genes from the Benchmarking Universal Single-Copy Orthologs database (BUSCO v5.4.7; [34]).

2.2. Illumina RNA-Seq Data Processing

R. heterophylla, an evergreen plant, was selected as the focus of our study on the relationship between large and small leaf size variation. To ensure a representative sample, we specifically selected nearly mature plants from this species that exhibited noticeable differences in leaf size. In September, samples were collected from these plants, and total RNA was extracted from both the large and small leaves. To ensure reliable results, we performed three biological replicates for each sample. The RNA extraction process involved homogenizing the tissue in TRIzol on a dry ice platform, following the recommended procedures provided by the manufacturer (Life Technologies, Carlsbad, CA, USA). To assess the quality and integrity of the extracted RNA, we utilized two methods: the Agilent 2100 Bioanalyzer and agarose gel electrophoresis. The Agilent 2100 Bioanalyzer is a capillary electrophoresis-based system that provides detailed information about the RNA sample, including the RNA integrity number (RIN). Electrophoresis on agarose gel allowed visualization of the distinct bands representing the different RNA species and provided additional confirmation of RNA integrity. The purity and concentration of the RNA were determined using a NanoDrop micro-spectrophotometer from Thermo Fisher (Waltham,

MA, USA). This instrument measures the absorbance of the RNA sample at both 260 nm and 280 nm wavelengths. The 260/280 nm absorbance ratio was calculated to determine the purity of the RNA, with values typically ranging between 1.8 and 2.0 considered indicative of high-quality RNA. Furthermore, the NanoDrop micro-spectrophotometer enabled us to determine the concentration of the extracted RNA using established calibration curves. We then processed the raw sequence data from the Illumina HiSeqTM 4000 system, applying stringent filters to exclude: (i) sequences that incorporated adapter contamination, (ii) sequences with over 10% of unidentified nucleotides (N), and (iii) sequences where the proportion of low quality bases (Q-value $\leq$ 20) exceeded 40%.

2.3. Gene Function Annotation and Gene Structure Analysis

The isoforms were subjected to BLAST analysis against various databases, including the NCBI non-redundant protein (Nr) database, the Swiss-Prot protein database, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and the COG/KOG database. We utilized the BLASTx tool available from the NCBI's BLAST website for the purpose of aligning our isoform sequences against the Nr database. For each isoform, the sequence that exhibited the minimum E-value was chosen to represent the analogous homolog. Subsequently, we identified the species associated with these homologous sequences and quantified the homologs attributable to each species.

To analyze the Gene Ontology (GO) annotation, we utilized the Blast2GO v6.0 software, which was run with the Nr annotation results on the isoforms [35], using an E-value filter of 1×10^{-5}. Subsequently, using the WEGO v2.0 software, we categorized the isoforms based on their GO functionalities. In order to classify the transcription factor (TF) families, we aligned the isoform-encoded protein sequences with the PlantTFDB v4.0 database using hmmscan [36].

Microsatellite mining in the entire transcriptome was performed using the MIcroSAtellite (MISA) tool. For the categorization of transcript isoforms into gene families and the subsequent assembly of a coding reference genome utilizing k-mer-based clustering and de Bruijn graph techniques, we utilized the Cogent v8.0 software. The assessment of alternative splicing across the transcript isoforms was executed using the SUPPA tool [37].

2.4. Differentially Expressed Genes (DEGs) and GO and KEGG Enrichment Analysis

We quantified the transcription levels of all the unigenes in both large and small leaves (with three replicates for each sample, totaling six replicates) by leveraging Illumina RNA-seq short reads, employing the comprehensive transcripts obtained from PacBio Iso-seq as references. The quantification of the gene expression was performed through the FPKM (fragments per kilobase of transcript per million mapped reads) approach, and we conducted the subsequent analyses employing the edgeR v3.3 software. The DEGs were pinpointed by setting a threshold of a log2 fold change of 2 or greater and a stringent FDR (false discovery rate) of less than 0.05 [38]. Following their identification, these DEGs were subjected to GO and KEGG pathway enrichment analyses to decipher their likely biological roles and associated molecular pathways [39].

3. Results

3.1. Full-Length Transcripts from PacBio Isoform Sequencing

The box and scatter plots depict the distribution of the leaf length for large and small leaves on a branch. The analysis reveals a significant difference in the leaf length between the two categories, with the large leaves averaging around 2.7 cm and the small leaves averaging around 1.2 cm (Figure 1b,c). We combined equal quantities of RNA extracted from the roots, stems, and leaves to create a composite sample for SMART sequencing, with the aim of compiling a comprehensive reference transcriptome database. After filtering (Full pass $\geq$ 3), we secured a total of 344,392 CCS. Out of these, 311,928 sequences (accounting for 90.57%) were recognized as FLNC, with an average length of 1802 base pairs (Figure 1b). Subsequent to the elimination of redundant sequences within the FLNC pool, we obtained

148,394 unpolished consensus isoforms. Further, by aligning the nFL reads to these raw isoforms, we distinguished 126,309 high quality and 21,592 lower quality isoforms. For the low-quality isoforms, we used Illumina RNA-seq data for the correction. This process yielded 141,286 polished isoforms, exhibiting an N50 value of 2166 bp. In the final phase, we applied CD-HIT v4.6.7 to remove any isoforms that were redundant or displayed high similarity, culminating in a collection of 88,546 distinct isoforms, with an improved N50 value of 2386 bp (Figure 1c). Based on a BUSCO search on the eudicots_odb10 dataset (2326 genes), the transcriptome is estimated to be 86.9% complete. Among the complete set of BUSCO genes, 67.4% were duplicated and 19.5% were single-copy BUSCOs (Table 1).
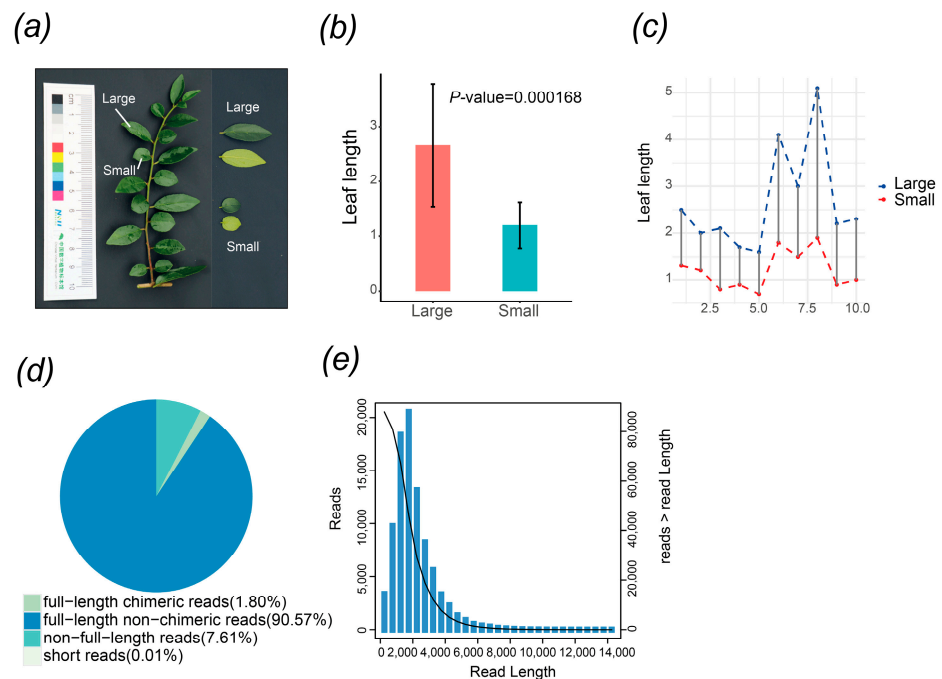

Figure 1. Leaf characteristics of *R. heterophylla* and transcriptome information. (**a**) The entire branch of *R. heterophylla* exhibits an alternating distribution of leaf sizes; (**b**) the box plot illustrates the distribution of leaf length for large and small leaves on a branch; (**c**) the scatter plot displays the distribution of leaf length for large and small leaves on a branch; (**d**) classification of the reads; (**e**) abundant isoforms.

Table 1. Transcriptome assembly completeness assessment on *R. heterophylla*, using BUSCO with the eudicots_odb10 dataset.

Metrics	Number
Complete BUSCOs (C)	2020 (86.9%)
Complete and single-copy BUSCOs (S)	453 (19.5%)
Complete and duplicated BUSCOs (D)	1567 (67.4%)
Fragmented BUSCOs (F)	34 (1.5%)
Missing BUSCOs (M)	272 (11.6%)
Total BUSCO groups searched	2326

3.2. Functional Annotation and Classification

Functional annotation was conducted utilizing multiple public databases, including NR, KOG, KEGG, GO, Swiss-Prot, and BLASTx. Of all the isoforms examined, a significant proportion, amounting to 83,725 or 94.56%, were effectively characterized with the aid of the referenced databases. The tally of isoforms that received annotations exhibited variation when considered across the different databases, ranging from 40,469 (45.70%, KEGG) to 83,187 (93.95%, NR). Nevertheless, a subset of 4821 isoforms (5.445%) remained

unannotated, suggesting the presence of novel or uncharacterized genes within the dataset (Table 2; Figure 2a). After aligning with the NR database, the annotation process enables the assessment of gene sequence similarity between *R. heterophylla* and closely related species. The highest similarity in gene sequences was observed between *R. heterophylla* and *Ziziphus jujuba*, with a total of 43,379 similar isoforms identified (52.17% of the aligned isoforms). This was followed by *Prunus mume* with 3288 isoforms (3.95%) and *Morus notabilis* with 3140 isoforms (3.77%) (Figure 2b).

Table 2. Summary table of annotation statistics in four databases.

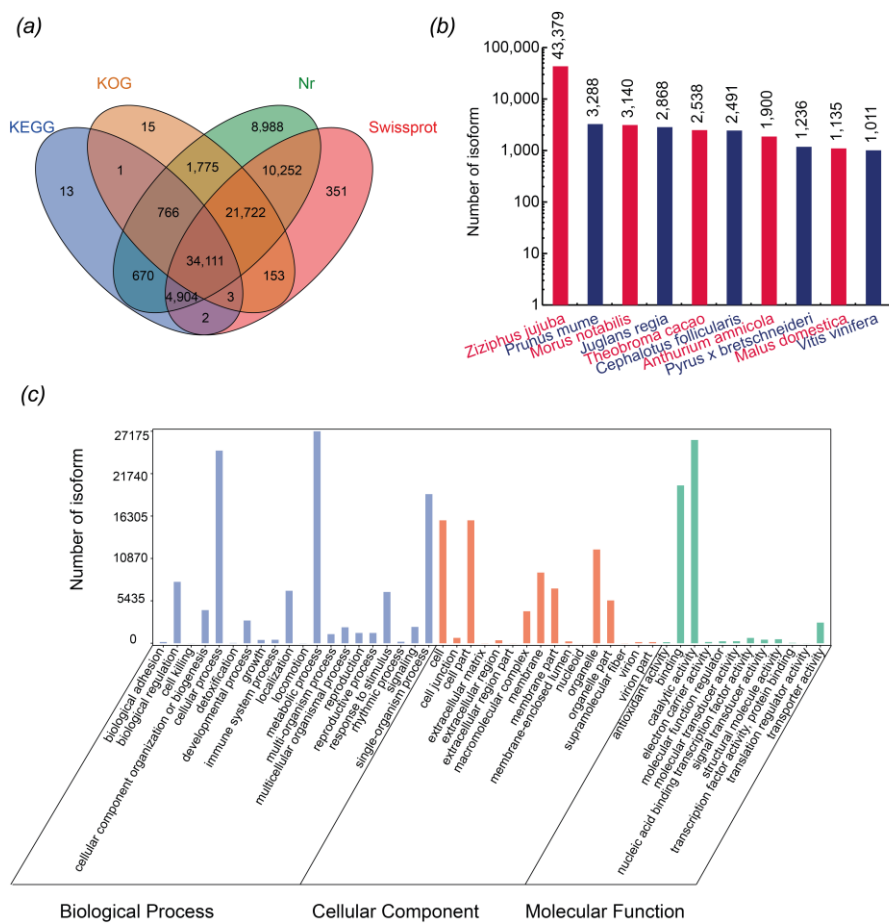| Total Isoform	NR	Swiss-Prot	KOG	KEGG	Annotation Gene	Without Annotation Gene Number
88,546	83,187 (93.95%)	71,497 (80.75%)	58,545 (66.12%)	40,469 (45.70%)	83,725 (94.56%)	4821 (5.44%)

Figure 2. Functional annotation and classification. (**a**) Venn plot showcasing functional annotations from 4 public databases; (**b**) statistics depicting the number of homologous sequences aligned to each species; (**c**) GO annotation classification for 31,187 genes.

By utilizing the annotations from NR, we employed the Blast2GO v6.0 software to acquire the GO annotation information for the isoforms. Among the GO annotations, a total of 83,725 isoforms were distributed into three distinct groups: biological processes, cellular components, and molecular functions (Figure 2b). Within the biological process category, the top three terms by abundance included 'metabolic process' (27,171), 'cellular process' (24,689), and 'single-organism process' (19,113) transcripts. For the category pertaining to cellular components, the terms with the highest representation were 'cell' (15,756), 'cell

part' (15,755), and 'organelle' (12,016). Regarding molecular functions, the classifications with the most transcripts were 'catalytic activity' (26,058) and 'binding' (20,231) (Figure 2c).

3.3. Gene Structure Prediction

We identified five different types of SSRs. Repeats consisting of di-nucleotides (15,231; 53.8%) were the predominant type observed, followed by tri-nucleotides (9707; 33.8%), tetra-nucleotides (1841; 3.6%), penta-nucleotides (589; 0.7%), and hexa-nucleotides (889; 1%) (Figure 3a,b). Alternative splicing (AS) increases the complexity of gene expression and protein diversity, and has a significant function in the growth and progression of plants. We identified 14,947 unigenes that underwent AS events, of which 20.65% experienced two AS events. The most abundant AS events were the retention of intron (RI) (Figure 3c,d).
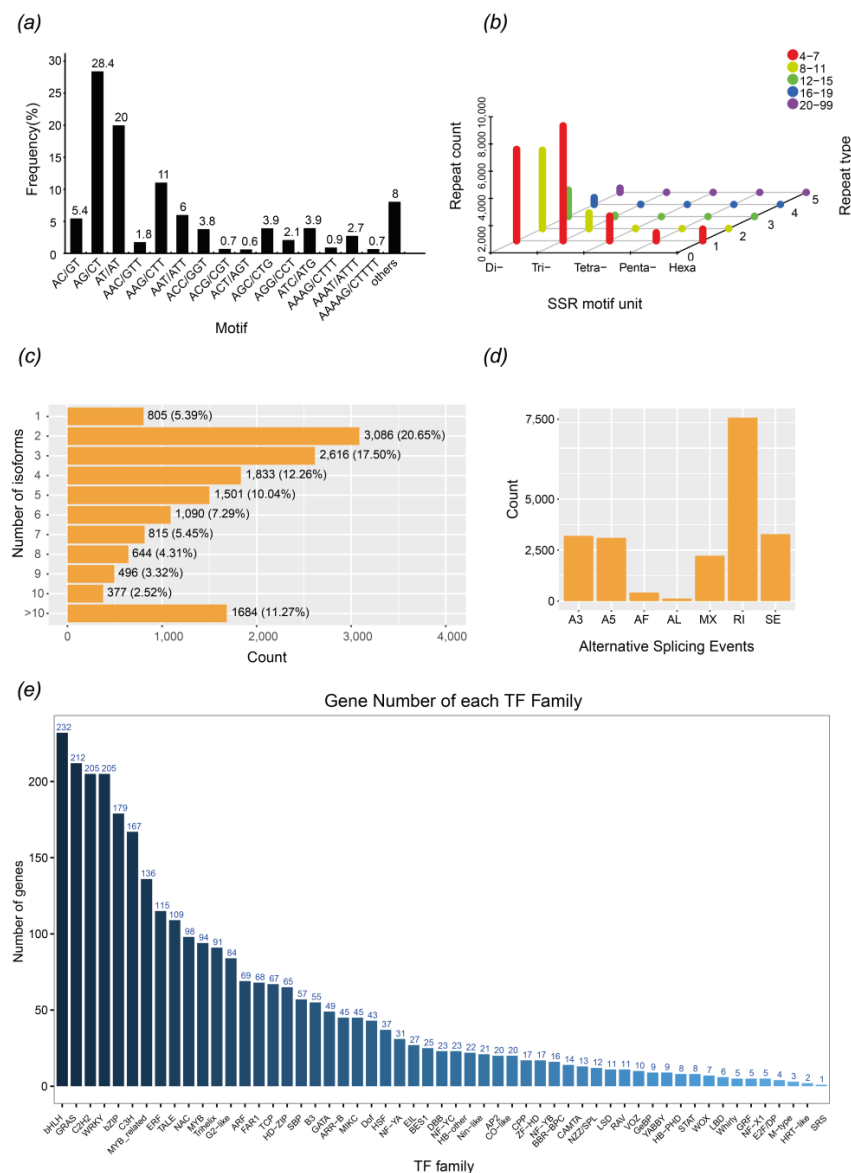

Figure 3. Gene structure prediction. (**a**) Proportional distribution of different types of tandem repeat units within the total SSR. (**b**) Distribution of SSR motifs in unigenes. (**c**) Statistics on isoforms. (**d**) Statistics on alternative splicing events. (**e**) Statistical analysis of the number of transcription factors in each family. A3: Alternative 3′ Splice Site, A5: Alternative 5′ Splice Site, AF: Alternative First Exon, AL: Alternative Last Exon, MX: Mutually Exclusive Exon, RI: Retained Intron, SE: Skipping Exon.

During the growth of leaves, transcription factors (TFs) are crucial in modulating the expression of genes. Identifying key TFs will aid in deepening our knowledge of the core regulatory processes involved. We classified all 2932 TFs into 55 gene families, with the top ten being bHLH (7.91%), GRAS (7.23%), C2H2 (6.99%), WRKY (6.99%), bZIP (6.11%), C3H (5.70%), MYB (4.64%), ERF (3.92%), TALE (3.72%), and NAC (3.34%) (Figure 3e).

3.4. Categorization and Examination of Genes Exhibiting Differential Expression

The Illumina HiSeq sequencing system (second-generation sequencing) was employed to obtain sequencing data for both large and small leaves of *R. heterophylla*, with three replicates for each sample. The transcriptomic information from these six replicates was then aligned to the reference transcriptome. The alignment of the sequencing reads to the reference sequences yielded proportions of 93.41%, 93.55%, 93.50%, 92.91%, 93.37%, and 90.79% for the respective samples (Table 3).

Table 3. Alignment of reads to the reference sequence.

Sample	All Reads Num	Unmapped Reads	Unique Mapped Reads	Multiple Mapped Reads	Mapping Ratio
Large1	46,664,928	3,073,767	3,131,403	40,459,758	93.41%
Large2	35,669,042	2,302,264	2,047,345	31,319,433	93.55%
Large3	40,439,690	2,627,144	2,626,124	35,186,422	93.50%
Small1	39,492,624	2,800,086	2,687,340	34,005,198	92.91%
Small2	37,522,320	2,487,999	2,381,952	32,652,369	93.37%
Small3	42,800,590	3,943,706	2,847,386	36,009,498	90.79%

In exploring the variations in the transcript expression associated with the leaf size in *R. heterophylla*, we aligned the RNA-seq data from large and small leaves, obtained through Illumina sequencing, with the comprehensive transcripts within the Iso-Seq database. We obtained a total of 982 DEGs, with 116 showing a notable increase in regulation, while 866 demonstrated a significant decrease (Figure 4).
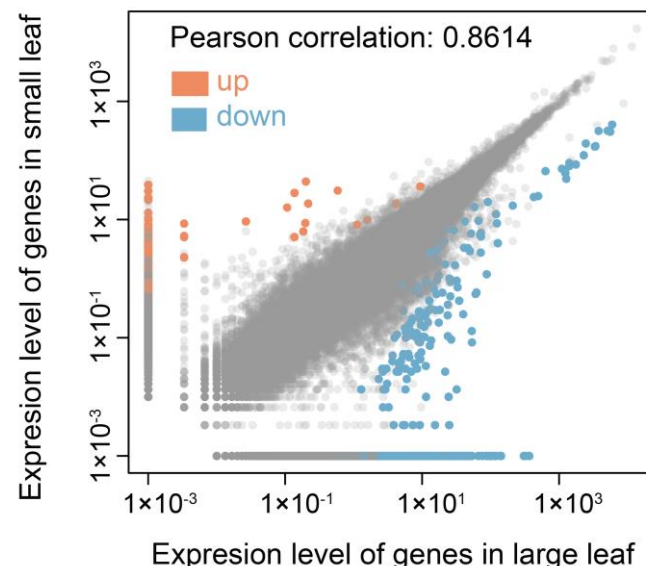

Figure 4. Scatter plot illustrating the DEGs between the large leaf and small leaf groups.

We determined the biological function of the DEGs between small and large leaves through functional enrichment analysis. The GO enrichment analysis classified the 982 DEGs into 42 categories (Figure 5a). Among them, 20 categories were involved in biological processes, 12 in cellular components, and 10 in molecular functions. Among the biological process categories, the most prevalent term was the metabolic process, with 48 genes up-regulated

and 365 genes down-regulated. In the cellular component category, the most abundant term was the cell part, comprising 18 up-regulated genes and 169 down-regulated genes. Within the molecular function category, the majority of DEGs were classified into catalytic activity, with 51 genes up-regulated and 327 genes down-regulated (Figure 5a). It is important to note that due to the multiplicity of GO annotations for individual genes, the number of up-regulated and down-regulated genes listed within each category reflects their occurrence in those specific categories and not the total number of unique DEGs.
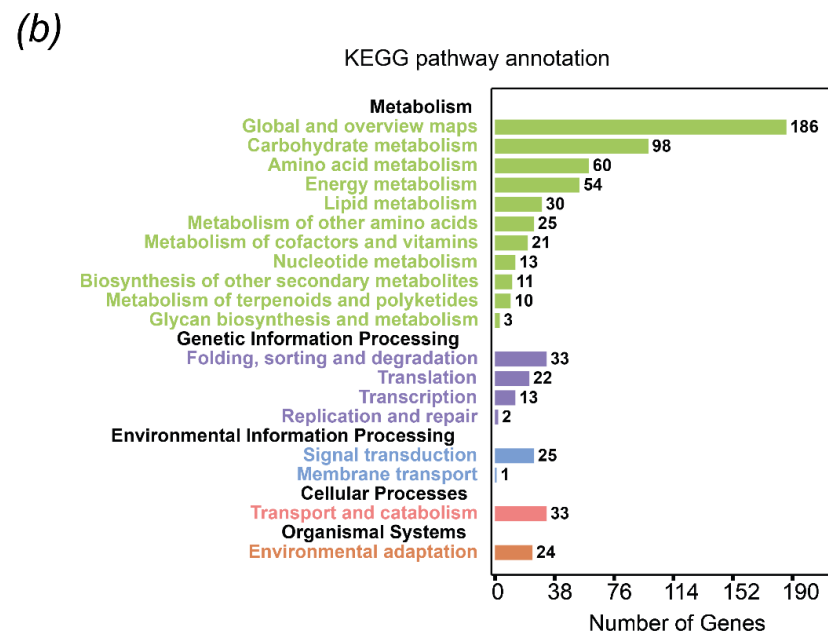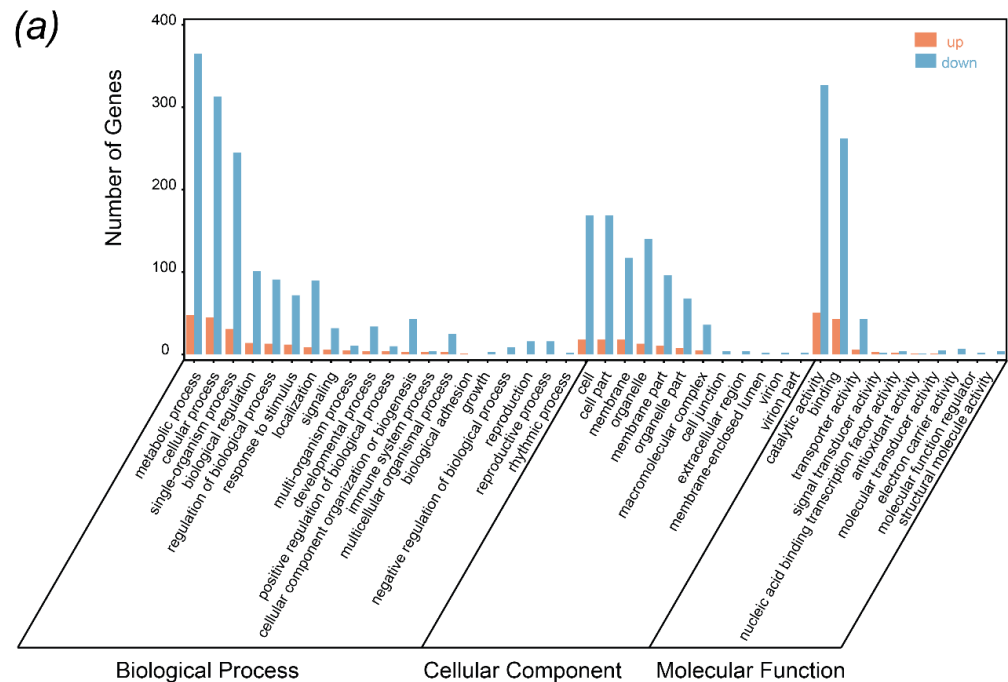

Figure 5. Go functional (**a**) and KEGG (**b**) pathway classification of unigenes differentially expressed between large and small leaves of *R. heterophylla*.

Through KEGG classification analysis, we determined the enriched pathways among the DEGs associated with leaf size variation in *R. heterophylla*. Notably, the top four pathways with the highest number of genes were global and overview maps (186), carbohydrate metabolism (98), amino acid metabolism (60), and energy metabolism (54) (Figure 5b). Metabolic pathways, including carbohydrate metabolism, amino acid metabolism, and energy metabolism, are known to play crucial roles in plant growth and development. Additionally, our analysis identified 25 DEGs that were enriched in the signal transduction pathway. The identification of these enriched pathways, among the DEGs, offers valuable insights into the underlying molecular mechanisms governing leaf size in *R. heterophylla*.

3.5. Genes Associated with the Transduction of Plant Hormone Signals

The transduction of plant hormone signals is vital in the process of leaf maturation. Within the scope of our examination of the 982 DEGs, a down-regulation was noted in numerous genes that participate in these hormone signaling cascades, shedding light on their relevance to leaf size regulation in *R. heterophylla*. Notably, genes such as IAA27, ARF2/6/9/19 (tryptophan metabolism), EBF1, and EIN2 (cysteine and methionine metabolism), ARR1 (zeatin biosynthesis), HAB1 and ABF2 (carotenoid biosynthesis), and *MYC2* (a key regulator of jasmonic acid signaling) were found to be down-regulated (Table 4). Furthermore, our study revealed the down-regulation of *TCP8* genes, known regulators of leaf or flower shape in *Arabidopsis thaliana* [40], in small leaves of *R. heterophylla* (Supporting File S1). Moreover, we observed the down-regulation of the *CDKE-1* gene, crucial for cell expansion, in *R. heterophylla* (Supporting File S1).

Table 4. List of DEGs involved in several hormone signaling pathways in *R. heterophylla*.

Gene	$\log^2$ Ratio (Small/Large Leaf)	*p* Value	FDR	Regulation	Function
BSK1	9.98	1.86×10^{-4}	1.91×10^{-2}	Up	BR-signaling kinases
EIL3	−11.17	3.69×10^{-4}	3.16×10^{-2}	Down	Regulators of ethylene signaling
HAB1	−11.86	1.40×10^{-4}	1.50×10^{-2}	Down	Regulates numerous ABA responses
EBF1	−13.42	3.34×10^{-8}	2.02×10^{-5}	Down	EIN3-binding F-box protein, ethylene signaling
EBF1	−14.28	4.95×10^{-7}	1.87×10^{-4}	Down	EIN3-binding F-box protein, ethylene signaling
TIFY6B	−6.59	1.47×10^{-6}	4.77×10^{-4}	Down	Repressor of jasmonate responses
ABF2	−12.10	2.25×10^{-4}	2.19×10^{-2}	Down	Regulates ABA-responsive gene expression
ARF9	−6.64	2.11×10^{-6}	6.23×10^{-4}	Down	Auxin response factor
IAA27	−10.18	5.84×10^{-4}	4.38×10^{-2}	Down	Auxin-responsive protein
ARR1	−11.75	4.60×10^{-4}	3.67×10^{-2}	Down	Cytokinin response regulator
ARF19	−11.70	1.36×10^{-4}	1.46×10^{-2}	Down	Auxin response factor
HAB1	−12.77	3.77×10^{-6}	9.95×10^{-4}	Down	ABA signaling pathway
MYC2	−13.83	2.67×10^{-7}	1.09×10^{-4}	Down	Regulators of JA signaling
EIN2	−11.80	5.67×10^{-5}	7.61×10^{-3}	Down	Ethylene-insensitive protein
ARF6	−10.59	6.15×10^{-4}	4.54×10^{-2}	Down	Auxin response factor
GRF8	−14.33	4.13×10^{-6}	1.04×10^{-3}	Down	Growth-regulating factor
ARF2	−13.08	3.80×10^{-6}	9.97×10^{-4}	Down	Auxin response factor
ARF6	−13.13	2.17×10^{-5}	3.75×10^{-3}	Down	Auxin response factor

4. Discussion

Plants depend on leaves as vital nutritional organs for photosynthesis and respiration, and the dynamic adjustment of leaf size serves as a crucial adaptive response by plants to changes in the external environment. It is fascinating that *R. heterophylla* presents an intriguing botanical phenomenon, characterized by the concurrent presence of differently sized leaves, often paired together along the same stem. This shrub, which thrives in the forested slopes of mountainous regions, demonstrates a potential evolutionary strategy for resource optimization. Its modest stature, often overshadowed by the canopy, may have driven the evolution of the variance in its leaf size, as a mechanism to maximize photosynthetic efficiency. Furthermore, the pairing of large and small leaves could serve as a defensive

adaptation, creating a confusing visual pattern that may deter herbivores or camouflage the plant against pests that rely on visual cues for host identification. However, the absence of the genetic background of *R. heterophylla* limits research on its leaf development and adaptation. To address this limitation, our project focused on constructing the full-length transcriptome of *R. heterophylla* using PacBio long-read isoform transcriptome sequencing. Additionally, we employed Illumina paired-end short-read RNA sequencing to identify key genes and pathways associated with leaf size heterophylly.

In recent years, extensive research has been conducted to investigate the regulatory mechanisms underlying heterophylly, a phenomenon characterized by the development of leaves with distinct shapes or sizes on the same plant [41–43]. For example, a study on *Ranunculus trichophyllus* found that heterophylly, the production of different leaf shapes under different growing conditions, is controlled by two plant hormones, *abscisic acid*, and ethylene. Aquatic leaves are characterized by a narrow shape, lack of stomata, and reduced vessel development due to *EIN3*-mediated overactivation of abaxial genes, while terrestrial leaves have established leaf polarity, and stomata and vessel development, due to *ABI3*-mediated activation of adaxial genes [43]. Similarly, a study on *Rorippa aquatica* suggests that the regulation of the gibberellin level via *KNOX1* genes is involved in regulating heterophylly [44]. These studies have advanced our understanding of how plants regulate the development of diverse leaf shapes.

In this study, we sought to expand on this knowledge by examining the expression levels of genes in large and small leaves in a particular plant species. Our findings revealed a total of 982 DEGs, with the majority consistently down-regulated in small leaves across multiple hormone signaling pathways. These findings suggest a crucial role of hormone signaling pathways in regulating leaf size heterophylly in this plant species. Specifically, our findings implicate the involvement of several hormone signaling pathways, including auxin, ethylene, brassinosteroid, jasmonic acid, and abscisic acid signaling pathways. These pathways are known to be involved in a wide range of plant developmental processes, such as cell division, elongation, and differentiation [45,46]. Moreover, the process of leaf size determination in *R. heterophylla* is complex and involves the interaction of multiple genes that regulate cell division and expansion. *EIN2*, *EBF1*, and *EIL1*, are key regulators of the ethylene signaling pathway, which plays a central role in controlling the final organ size by restricting cell expansion [47,48]. In our study, these regulators were significantly down-regulated in small leaves, indicating their critical role in regulating leaf size development in *R. heterophylla*. Additionally, the expression of genes involved in auxin signaling pathways, such as *ARF2, ARF6, ARF9, ARF19*, and *GRF8*, were significantly reduced in small leaves, consistent with the finding that these genes are critical regulators of leaf development through influencing cell division and expansion [22,49]. In addition, other hormone signaling pathways have also been implicated in the regulation of leaf size development.

These findings propose that multiple hormone signaling pathways and genes are involved in the regulation of leaf size development in *R. heterophylla*. However, to gain a more comprehensive understanding of the mechanisms that determine leaf size development in this plant species, it is necessary to investigate the specific interactions between these genes and hormonal pathways. Future studies should focus on exploring the mechanisms of ethylene and auxin hormone pathways in more detail, as they may play a particularly important role in leaf size regulation in *R. heterophylla*. By doing so, we can gain a deeper understanding of the molecular mechanisms that underlie leaf size development in this plant species, which could have broader implications for the field of plant biology.

5. Conclusions

The study of leaf development and adaptation in plants is of great importance for understanding plant growth and development, as well as for improving crop productivity and environmental adaptation. Using PacBio long-read isoform transcriptome sequencing and Illumina paired-end short-read RNA sequencing, we constructed the full-length transcriptome of *R. heterophylla* and identified key genes and pathways involved in leaf

size heterophylly. Our findings indicate that hormone signaling pathways play a crucial role in the regulation of leaf size. These results offer valuable insights into the molecular mechanisms underlying leaf development and adaptation in *R. heterophylla* and, potentially, other plant species. In addition to highlighting the central role of hormone signaling, our research underscores the complexity of genetic regulation in leaf size variation. The significant number of differentially expressed genes identified between large and small leaves suggests a multifaceted genetic network, where interactions between genes and environmental factors may dictate leaf morphology. This complexity points to the potential for developing targeted breeding or genetic engineering strategies to optimize leaf traits for specific agricultural or environmental needs. Furthermore, the decrease in expression of auxin-related genes in smaller leaves provides a tangible link between hormonal control and physical leaf development, suggesting potential targets for manipulation to achieve desired plant phenotypes. While further investigation is necessary to confirm these findings and elucidate the specific mechanisms by which these genes affect leaf growth, this study provides a basal genetic resource for future research in this field.

References

1. McConnell, J.R.; Barton, M.K. Leaf polarity and meristem formation in *Arabidopsis*. *Development* **1998**, *125*, 2935–2942. [CrossRef]
2. Beemster, G.T.; Fiorani, F.; Inzé, D. Cell cycle: The key to plant growth control? *Trends Plant Sci.* **2003**, *8*, 154–158. [CrossRef]
3. Barkoulas, M.; Hay, A.; Kougioumoutzi, E.; Tsiantis, M. A developmental framework for dissected leaf formation in the *Arabidopsis* relative *Cardamine hirsuta*. *Nat. Genet.* **2008**, *40*, 1136–1141. [CrossRef]
4. Waites, R.; Hudson, A. phantastica: A gene required for dorsoventrality of leaves in *Antirrhinum majus*. *Development* **1995**, *121*, 2143–2154. [CrossRef]
5. Braybrook, S.A.; Kuhlemeier, C. How a plant builds leaves. *Plant Cell* **2010**, *22*, 1006–1018. [CrossRef]
6. Bayer, E.M.; Smith, R.S.; Mandel, T.; Nakayama, N.; Sauer, M.; Prusinkiewicz, P.; Kuhlemeier, C. Integration of transport-based models for phyllotaxis and midvein formation. *Genes Dev.* **2009**, *23*, 373–384. [CrossRef]
7. Guenot, B.; Bayer, E.; Kierzkowski, D.; Smith, R.S.; Mandel, T.; Žádníková, P.; Benková, E.; Kuhlemeier, C. Pin1-independent leaf initiation in Arabidopsis. *Plant Physiol.* **2012**, *159*, 1501–1510. [CrossRef]
8. Sablowski, R.; Carnier Dornelas, M. Interplay between cell growth and cell cycle in plants. *J. Exp. Bot.* **2014**, *65*, 2703–2714. [CrossRef]
9. Blilou, I.; Xu, J.; Wildwater, M.; Willemsen, V.; Paponov, I.; Friml, J.; Heidstra, R.; Aida, M.; Palme, K.; Scheres, B. The PIN auxin efflux facilitator network controls growth and patterning in *Arabidopsis* roots. *Nature* **2005**, *433*, 39–44. [CrossRef]
10. Velasquez, S.M.; Barbez, E.; Kleine-Vehn, J.; Estevez, J.M. Auxin and Cellular Elongation. *Plant Physiol.* **2016**, *170*, 1206–1215. [CrossRef]
11. Majda, M.; Robert, S. The role of auxin in cell wall expansion. *Int. J. Mol. Sci.* **2018**, *19*, 951. [CrossRef]

12. Castro-Camba, R.; Sánchez, C.; Vidal, N.; Vielba, J.M. Plant development and crop yield: The role of gibberellins. *Plants* **2022**, *11*, 2650. [CrossRef]

13. Chen, S.; Wang, X.J.; Tan, G.F.; Zhou, W.Q.; Wang, G.L. Gibberellin and the plant growth retardant Paclobutrazol altered fruit shape and ripening in tomato. *Protoplasma* **2020**, *257*, 853–861. [CrossRef]

14. Nagai, K.; Mori, Y.; Ishikawa, S.; Furuta, T.; Gamuyao, R.; Niimi, Y.; Hobo, T.; Fukuda, M.; Kojima, M.; Takebayashi, Y.; et al. Antagonistic regulation of the gibberellic acid response during stem growth in rice. *Nature* **2020**, *584*, 109–114. [CrossRef]

15. Schaller, G.E.; Street, I.H.; Kieber, J.J. Cytokinin and the cell cycle. *Curr. Opin. Plant Biol.* **2014**, *21*, 7–15. [CrossRef]

16. Werner, T.; Schmülling, T. Cytokinin action in plant development. *Curr. Opin. Plant Biol.* **2009**, *12*, 527–538. [CrossRef]

17. Wybouw, B.; De Rybel, B. Cytokinin—A Developing Story. *Trends Plant Sci.* **2019**, *24*, 177–185. [CrossRef]

18. Wu, W.; Du, K.; Kang, X.; Wei, H. The diverse roles of cytokinins in regulating leaf development. *Hortic. Res.* **2021**, *8*, 118. [CrossRef]

19. Wei, Z.; Li, J. Brassinosteroids regulate root growth, development, and symbiosis. *Mol. Plant* **2016**, *9*, 86–100. [CrossRef]

20. Pekker, I.; Alvarez, J.P.; Eshed, Y. Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *Plant Cell* **2005**, *17*, 2899–2910. [CrossRef]

21. Koenig, D.; Bayer, E.; Kang, J.; Kuhlemeier, C.; Sinha, N. Auxin patterns solanum lycopersicum leaf morphogenesis. *Development* **2009**, *136*, 2997–3006. [CrossRef]

22. Peng, Y.; Fang, T.; Zhang, Y.; Zhang, M.; Zeng, L. Genome-wide identification and expression analysis of auxin response factor (ARF) gene family in Longan (*Dimocarpus longan* L.). *Plants* **2020**, *9*, 221. [CrossRef]

23. Horiguchi, G.; Kim, G.T.; Tsukaya, H. The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J.* **2005**, *43*, 68–78. [CrossRef]

24. Viola, I.L.; Alem, A.L.; Jure, R.M.; Gonzalez, D.H. Physiological roles and mechanisms of action of class I TCP transcription factors. *Int. J. Mol. Sci.* **2023**, *24*, 5437. [CrossRef]

25. Liu, Y.; Guan, X.; Liu, S.; Yang, M.; Ren, J.; Guo, M.; Huang, Z.; Zhang, Y. Genome-wide identification and analysis of TCP transcription factors involved in the formation of leafy head in Chinese cabbage. *Int. J. Mol. Sci.* **2018**, *19*, 847. [CrossRef]

26. Sorrell, D.A.; Menges, M.; Healy, J.M.; Deveaux, Y.; Amano, C.; Su, Y.; Nakagami, H.; Shinmyo, A.; Doonan, J.H.; Sekine, M.; et al. Cell cycle regulation of cyclin-dependent kinases in tobacco cultivar Bright Yellow-2 cells. *Plant Physiol.* **2001**, *126*, 1214–1223. [CrossRef]

27. Hu, Y.; Tang, F.; Zhang, D.; Shen, S.; Peng, X. Integrating genome-wide association and transcriptome analysis to provide molecular insights into heterophylly and eco-adaptability in woody plants. *Hortic. Res.* **2023**, *10*, uhad212. [CrossRef]

28. He, D.; Guo, P.; Gugger, P.F.; Guo, Y.; Liu, X.; Chen, J. Investigating the molecular basis for heterophylly in the aquatic plant *Potamogeton octandrus* (Potamogetonaceae) with comparative transcriptomics. *Peer J.* **2018**, *6*, e4448. [CrossRef]

29. Nakayama, H.; Sinha, N.R.; Kimura, S. How do plants and phytohormones accomplish heterophylly, leaf phenotypic plasticity, in response to environmental cues. *Front. Plant Sci.* **2017**, *8*, 1717. [CrossRef]

30. Wang, L.; Fan, S.; Wang, X.; Wang, X.; Yan, X.; Shan, D.; Xiao, W.; Ma, J.; Wang, Y.; Li, X.; et al. Physicochemical aspects and sensory profiles as various potential factors for comprehensive quality assessment of Nü-Er-Cha produced from *Rhamnus heterophylla* Oliv. *Molecules* **2019**, *24*, 3211. [CrossRef]

31. Gordon, S.P.; Tseng, E.; Salamov, A.; Zhang, J.; Meng, X.; Zhao, Z.; Kang, D.; Underwood, J.; Grigoriev, I.V.; Figueroa, M.; et al. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS ONE* **2015**, *10*, e0132628. [CrossRef]

32. Salmela, L.; Rivals, E. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* **2014**, *30*, 3506–3514. [CrossRef]

33. Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* **2010**, *26*, 680–682. [CrossRef]

34. Manni, M.; Berkeley, M.R.; Seppey, M.; Simão, F.A.; Zdobnov, E.M. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [CrossRef]

35. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [CrossRef]

36. Jin, J.; Tian, F.; Yang, D.C.; Meng, Y.Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2017**, *45*, D1040–D1045. [CrossRef]

37. Alamancos, G.P.; Pagès, A.; Trincado, J.L.; Bellora, N.; Eyras, E. Leveraging transcript quantification for fast computation of alternative splicing profiles. *Rna* **2015**, *21*, 1521–1531. [CrossRef]

38. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* **1995**, *57*, 289–300. [CrossRef]

39. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36*, D480–D484. [CrossRef]

40. Zhang, L.; Zhou, L.; Yung, W.S.; Su, W.; Huang, M. Ectopic expression of Torenia fournieri TCP8 and TCP13 alters the leaf and petal phenotypes in *Arabidopsis thaliana*. *Physiol. Plant* **2021**, *173*, 856–866. [CrossRef]

41. Koga, H.; Kojima, M.; Takebayashi, Y.; Sakakibara, H.; Tsukaya, H. Identification of the unique molecular framework of heterophylly in the amphibious plant *Callitriche palustris* L. *Plant Cell* **2021**, *33*, 3272–3292. [CrossRef]

42. Qin, S.W.; Bao, L.H.; He, Z.G.; Li, C.L.; La, H.G.; Zhao, L.F. Identification and regulatory network analysis of SPL family transcription factors in *Populus euphratica* Oliv. heteromorphic leaves. *Sci. Rep.* **2022**, *12*, 2856. [CrossRef]

43. Kim, J.; Joo, Y.; Kyung, J.; Jeon, M.; Park, J.Y.; Lee, H.G.; Chung, D.S.; Lee, E.; Lee, I. A molecular basis behind heterophylly in an amphibious plant, *Ranunculus trichophyllus*. *PLoS Genet.* **2018**, *14*, e1007208. [CrossRef]

44. Nakayama, H.; Nakayama, N.; Seiki, S.; Kojima, M.; Sakakibara, H.; Sinha, N.; Kimura, S. Regulation of the KNOX-GA gene module induces heterophyllic alteration in North American lake cress. *Plant Cell* **2014**, *26*, 4733–4748. [CrossRef]

45. Takatsuka, H.; Umeda, M. Hormonal control of cell division and elongation along differentiation trajectories in roots. *J. Exp. Bot.* **2014**, *65*, 2633–2643. [CrossRef]

46. Zluhan-Martínez, E.; López-Ruíz, B.A.; García-Gómez, M.L.; García-Ponce, B.; de la Paz Sánchez, M.; Álvarez-Buylla, E.R.; Garay-Arroyo, A. Integrative roles of phytohormones on cell proliferation, elongation and differentiation in the *Arabidopsis thaliana* primary root. *Front. Plant Sci.* **2021**, *12*, 659155. [CrossRef]

47. Merchante, C.; Brumos, J.; Yun, J.; Hu, Q.; Spencer, K.R.; Enríquez, P.; Binder, B.M.; Heber, S.; Stepanova, A.N.; Alonso, J.M. Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. *Cell* **2015**, *163*, 684–697. [CrossRef]

48. Feng, G.; Liu, G.; Xiao, J. The *Arabidopsis* EIN2 restricts organ growth by retarding cell expansion. *Plant Signal. Behav.* **2015**, *10*, e1017169. [CrossRef]

49. Kalve, S.; De Vos, D.; Beemster, G.T. Leaf development: A cellular perspective. *Front. Plant Sci.* **2014**, *5*, 362. [CrossRef]