Biobreeding

SCIEPublish

Article

PacBio Iso-Seq and RNA-Seq Reveal Gene Expression Divergence in Early-Flowering Gynoecious and Androecious Ramie Mutants

Hongyue Qu [1,2,3,†], Shuo Zhang [1,†], Jue Zhao [1], Qinxu Song [1,2] and Hucheng Xing [1,2,4,*]

[1] Ramie Research Institute, Hunan Agricultural University, Changsha 410128, China; sboping@stu.hunau.edu.cn (H.Q.); zs15145819770@stu.hunau.edu.cn (S.Z.); zhaojue66@stu.hunau.edu.cn (J.Z.); songqinxu@stu.hunau.edu.cn (Q.S.)

[2] Yuelushan Laboratory, Changsha 410128, China

[3] Xiangxi Tujia and Miao Autonomous Prefecture Academy of Agricultural Sciences, Jishou 416000, China

[4] Hunan Key Laboratory of Germplasm Resources Innovation and Resource Utilization Crop Breeding Center, Changsha 410128, China

* Corresponding author. E-mail: xinghucheng@hunau.edu.cn (H.X.)

† These authors contributed equally to this work.

Received: 11 June 2026; Revised: 30 July 2026; Accepted: 24 August 2026; Available online: 9 September 2026

ABSTRACT: Ramie (*Boehmeria nivea* L.) is an economically valuable bast fiber crop, yet a high-quality full-length transcriptome reference for sex-divergent lines remains lacking. To fill this gap, we combined PacBio Iso-Seq with Illumina RNA-Seq to construct a full-length transcriptome for the gynoecious mutant GBN09 and compared transcriptomes of four tissues (leaves, bast fibers, stems, and flowers) between GBN09 and the androecious mutant GBN10. The analysis generated 63,082 high-quality isoforms and annotated 30,885 genes. Regulatory element mining revealed 565 transcription factors (predominantly IAA, AP2/ERF, and WRKY families), 16,617 long non-coding RNAs (lncRNAs), and 31,759 simple sequence repeats (SSRs). Transcriptomic comparisons identified 8414 differentially expressed genes (DEGs), with the most profound transcriptional shifts in floral and young stem tissues. Enrichment analysis linked these DEGs primarily to linoleic acid metabolism, phytohormone signaling cascades, and secondary metabolite biosynthesis, highlighting their roles in sexual differentiation. Furthermore, we detected 274 sex-biased differential alternative splicing (DAS) events, many of which mapped to cell wall biogenesis and core auxin signaling components such as AFB2 and NPY2. These results indicate that post-transcriptional modifications provide an additional regulatory layer during ramie floral sex differentiation. Validated by qRT-PCR, this dataset establishes a robust molecular resource and identifies candidate genes to elucidate sex differentiation and support molecular breeding in ramie.

Keywords: Ramie; Sex differentiation; PacBio Iso-Seq; RNA-Seq

1. Introduction

Boehmeria nivea (L.) Gaudich., widely known as ramie, belongs to the Urticaceae family as a perennial herbaceous species. Representing China's second most extensively cultivated fiber crop after cotton, it

possesses significant agricultural and economic importance [1]. Historically prioritized for bast fiber extraction, ramie is highly prized by the textile industry for its superior fiber length, exceptional tensile strength, and characteristic natural luster [2–4]. Recently, however, breeding initiatives have shifted toward a multi-purpose utilization model. Due to its rapid vegetative growth and substantial protein accumulation, ramie provides a viable alternative for livestock forage. Furthermore, the species demonstrates remarkable efficacy in phytoremediation, specifically in rehabilitating soils contaminated by heavy metals [2,5,6]. Pharmacologically, ramie tissues—particularly roots and leaves—synthesize bioactive secondary metabolites, notably chlorogenic acid and specific flavonoids, which have historically been used for antipyretic and detoxification purposes [7]. To fully capitalize on these multifaceted agronomic traits, a comprehensive elucidation of the underlying genetic mechanisms is imperative. Advancing our molecular understanding will directly accelerate germplasm exploitation and support sustainable cultivation strategies [8].

While the economic significance of ramie is well established, the genetic architectures governing its fundamental agronomic traits—particularly floral organogenesis and sexual divergence—remain largely obscure. Although recent assemblies of draft genomes have provided preliminary reference maps, the profound genetic heterozygosity inherent to this species severely limits the utility of relying on any single reference genome. This limitation becomes exceptionally acute when attempting to map the regulatory networks of spontaneous mutants exhibiting atypical reproductive phenotypes. Transcriptomic profiling circumvents these genomic blind spots. Traditional short-read platforms (e.g., Illumina) reliably deliver cost-effective, high-throughput quantitative expression metrics [9]; yet, they inherently struggle with complex transcript reconstruction. Short reads frequently collapse or misassemble structural variants. Third-generation single-molecule sequencing, notably the PacBio platform, resolves this bottleneck by capturing full-length mRNA transcripts without fragmentation. Bypassing the computational assembly step allows for the direct and precise identification of alternative polyadenylation (APA) sites, transcript fusions, and intricate alternative splicing (AS) architectures—critical post-transcriptional modifications that are notoriously difficult to resolve using short-read datasets alone [10–12]. Therefore, deploying a hybrid analytical pipeline that merges Illumina's sequencing depth with PacBio's structural fidelity has emerged as the most robust strategy for decoding the complex transcriptomes of highly heterozygous non-model crops and their corresponding sex-divergent mutants.

To avoid conceptual confusion, we use the following definitions throughout this study: sex determination refers to the genetic and developmental decision that establishes male or female floral fate; sex differentiation refers to the downstream developmental processes that produce morphologically distinct male or female floral organs; and sex expression refers to the observable phenotypic manifestation of sex at the whole-plant or floral level. Clarifying these processes is essential for ramie breeding because floral sex directly influences fiber yield and breeding strategies.

Up to now, molecular research in this crop has largely concentrated on bast fiber formation [13], plant responses to environmental stresses [14–19], and nutrient uptake [20]. By contrast, very little information is available regarding the genetic mechanisms that control flower formation and sex expression, even though these reproductive processes are critical for seed yield and practical breeding programs. Under natural conditions along the Yangtze River, typical monoecious ramie plants bloom in autumn. To explore this reproductive biology, our current work took advantage of two naturally occurring mutant lines: a purely female (gynoecious) line named GBN09, and its closely related all-male (androecious) derivative, GBN10. A defining phenotypic hallmark of these two specific mutant lines is their radically altered anthesis timeline. Deviating from the typical September flowering period characteristic of wild-type ramie, these accessions exhibit an extreme temporal shift, initiating reproductive development as early as May. This distinct precocious-flowering phenotype establishes them as invaluable natural models for deciphering the synergistic regulation of floral transition and sexual dimorphism. To elucidate the underlying molecular architecture driving these traits, we implemented an integrated sequencing strategy that merged PacBio

full-length transcriptomics with Illumina short-read depth. We profiled a spatially resolved panel encompassing four primary tissues—leaves, main stems, bast fibers, and inflorescences—harvested from both sexually divergent morphs. The fundamental objective of this investigation was to systematically identify pivotal candidate loci and dissect the core metabolic circuitries governing floral organogenesis and sex determination. Ultimately, this comprehensive full-length reference framework will supply precise genetic targets, significantly accelerating future marker-assisted breeding and directed trait enhancement in this species.

2. Materials and Methods

2.1. Plant Material and Sampling

The experimental germplasm comprised two distinct *Boehmeria nivea* (L.) Gaudich. lines: a gynoecious (all-female) accession designated GBN09, and its androecious (all-male) counterpart, GBN10. Because GBN10 originated as a spontaneous adventitious shoot arising directly from a GBN09 parent plant, these two morphs share a clonally related genetic background. Both lines were cultivated at the Yunyuan Experimental Station of Hunan Agricultural University, beginning in 2017. Field propagation was maintained utilizing a standard 50-cm inter-row spacing under conventional agricultural regimens, including regular irrigation, fertilization, and weed management.

For the construction of the PacBio SMRT full-length transcriptome library, tissue harvesting was conducted on 20 September 2019, during the active anthesis stage. An equal-weight pooling strategy was employed, combining distinct tissues (leaves, main stems, inflorescences, and epidermal peels) excised from six robust GBN09 individuals to generate a comprehensive bulk sample. Subsequently, for Illumina-based comparative expression profiling, four specific tissues—leaves, inflorescences, young apical stems, and middle bast fibers—were systematically harvested from both GBN09 and GBN10 lines. The “middle bast fiber” specimens were specifically defined as the decorticated bark layers stripped from a 20-cm longitudinal section encompassing the physical midpoint of the main stem. All sampling procedures were performed strictly in triplicate to provide independent biological replicates. Immediately upon excision, all tissues were instantaneously quenched in liquid nitrogen to inhibit endogenous RNase activity and subsequently archived at −80 °C pending total RNA extraction.

2.2. PacBio Iso-Seq Library Preparation and Sequencing

Ramie tissues naturally contain high levels of polysaccharides and polyphenols. To address this, total RNA was extracted using the RNAprep Pure Plant Kit (Takara, Dalian, China). RNA integrity was strictly evaluated. Only samples achieving an RNA Integrity Number (RIN) ≥ 8.0 on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) proceeded to downstream steps. Subsequently, full-length cDNA was synthesized from 2 μg of qualified total RNA with the SMARTer PCR cDNA Synthesis Kit manufactured by Takara. PCR cycles were carefully optimized during this step to prevent downstream biases caused by over-amplification.

The resulting cDNA required length-based fractionation. We applied the BluePippin protocol (Sage Science, Beverly, MA, USA) to sort the molecules into three distinct size groups: 1–2, 2–3, and 3–6 kilobases. Using 1 to 2 μg of these size-selected cDNA fractions, we constructed the SMRTbell libraries. This procedure strictly followed the Pacific Biosciences DNA Template Prep Kit 2.0 instructions. The annealing and polymerase association steps utilized the P4 Binding Kit and v3 primers. Finally, the prepared libraries were loaded onto a PacBio Sequel sequencer. Sequencing was performed using C3 chemistry with a movie collection time of 120 min [21].

Raw bioinformatic outputs required multiple filtering steps. Initially, reads shorter than 50 bp or showing an accuracy score below 0.75 were discarded. Because SMRT long reads inherently carry a certain

error margin, we integrated our high-quality Illumina short reads to correct these alignments using the LSC 1. alpha algorithm [22]. Next, reads of insert (ROIs) were classified by identifying both 5′ and 3′ adapters. Full-length non-chimeric (FLNC) sequences containing both adapters were then processed using the Iterative Clustering for Error Correction (ICE) tool. This step generated consensus transcripts exceeding 99% accuracy. To eliminate sequence redundancy, the CD-HIT program was applied, yielding the final non-redundant isoform dataset [23].

Parallel to the PacBio processing, the Illumina short reads were assembled *de novo*. We utilized Trinity (v2.0.6) and set the minimum contig length threshold to 150 bp. The resulting contigs were then clustered using the TGICL tool from the Trinity package, which automatically filtered out sequences shorter than 200 bp. Finally, the remaining reads were merged and dereplicated to produce the definitive short-read consensus dataset.

2.3. Functional Annotation and Structural Analysis

To determine the biological roles of the assembled transcripts, we performed homology alignments against several major databases using BLAST algorithms [24]. Rather than utilizing a single reference, we searched our data across the Swiss-Prot repository, the Non-redundant protein (Nr) and Nucleotide (Nt) archives, as well as the Gene Ontology (GO), Eukaryotic Orthologous Groups (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

For regulatory gene analysis, the iTAK software (version 1.7a) was applied to screen and categorize potential plant transcription factors (TFs) [25]. To discover long non-coding RNAs (lncRNAs), we designed a stringent four-step filtering pipeline. Specifically, we evaluated the protein-coding capacity of the transcripts using CNCI (v2) [26], CPC (v0.9-r2, setting the e-value threshold to 1×10^{-10}) [27], PfamScan (with an e-value limit of 0.001), PfamScan (e-value 0.001) [28], and PLEK (v1.2, restricting the minimum sequence length to 200 bp) [29]. Only those sequences universally predicted by all four algorithms as lacking coding potential were preserved as true lncRNA candidates.

Furthermore, to investigate transcriptomic structural features, we mined simple sequence repeats (SSRs) by running the MISA script (v1.0) under its standard default parameters [30]. Finally, complex AS patterns were computationally mapped and classified relying on the AStalavista platform (version 4.0) [31,32].

2.4. Preparation and Sequencing of Short-Read Libraries

Short-read sequencing libraries were prepared using the NEBNext Ultra RNA Library Prep Kit alongside matching Multiplex Oligos (New England Biolabs, Ipswich, MA, USA). Initially, mRNA was isolated and subjected to chemical fragmentation. The target fragment size was strictly controlled at 250 to 300 base pairs.

These cleaved RNA segments served as templates for first-strand cDNA synthesis. Second-strand cDNA was subsequently generated. The resulting double-stranded cDNA underwent end-repair and blunting. A single ‘A’ nucleotide was then added to the 3′ ends. This modification facilitated the specific ligation of sequencing adapters.

Unreacted components were removed using AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA). This clean-up step also served to size-select the target DNA fragments. The selected fragments were then enriched through a standard PCR amplification step. Finally, the purified libraries were loaded onto the Illumina HiSeq X platform to generate high-throughput raw sequencing reads.

2.5. Identification and Functional Profiling of Differentially Expressed Genes

The *de novo* PacBio Iso-Seq assembly served as the reference transcriptome. Transcript abundance was initially quantified using raw read counts. Differential expression analysis was then conducted utilizing the edgeR package [33].

To control for false-positive results during multiple comparisons, the initial *p*-values were strictly adjusted via the False Discovery Rate (FDR) method. Differentially expressed genes (DEGs) were subsequently identified based on two absolute thresholds. First, the absolute log2 fold change required a minimum value of 1.0. Second, the FDR-adjusted *p*-value had to remain strictly below 0.05.

Subsequent functional annotations characterized these specific DEGs. The Blast2GO pipeline was deployed to assign GO terms [34]. Parallel metabolic pathway assignments were retrieved directly from the KEGG database [35]. Finally, statistical enrichment of specific GO categories and KEGG pathways was calculated using standard hypergeometric tests. In these statistical models, the entire assembled transcriptome functioned as the background reference.

2.6. Verification of Transcriptomic Data via qRT-PCR

To validate the reliability of the RNA-Seq results, 16 representative DEGs were selected for independent qRT-PCR analysis. Specific amplification primers were designed using Primer Premier 5.0 software. Detailed primer sequences are provided in Table S1.

Independent fresh RNA samples were extracted from leaves, stems, bast fibers, and both floral types. This extraction utilized a Plant RNA Extraction Kit (Vazyme, Nanjing, China). First-strand cDNA was immediately synthesized using the All-In-One 5× RT MasterMix (Applied Biological Materials Inc., Shanghai, China).

Quantitative PCR assays were executed on a QuantStudio 1 Real-Time PCR System. Reaction mixtures were prepared utilizing the Blastaq 2× qPCR MasterMix. The BnActin gene functioned as the internal reference for target gene normalization. All amplification reactions were strictly performed utilizing three biological replicates and corresponding technical replicates.

2.7. Characterization of AS and Differential Splicing Patterns

AS events were identified from the non-redundant PacBio dataset using SUPPA2 [36]. This tool classified structural variations into seven standard categories. These categories included retained introns (RI), skipping exons (SE), alternative 3′ and 5′ splice sites (A3/A5), mutually exclusive exons (MX), and alternative first or last exons (AF/AL).

Quantification of splicing events required Percent Spliced In (PSI) values. These PSI values were calculated using transcript per million (TPM) data. The TPM data originated directly from the Illumina short-read libraries.

Differential alternative splicing (DAS) analysis was performed to compare male and female flowers. The diffSplice function within SUPPA2 executed this comparison. Strict thresholds defined significant DAS events. Specifically, the absolute change in PSI (|ΔPSI|) had to exceed 0.1. Simultaneously, the *p*-value required a threshold strictly below 0.05.

Venn diagrams visualized overlapping DAS events across different male germplasms relative to the female tissues. Finally, the specific DAS-associated genes underwent functional annotation. This annotation process utilized both the GO framework and the Swiss-Prot database.

2.8. Data Processing and Statistical Evaluation

The relative expression levels of genes were calculated via the standard $2^{-\Delta\Delta Ct}$ method. Data visualization and graph construction were performed utilizing Origin 2024 software. All statistical

evaluations relied on Analysis of Variance (ANOVA). A predefined threshold of $p < 0.05$ was used to establish statistical significance. Within the generated figures, distinct alphabetical letters designate significant differences among the analyzed groups.

To assess genetic identity between GBN09 and GBN10, high-quality paired-end RNA-Seq reads from floral tissues were aligned to the ramie reference genome using HISAT2 (v2.2.1). Transcriptome-wide single nucleotide polymorphism (SNP) variant calling was performed using BCFtools (v1.17) mpileup with parameters -I -d 500, followed by variant filtering (QUAL > 20, MQ > 30). Fixed homozygous divergent SNPs between GBN09 and GBN10 were identified, and the sequence divergence rate was quantified across the callable expressed transcriptome.

3. Result

3.1. Characterization of the GBN09 Transcriptome via PacBio SMRT Sequencing

A single SMRTbell library was constructed using pooled RNA from the stems, leaves, flowers, and epidermis of the GBN09 line. Sequencing on the PacBio Sequel platform generated 368.2 Gb of raw data. Adapter sequences, low-quality reads (score < 0.75), and short fragments (<50 bp) were removed. This filtering step yielded 6,826,654 valid polymerase reads (mean length: 53,936 bp; N50: 119,428 bp). Subsequent processing generated 66,357,794 subreads (mean length: 984 bp; N50: 1596 bp). The ROI length distribution matched the expected library size (Figure S1).

ROIs were analyzed to identify full-length non-chimeric (FLNC) sequences. FLNC identification required the simultaneous presence of 5′ primers, 3′ primers, and poly-A tails. A total of 660,577 FLNC reads were identified, accounting for 71.38% of all ROIs (Table 1; Figure S2). The average FLNC length was 1420 bp, and the mean insert quality score was 0.9814. These metrics suggest minimal artificial concatemer formation in the library.

Illumina short-read data were utilized to correct errors in the FLNC dataset. Following hybrid error correction and redundancy removal, 63,082 high-quality (HQ) consensus isoforms (mean length: 1520 bp) and 350 low-quality (LQ) isoforms (mean length: 872 bp) were generated. Distributions of subread lengths and FLNC density are presented in Figure 1A,B.

To directly confirm the clonal origin and homogeneous genetic background between GBN09 and GBN10, we performed transcriptome-wide SNP variant calling using high-depth RNA-Seq data aligned to the reference genome (Table S2). Out of 499,438 high-quality SNPs identified against the reference, GBN09 and GBN10 shared a 0.9965 identical genetic background. Across the ~35,000,000 bp expressed callable transcriptome, only 1723 fixed homozygous divergent SNPs were detected between the two lines, representing an extremely low sequence divergence rate of 0.0049%. This genomic evidence unequivocally verifies that GBN10 is a clonal mutant of GBN09, and demonstrates that the subsequent divergence in gene expression predominantly reflects sex-specific organ differentiation and metabolic reprogramming rather than genetic heterogeneity.

Table 1. Summary of PacBio SMRT sequencing statistics for GBN09.

Category	Read Count	Total Bases (bp)	Average Length (bp)
Polymerase Reads	6,826,654	368,202,410,144	53,936
Subreads	66,357,794	65,313,589,080	984
ROIs	925,367	1,421,634,442	1536
Number of FLNC	660,577	937,766,235	1420
HQ Isoforms	63,082	95,874,316	1520
LQ Isoforms	350	305,226	872

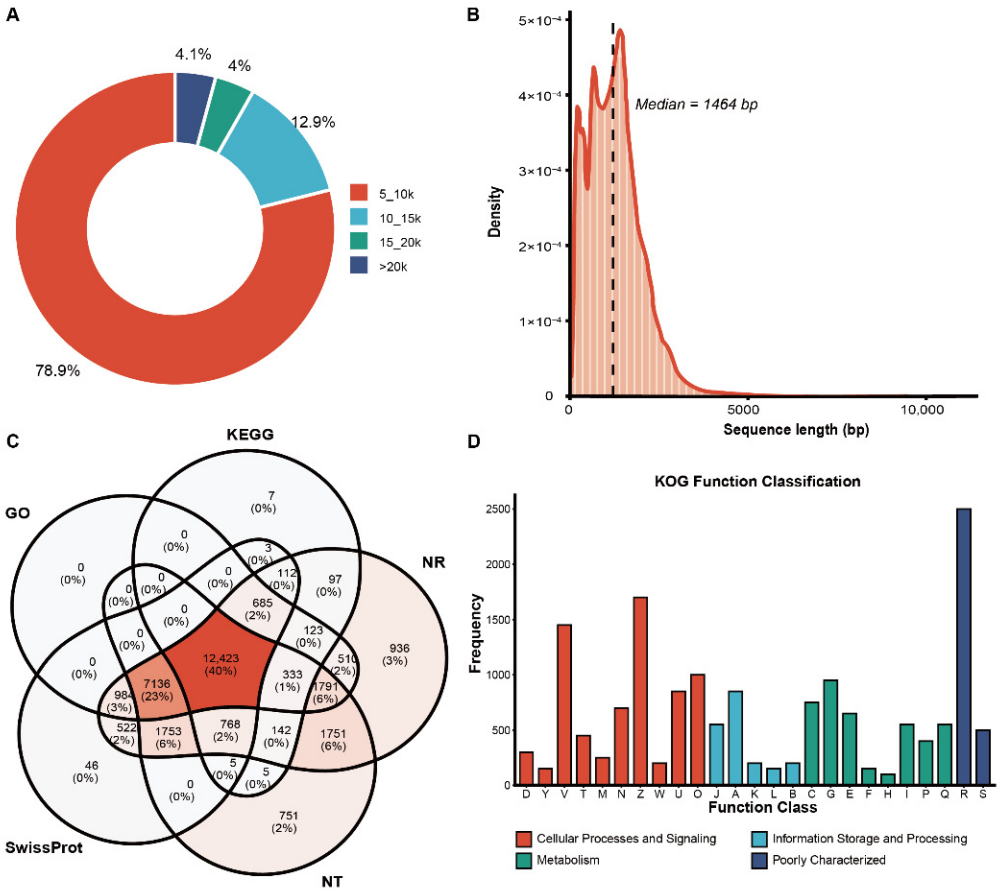

Figure 1. Sequence features and functional annotation of the GBN09 full-length transcriptome. (**A**) Subread length distribution. (**B**) Density plot of full-length non-chimeric (FLNC) read lengths. (**C**) Venn diagram of gene annotations across Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Nucleotide (Nt), and Non-redundant protein (Nr) databases. (**D**) Eukaryotic Orthologous Groups (KOG) functional classification of consensus sequences. Bar colors represent four main categories: red for cellular processes and signaling; cyan for information storage and processing; green for metabolism; and blue for poorly characterized functions. Detailed KOG subgroup labels: [A] RNA modification and processing machinery; [B] Dynamics and structure of chromatin; [C] Conversion and production of cellular energy; [D] Chromosome partitioning, cell division, and cell cycle control; [E] Metabolism and transport of amino acids; [F] Metabolism and transport of nucleotides; [G] Metabolism and transport of carbohydrates; [H] Metabolism and transport of coenzymes; [I] Metabolism and transport of lipids; [J] Ribosomal biogenesis, structure, and translation processes; [K] Transcriptional regulation; [L] DNA repair, recombination, and replication; [M] Biogenesis of the cell wall, envelope, and membrane; [N] Cellular motility; [O] Chaperone functions, protein turnover, and posttranslational modifications; [P] Metabolism and transport of inorganic ions; [Q] Catabolism, transport, and biosynthesis of secondary metabolites; [R] Predictions of general function only; [S] Unknown functional roles; [T] Mechanisms of signal transduction; [U] Vesicular transport, secretion, and intracellular trafficking; [V] Cellular defense mechanisms; [W] Structures of the extracellular matrix; [Y] Structural components of the nucleus; [Z] Cytoskeletal framework.

3.2. Comprehensive Functional Assignment for High-Quality Isoforms

The assembled high-quality transcripts were annotated using six public databases: KOG, KEGG, GO, Swiss-Prot, Nt, and Nr (Table S3). A total of 30,885 distinct isoforms received at least one functional annotation (Table S4).

The Nr database yielded the highest number of mapped sequences (30,066 transcripts, 97.35%). Subsequent annotation proportions were observed in Nt (86.96%), Swiss-Prot (79.12%), GO (77.67%), and KEGG (47.61%) (Figure 1C). Transcript length analysis indicated that 99.8% of the annotated sequences were shorter than 6 kb. Sequences ranging from 1000 to 2000 bp constituted the largest subgroup, representing approximately 44% (13,585 transcripts) of the dataset (Table S4).

Annotation overlap was evaluated across the databases (Figure 1C). Simultaneous alignments in Swiss-Prot, Nt, and Nr occurred for 22,081 sequences. A total of 12,424 transcripts shared annotations across all five primary databases. Exclusive alignments were minimal; only 936 and 751 sequences mapped uniquely to Nr and Nt, respectively. No transcripts were exclusively annotated by the GO database.

KOG classification was applied to 13,079 sequences. These transcripts were assigned to 25 distinct KOG categories within four primary domains: cellular processes and signaling, information storage and processing, metabolism, and poorly characterized functions (Figure 1D). The most heavily represented categories included "General function prediction only" (Cluster R), "Posttranslational modification, protein turnover, chaperones" (Cluster O), and "Signal transduction mechanisms" (Cluster T). These dominant functional assignments align with robust active protein synthesis and complex signaling networks in the sampled tissues.

3.3. Profiling of Regulatory Elements: TFs, Structural Variants, lncRNAs, and Microsatellites

A total of 565 putative TFs were identified within the assembled transcriptome. The IAA family constituted the largest group, comprising 61 members. The AP2/ERF and WRKY families were subsequently ranked, containing 45 and 44 members, respectively (Figure 2A).

To assess transcript diversity in the absence of a high-quality reference genome, Iso-Seq reads were aligned to the newly generated UniTransModels. Single isoforms were produced by 8774 loci. AS events generated multiple variants at the remaining loci: 2657 and 960 loci produced two and three isoforms, respectively. Loci generating more than three isoforms accounted for 5.18% of the total dataset (Figure S3). The AStalavista analysis identified 311 distinct AS events (Table S5). RI was the predominant AS type (110 events), followed by A3 splice sites (96 events) and A5 splice sites (74 events). SE events were detected in 19 cases (Figure 2B).

GO and KEGG functional enrichment analyses were conducted on genes undergoing AS. Within the molecular function category, significant enrichment was observed for 'catalase activity' and 'phosphoethanolamine N-methyltransferase activity' (Table S6). Enriched biological processes primarily involved 'mRNA/RNA processing' and 'phosphatidylcholine biosynthetic process' (Table S7). Significant cellular components included 'viral nucleocapsid' and 'glyoxysome' (Table S8; Figure S4). KEGG mapping assigned the 311 AS-associated genes to 48 distinct metabolic and signaling pathways (Table S9).

A total of 16,617 lncRNAs were predicted from the consensus sequences, with maximum lengths reaching 7100 bp (Figure 2C). Sequence homology in the Swiss-Prot database was confirmed for 9713 lncRNAs. GO and KEGG alignments provided annotations for 10,135 and 6123 lncRNAs, respectively (Table S10). Pathway enrichment indicated significant lncRNA associations with 'Photosynthesis', 'Oxidative phosphorylation', 'Spliceosome', and 'Ribosome' (Figure S5).

Microsatellite analysis via MISA identified 31,759 SSRs (Table S11). Dinucleotide repeats represented the predominant motif type (32.07%, 10,158 SSRs), followed by trinucleotide repeats (14.37%, 4565 SSRs). Frequencies for longer motif types were minimal: tetranucleotides (1.03%), pentanucleotides (0.42%), and hexanucleotides (0.39%) (Figure 2D).

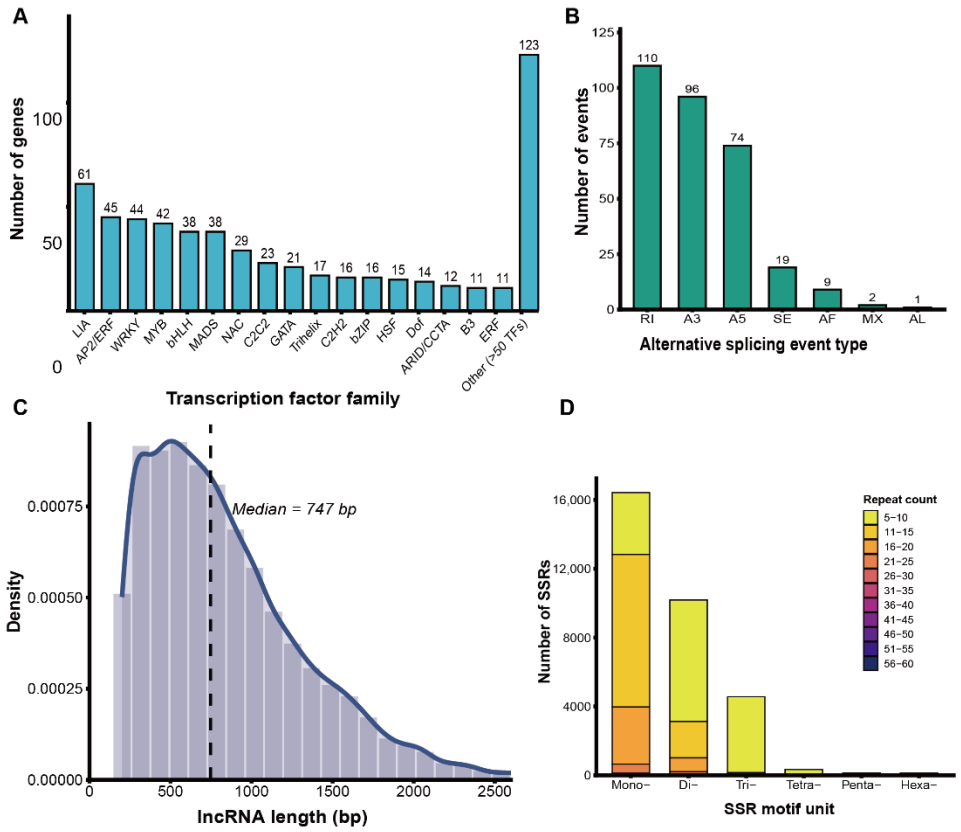

Figure 2. Characteristics of regulatory elements and structural variants in the full-length transcriptome. (**A**) Distribution of transcription factor (TF) families across the assembled transcripts. (**B**) Classification and occurrence of alternative splicing (AS) modes. The AS categories include skipping exon (SE), retained intron (RI), mutually exclusive exon (MX), alternative 5′ splice site (A5), alternative 3′ splice site (A3), alternative first exon (AF), and alternative last exon (AL). (**C**) Size distribution of the predicted long non-coding RNAs (lncRNAs). (**D**) Frequencies of simple sequence repeat (SSR) motifs.

3.4. Organ-Specific Transcriptional Dynamics and Differential Expression Profiles in Sex-Divergent Ramie Lines

To quantify transcript abundance, the PacBio full-length isoforms served as the reference for mapping Illumina short reads via the RSEM algorithm utilizing Bowtie2. Approximately 1.59 billion clean reads were successfully aligned. The global mapping efficiency exceeded 73% across all biological samples. Specifically, the average unique mapping rate reached 73.52%, while multi-mapped reads constituted 26.48% (Table S12).

Following the normalization of gene expression levels, DEGs were detected using the DEGseq package. Principal Component Analysis (PCA) based on the global expression profiles demonstrated tight clustering among biological replicates. The primary sources of variance, PC1 and PC2, accounted for 25.7% and 20.0% of the total variance, respectively, effectively discriminating the samples by tissue origin and genetic background (Figure S6).

Relative to the reference genome, a total of 8414 DEGs were identified across the four sampled tissues (leaves, young stems, flowers, and middle bast fibers), comprising 4161 up-regulated and 4253 down-regulated transcripts. The abundance of DEGs exhibited substantial tissue specificity. The most pronounced transcriptional divergence between the gynoecious (GBN09) and androecious (GBN10) lines occurred in floral tissues, encompassing 2553 DEGs (representing 30.34% of the total differential transcriptome). This was followed by young stems (1156 DEGs, 13.74%), leaves (257 DEGs, 3.05%), and middle bast fibers (195 DEGs, 2.32%) (Figure 3A).

Venn diagram analysis highlighted profound tissue-specificity among the differential transcripts (Figure 3B). The vast majority of DEGs were spatially restricted: 5161 loci were specific to flowers, 1121

to young stems, 668 to bast fibers, and 187 to leaves. A mere 18 transcripts were constitutively differentially expressed across all four tissues. Within this core set, 16 genes were universally up-regulated (Figure 3C), whereas zero transcripts showed universal down-regulation (Figure 3D). In the maternal GBN09 line, 2881 loci displayed unique up-regulation in floral organs, contrasting with 646, 188, and 39 genes in fibers, stems, and leaves, respectively.

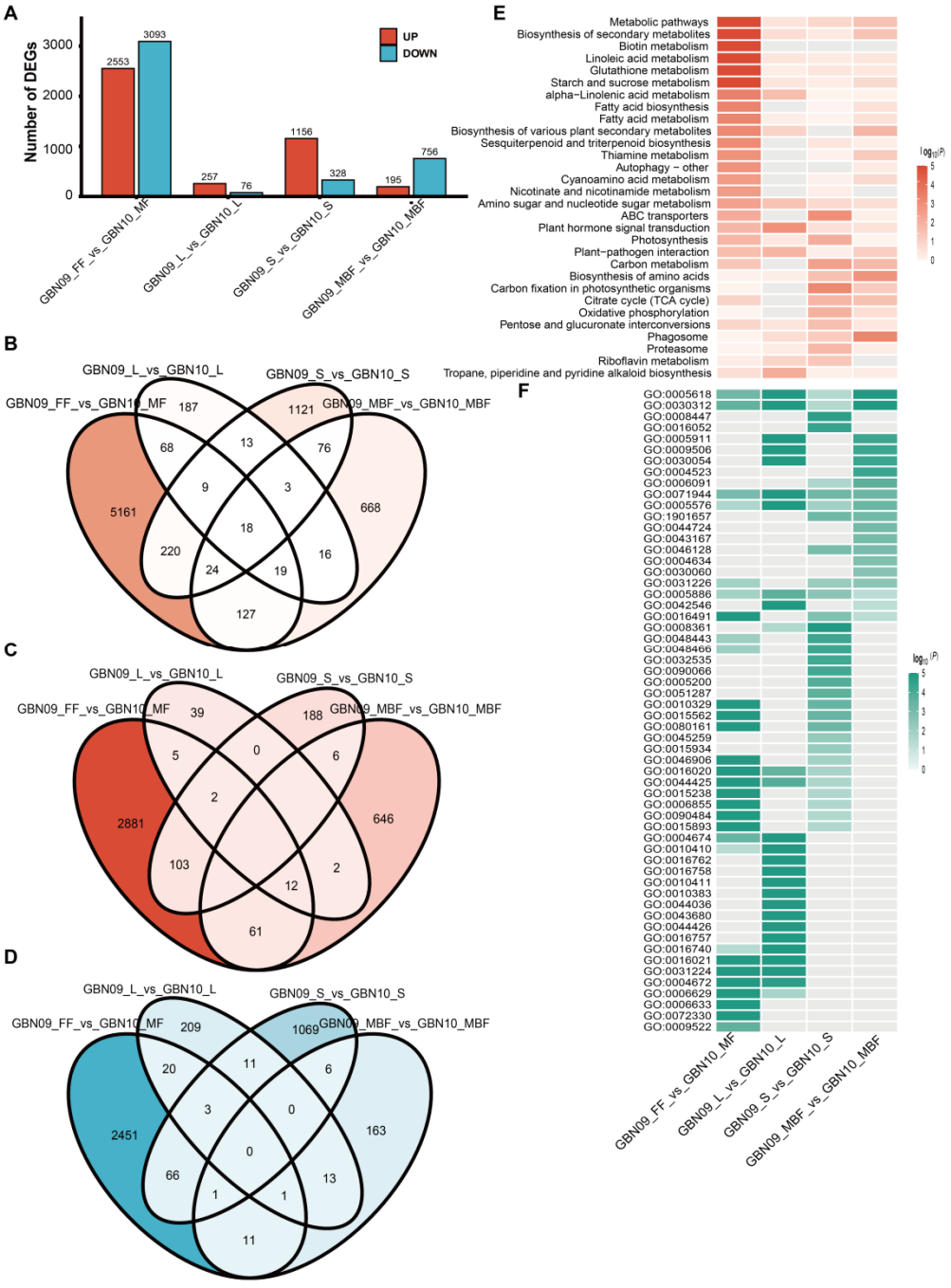

Figure 3. Pathway enrichment and GO analysis for DEGs in different organs. (**A**) Summary of up-regulated and downregulated genes. L indicates leaf; FF indicates female flower; MF indicates male flower; S indicates young stem; MBF indicates middle bast fiber. (**B**) Venn diagrams of all DEGs. (**C**) Venn diagrams of up-regulated DEGs. (**D**) Venn diagrams of down-regulated DEGs. (**E**) KEGG pathway enrichment analysis of DEGs ($p < 0.05$, the data were processed by −log10 (p value)). (**F**) GO enrichment analysis of DEGs.

KEGG enrichment analysis was executed to delineate the metabolic pathways governing these DEGs. In floral tissues, transcripts up-regulated in the female line showed significant enrichment in pathways associated with linoleic acid metabolism, photosynthesis, plant hormone signal transduction, and secondary metabolite biosynthesis. Conversely, down-regulated transcripts in female flowers were predominantly mapped to fatty acid biosynthesis and degradation, biotin metabolism, and nicotinate and nicotinamide metabolism (Figure 3E; Table S13). These metabolic shifts imply that female floral development necessitates intensive carbon fixation, phytohormone signaling, and secondary metabolism.

Vegetative organs displayed fewer significantly enriched pathways (Figure 3E; Tables S14 and S15). However, DEGs identified in the middle bast fibers were specifically clustered within 'phagosome', 'aminoacyl-tRNA biosynthesis', and 'cysteine and methionine metabolism', underscoring the specialized biochemical demands of fiber cell development (Figure 3E; Table S16).

GO functional categorization partitioned the DEGs into biological process, cellular component, and molecular function domains. Leaves contained the highest number of significantly enriched GO terms (189 terms, Table S17), followed by flowers (177, Table S18), stems (96, Table S19), and middle bast fibers (16, Table S20). Notably, floral DEGs were heavily enriched in monocarboxylic and fatty acid biosynthetic processes, alongside 'photosystem I' (GO:0009522). Leaf-enriched terms centered on cell wall remodeling activities, such as 'xyloglucan:xyloglucosyl transferase activity' and 'cell wall polysaccharide metabolic process'. Stem-specific DEGs were primarily linked to cellular architecture and energetics, including 'structural constituent of cytoskeleton', 'NAD binding', and 'proton-transporting ATP synthase activity'. Finally, DEGs in bast fibers showed significant enrichment for 'phosphopyruvate hydratase activity' and 'L-ascorbate oxidase activity', indicating vigorous carbohydrate turnover and cell wall modification during fiber elongation (Figure 3F).

3.5. Orthogonal Verification of Transcriptomic Outputs Using Real-Time PCR

To validate the reliability of the RNA-Seq data, 16 DEGs representing various tissues (leaves, stems, bast fibers, and flowers) were selected for qRT-PCR analysis. As presented in Figure 4, the qRT-PCR results confirmed the differential expression patterns of all 16 candidate genes between the GBN09 and GBN10 lines ($p < 0.05$).

Specifically, in leaves, transcript 33288 (B09hun_1-10k_transcript/33288) was significantly up-regulated in the female line (GBN09), whereas transcripts 47493, 46072, and 48813 exhibited higher expression in the male line (GBN10). In the middle bast fibers, transcripts 46188, 40888, and 20088 were highly expressed exclusively in GBN09. In stem tissues, transcripts 13579 and 13509 showed higher expression in GBN09, while transcripts 55424 and 50783 were up-regulated in GBN10.

In floral tissues, transcripts 37069, 22424, and 19585 were predominantly expressed in female flowers (GBN09). Conversely, transcripts 16529 and 12757 were highly up-regulated in male flowers (GBN10).

Overall, the relative expression levels and fold-change trends measured by qRT-PCR demonstrated a high degree of consistency with the read count/FPKM values obtained from the RNA-Seq data. This concordance confirms the accuracy and reliability of the transcriptomic sequencing and bioinformatic analyses in this study.

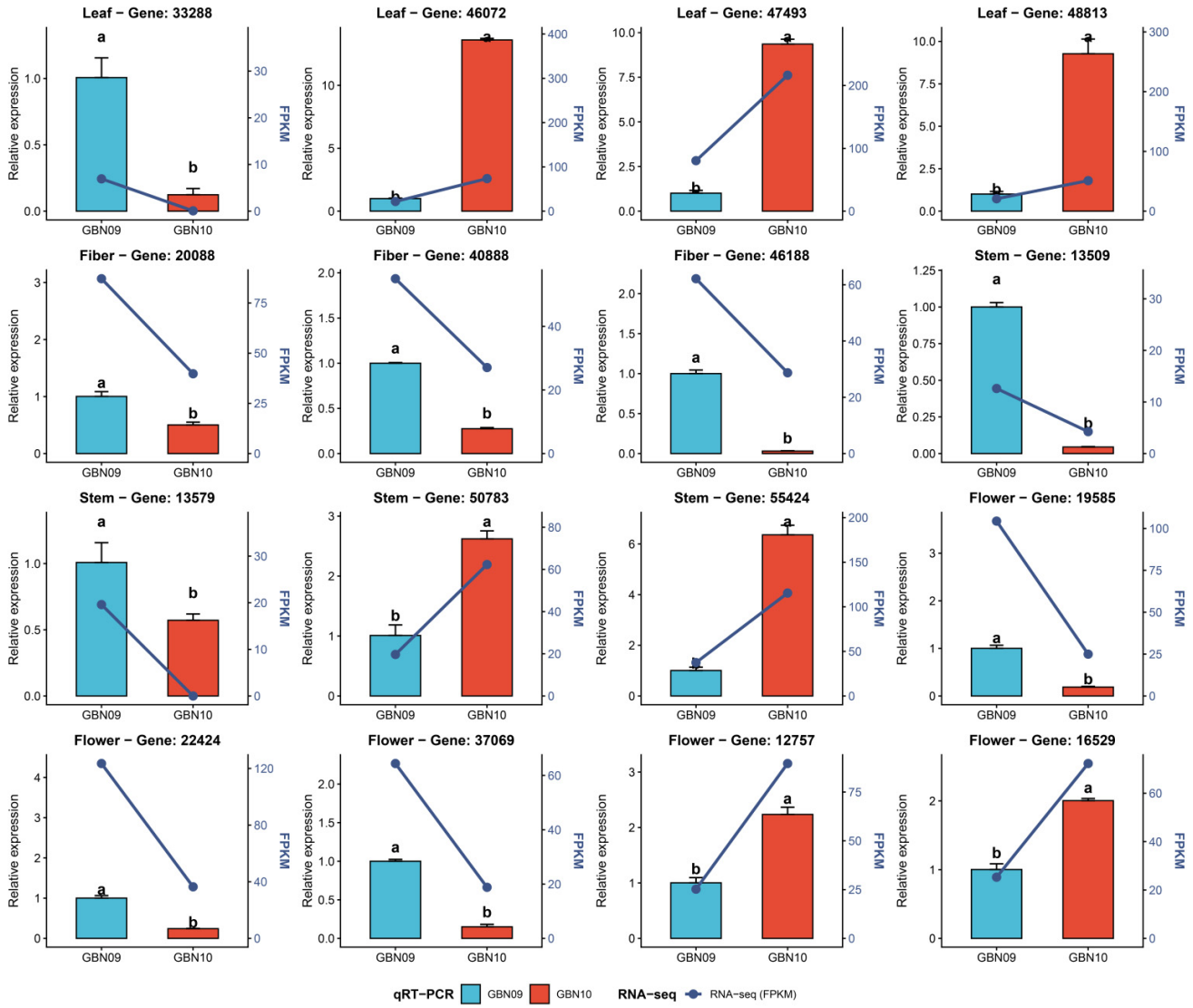

Figure 4. The qRT-PCR expression and RNA-Seq FPKM analysis of 16 DEGs in leaf, fiber, stem, and flower of ramie GBN09 and GBN10.

3.6. Alternative Splicing Analysis in Sexually Dimorphic Ramie Flowers

In addition to differential gene expression, AS is a vital post-transcriptional regulatory mechanism. Here, we adopted the full-length transcriptome as a reference to analyze the AS events in ramie flowers.

A total of 281 AS events were identified in the floral transcriptomes (Figure 5). Among the seven classical AS types, RI was the most frequent, followed by A3 splice site and A5 splice site. This high frequency of RI is consistent with typical plant transcriptomes, which differ from animal systems where SE is more prevalent.

To investigate sex-specific splicing, we identified 274 DAS events between male and female flowers ($p < 0.05$, $|\Delta PSI| > 0.1$). These DAS events corresponded to over 200 specific genes, indicating that AS plays a role in floral sex differentiation in ramie.

According to the annotation data (Table 2), a large portion of the identified DAS genes map to phytohormone pathways—especially auxin signaling—and cell wall formation. Looking closely at specific targets, the Auxin Signaling F-Box 2 (AFB2, PB.5309) transcript favored an A3 splicing site in male flowers, where its inclusion rate (PSI = 0.50) was more than double that of female flowers (PSI = 0.22). We observed a comparable male-biased A3 pattern ($\Delta PSI = -0.32$) in NPY2 (PB.5689), another key player

in auxin patterning. This implies that alternative splicing directly interferes with auxin dynamics as the floral organs mature. Furthermore, to provide direct physical evidence for these sex-biased DAS events at the read level, we generated Sashimi plots for the core auxin-related regulators, AFB2 and NPY2 (Figure S7). The mapping of full-length reads visually confirmed the distinct splice junction usages between female and male flowers. Meanwhile, isoform-specific FPKM quantification revealed robust expression shifts corresponding to these structural variations, providing reliable read-level validation for the AS-mediated post-transcriptional regulation of ramie sex determination. We performed protein-level computational analysis of the splice isoforms of AFB2 and NPY2 (Table S21). For NPY2 (PB.5689), the A3 event produces substantially different proteins: the male-biased isoform (PB.5689.2) encodes a 639-aa protein with a unique 90-aa N-terminal extension compared to the female-preferred isoform (PB.5689.1, 549 aa). The C-termini (containing the BTB/POZ domain) are shared, but the 90-aa N-terminal extension in the male-biased isoform is an entirely novel sequence absent from the female isoform, suggesting that sex-biased A3 splicing generates functionally distinct NPY2 protein variants. For AFB2 (PB.5309), the A3 event is located in the 5′ UTR, producing identical 573-aa proteins but potentially affecting translational regulation through 236-bp UTR length variation.

Beyond hormone signaling, structural and proliferative genes were also subject to sex-specific splicing. For instance, the Glucuronosyltransferase 14B (GlcAT14B, PB.1367) transcript retained an intron (RI event) much more frequently in female tissues (PSI = 0.93) compared to male tissues (PSI = 0.52). In contrast, the aquaporin PIP1-4 (PB.5641) underwent an SE event predominantly in males. A significant splicing shift (ΔPSI = −0.13) was also captured in GRF1-Interacting Factor 3 (GIF3, PB.3314). Such structural alterations in the transcripts likely help tailor cell growth to the specific needs of male versus female flowers. Collectively, these findings suggest that alternative splicing acts as an additional regulatory layer, functioning alongside gene expression networks to modulate sexual dimorphism in ramie flowers.

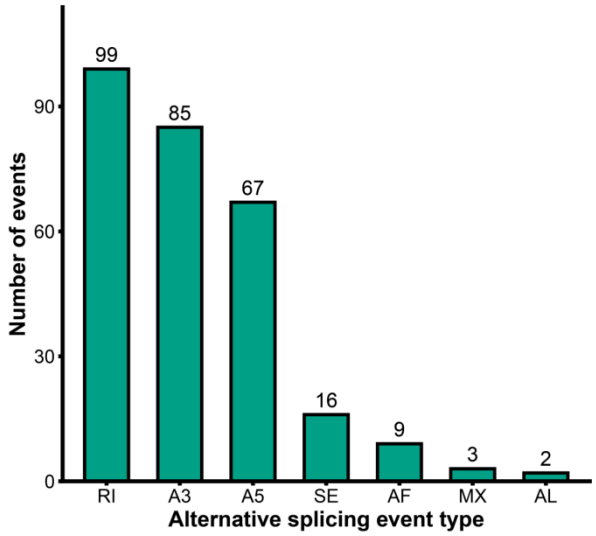

Figure 5. Classification and frequency of alternative splicing (AS) events in ramie flowers. The bar chart shows the number of seven AS types: intron retention (RI), alternative 3′ splice site (A3), alternative 5′ splice site (A5), exon skipping (SE), mutually exclusive exons (MX), alternative first exon (AF), and alternative last exon (AL).

Table 2. Key differentially spliced genes involved in ramie sex determination.

Gene ID	AS Type	PSI (Male)	PSI (Female)	ΔPSI	*p*-Value	Annotation & Function
PB.5309	A3	0.50	0.22	−0.28	0.031	AFB2 (Auxin Signaling F-Box 2)
PB.5689	A3	0.32	0.00	−0.32	0.026	NPY2 (Auxin-mediated organogenesis)
PB.3314	A3	0.40	0.27	−0.13	0.046	GIF3 (GRF-interacting factor 3, Cell proliferation)
PB.5641	SE	0.64	0.30	−0.34	0.007	PIP1-4 (Aquaporin, Water transport)

PB.1367	RI	0.52	0.93	0.41	0.025	GlcAT14B (Cell wall xylan synthesis)
PB.2706	A3	0.35	0.60	0.25	0.036	PRP (Proline-rich protein, Cell wall structure)
PB.1752	RI	0.50	0.13	−0.37	0.027	GDSL (Esterase/lipase, Lipid metabolism)

Note: The inclusion rate is calculated as the Percent Spliced In (PSI, ranging from 0 to 1). ΔPSI = PSI(Female flower) − PSI(Male flower); positive values indicate higher inclusion in female flowers and negative values indicate higher inclusion in male flowers. Significant DAS events were defined by $p < 0.05$ and $|\Delta PSI| > 0.1$.

4. Discussion

Ramie is a crop of considerable economic significance, whose fiber yield and quality are closely associated with its floral sex. Due to the lack of high-quality reference genomes for specific sex-divergent lines, full-length transcriptome sequencing is necessary to understand their genetic basis. In this study, we combined PacBio Iso-Seq with Illumina RNA-Seq to construct a comprehensive transcriptome for the gynoecious ramie line GBN09. This provides a fundamental reference for studying organ development and sex differentiation in this species.

Our sequencing generated 63,082 high-quality isoforms with an average length of 1520 bp. Among them, 30,885 unigenes were successfully annotated using KOG, KEGG, GO, Swiss-Prot, Nt, and Nr databases. The high annotation rate confirms the reliability of the assembled transcripts and provides extensive data for downstream functional analyses.

A total of 565 TFs were identified, with the IAA, AP2/ERF, and WRKY families being the most abundant. IAA proteins are key repressors in the auxin signaling pathway, regulating plant developmental processes [37]. Other prominently expressed regulators included AP2/ERF and WRKY transcription factors. Members of the AP2/ERF family heavily influence secondary metabolic cascades—driving the production of terpenes, phenolics, and alkaloids [38]. Meanwhile, WRKY factors act as vital networking hubs for hormonal signaling, bridging ABA, ethylene, and auxin dynamics during the transition to flowering [39]. Given their robust transcriptional output in our dataset, these particular TF families likely orchestrate the delicate balance between developmental cues and stress adaptation in growing ramie organs.

Unlike the systemic sexual dimorphism seen in animals, sex differences in plants generally manifest locally within the flowers [40]. Deciphering how this localized sex determination happens remains a central puzzle in developmental genetics [41]. From an agricultural perspective, however, unlocking these reproductive pathways is more than an academic exercise; it is an essential step for breeding commercially valuable crops whose field yields are directly tied to floral sex.

When comparing the transcriptomic profiles of the GBN09 and GBN10 lines, we captured a massive shift of 8414 differentially expressed genes, with the most dramatic transcriptional rewiring localizing to the stems and flowers. A closer look at the female floral transcriptome showed a clear functional bias: transcripts tied to phytohormone networking, linoleic acid processing, and broader secondary metabolism were disproportionately activated. This targeted up-regulation strongly implies that driving a ramie plant toward a female phenotype requires tight synchronization between specialized metabolic synthesis and localized hormonal signaling.

The reliance on specific hormones and metabolic shifts to steer floral sex is not unique to ramie; it appears to be a deeply conserved evolutionary strategy across diverse plant lineages. For example, early floral differentiation in spinach leans heavily on precise gibberellin and auxin signaling [42], while the sex of Salix variegata blossoms is tightly linked to phenylpropanoid and flavonoid fluxes [43]. Similarly, Fraxinus mandshurica manages its floral dimorphism through a complex interplay of energy metabolism, photoperiod cues, and hormonal triggers [44]. Seeing analogous secondary metabolism and phytohormone networks dominate our ramie dataset strongly suggests that Boehmeria nivea utilizes these same fundamental regulatory circuits to establish its sexual phenotypes.

While raw transcript abundance sets the baseline for organogenesis, post-transcriptional recalibration via AS fine-tunes the final functional output. When we mapped the structural variations between male and female ramie flowers, we captured 274 distinct differential AS events. Interestingly, a large portion of these splicing shifts targeted transcripts involved in auxin signaling and cell wall architecture (Table 2). A prime example is the auxin receptor AFB2 and the developmental regulator NPY2, both of which underwent sex-biased splicing. By generating distinct protein isoforms, this structural variation likely alters cellular sensitivity to local auxin gradients without drastically changing the overall gene expression level. We noticed a similar post-transcriptional pattern in structural components like the GlcAT14B gene. Ultimately, establishing the distinct physical traits of male versus female flowers requires this dual-action mechanism: broad transcriptional shifts running in parallel with highly targeted splicing modifications. It should be noted that the DAS events identified here are computational predictions based on PSI values; experimental validation by isoform-specific qPCR, fragment electrophoresis, or transcript cloning is required to confirm differences in abundance among distinct splice isoforms between male and female flowers.

Moving forward, the extensive full-length transcriptomic catalog generated here provides a massive foundational dataset, but it remains largely descriptive at this stage. To definitively lock down the biological roles of the candidate genes governing ramie sex and organogenesis, we must transition from *in silico* predictions to *in vivo* functional assays. Isolating and manipulating specific targets via CRISPR/Cas9, RNA interference, or transgenic overexpression will be the critical next steps to prove their exact developmental functions. Integrating transcriptomic data with proteomic and metabolomic analyses will further improve our understanding of the molecular mechanisms underlying ramie organogenesis and sex determination. We acknowledge that the DAS events identified through SUPPA2 analysis of PacBio–Illumina hybrid data primarily represent transcriptomic evidence. While PacBio Iso-Seq provides direct observation of full-length isoform structures, and our Sashimi plots (Figure S7) confirm splice junction support at the raw-read level, the PSI-based DAS classification should be further corroborated by isoform-specific qPCR or transcript cloning in future molecular studies. Nevertheless, the present study establishes a robust transcriptomic foundation and identifies priority candidate genes—particularly NPY2, whose splice isoforms produce substantially different proteins—for such functional validation.

From a breeding perspective, the sex-biased splicing of auxin-responsive genes such as AFB2 and NPY2 provides candidate molecular markers for early sex identification in ramie seedlings, well before floral morphogenesis becomes visible. Splice isoform-specific markers derived from these loci could be developed into PCR-based sex diagnostic assays. In addition, because female flowers are generally associated with higher fiber yield, manipulating the splicing or expression of AFB2/NPY2 via CRISPR/Cas9 or RNA interference may offer a route to control sex expression and enhance fiber productivity in monoecious or female-biased breeding lines.

5. Conclusions

In this study, we combined PacBio Iso-Seq and Illumina RNA-Seq to construct a full-length transcriptome for the ramie GBN mutant lines. This approach identified numerous novel isoforms and lncRNAs, significantly improving the existing genomic resources for ramie. Our transcriptomic analysis revealed that floral sex differentiation is closely associated with differential gene expression in fatty acid metabolism and in gibberellin and auxin signaling pathways. Furthermore, post-transcriptional regulation via AS plays a critical role in this process. Specifically, AS events in key auxin-responsive genes, such as AFB2 and NPY2, suggest that alternative splicing modulates hormone sensitivity during male and female floral organogenesis. Together, these transcriptional and post-transcriptional regulatory networks provide new insights into the molecular basis of the early-flowering and sex-divergent phenotypes in ramie. These findings offer valuable candidate genes for future molecular breeding aimed at controlling sex expression in ramie cultivation.

Supplementary Materials

The following supporting information can be found at: https://www.sciepublish.com/article/pii/1206, Figure S1: Histogram of ROI read length; Figure S2: Density distribution of FLNC read lengths; Figure S3: Distribution of isoform counts per UniTransModel; Figure S4: Top 25 GO enrichment terms for genes undergoing Alternative Splicing (AS); Figure S5: Top 25 KEGG enrichment in LncRNA; Figure S6: All_gene.PCA; Figure S7: Transcriptomic evidence for sex-biased splicing of AFB2 and NPY2. Table S1: List of primers used in the study; Table S2: Transcriptome-wide genetic variation analysis confirming the near-isogenic background between the gynoecious (GBN09) and androecious (GBN10) mutants; Table S3: Functional annotation of HQ isoforms by the PacBio Sequel platform; Table S4: Function annotation statisticians; Table S5: Identification of AS events in GBN09; Table S6: GO enrichment results of Molecular Function; Table S7: GO enrichment results of Biological Process; Table S8: GO enrichment results of Cellular Component; Table S9: KEGG pathway results of 311 AS genes; Table S10: Annotation of lncRNAs; Table S11: Total number of identified SSRs by the MISA software; Table S12: Total mapped statix of Illumina sequence; Table S13: KEGG enrichment pathway of GBN09 female flower; Table S14: KEGG enrichment pathway of GBN09 leaf; Table S15: KEGG enrichment pathway of GBN09 stem; Table S16: KEGG enrichment pathway of GBN09 bast fiber; Table S17: GO enrichment of GBN09 leaf; Table S18: GO enrichment of GBN09 female flower; Table S19: GO enrichment of GBN09 stem; Table S20: GO enrichment of GBN09 bast fiber; Table S21: Protein-level analysis of splice isoforms for key DAS genes.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the authors used DeepL and Grammarly for language translation, language polishing, and grammar checking. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Author Contributions

Software, H.Q. and J.Z.; Validation, H.Q.; Formal Analysis, H.Q., S.Z. and J.Z.; Investigation, J.Z. and Q.S.; Data Curation, H.Q., S.Z. and Q.S.; Writing—Original Draft Preparation, H.Q.; Writing—Review & Editing, S.Z. and H.X.; Supervision, H.X.; Project Administration, H.X.; Funding Acquisition, H.X.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Raw PacBio Iso-Seq and Illumina RNA-Seq reads will be deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNAxxxxxx after online publication. Processed isoform annotations, gene expression matrices, and alternative splicing results are available from the corresponding author upon reasonable academic request.

Funding

This research was funded by the Yuelushan Laboratory Breeding Program (YLS-2025-ZY02056), the Hunan Provincial Natural Science Foundation (2025JJ50167), and the National Natural Science Foundation of China (Grant No.32272137).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Huang C, Zhou J, Jie Y, Xing H, Zhong Y, Yu W, et al. A Ramie bZIP Transcription Factor *BnbZIP2* Is Involved in Drought, Salt, and Heavy Metal Stress Response. *DNA Cell Biol.* **2016**, *35*, 776–786. DOI:10.1089/dna.2016.3251
2. Fu H, Wang C, Cui G, She W, Zhao L. Ramie Yield Estimation Based on UAV RGB Images. *Sensors* **2021**, *21*, 669. DOI:10.3390/s21020669
3. Li F, Zeng Z, Huang R, Wang Y, Liu T. Identification of proteins associated with bast fiber growth of ramie by differential proteomic analysis. *BMC Genom.* **2021**, *22*, 865. DOI:10.1186/s12864-021-08195-9
4. Zeng Z, Li F, Huang R, Wang Y, Liu T. Phosphoproteome analysis reveals an extensive phosphorylation of proteins associated with bast fiber growth in ramie. *BMC Plant Biol.* **2021**, *21*, 473. DOI:10.1186/s12870-021-03252-7
5. Kandimalla R, Kalita S, Choudhury B, Devi D, Kalita D, Kalita K, et al. Fiber from ramie plant (*Boehmeria nivea*): A novel suture biomaterial. *Mater. Sci. Eng. C* **2016**, *62*, 816–822. DOI:10.1016/j.msec.2016.02.040
6. Liu T, Zhu S, Tang Q, Tang S. Genome-wide transcriptomic profiling of ramie (*Boehmeria nivea* L. Gaud) in response to cadmium stress. *Gene* **2015**, *558*, 131–137. DOI:10.1016/j.gene.2014.12.057
7. Mu L, Cai M, Wang Z, Liu J, Liu T, Wanapat M, et al. Assessment of ramie leaf (*Boehmeria nivea* L. Gaud) as an animal feed supplement in P.R. China. *Trop. Anim. Health Prod.* **2020**, *52*, 115–121. DOI:10.1007/s11250-019-01997-w
8. Zhang H, He Y, Yu L, Tang S, Tan Z, Zhou C, et al. The potential of ramie as forage for ruminants: Impacts on growth, digestion, ruminal fermentation, carcass characteristics and meat quality of goats. *Anim. Sci. J.* **2019**, *90*, 481–492. DOI:10.1111/asj.13165
9. Bjornson M, Kajala K, Zipfel C, Ding P. Low-cost and High-throughput RNA-Seq Library Preparation for Illumina Sequencing from Plant Tissue. *Bio-Protocol* **2020**, *10*, e3799. DOI:10.21769/BioProtoc.3799
10. Liu YH, Wang L, Yu L. The principle and application of the single-molecule real-time sequencing technology. *Hereditas* **2015**, *37*, 259–268. DOI:10.16288/j.yczz.14-323
11. Li W, Liu Z, Feng H, Yang J, Li C. Characterization of the Gene Expression Profile Response to Drought Stress in *Populus ussuriensis* Using PacBio SMRT and Illumina Sequencing. *Int. J. Mol. Sci.* **2022**, *23*, 3840. DOI:10.3390/ijms23073840
12. Savarese M, Qureshi T, Torella A, Laine P, Giugliano T, Jonson PH, et al. Identification and Characterization of Splicing Defects by Single-Molecule Real-Time Sequencing Technology (PacBio). *J. Neuromuscul. Dis.* **2020**, *7*, 477–481. DOI:10.3233/JND-200523
13. Chen J, Pei Z, Dai L, Wang B, Liu L, An X, et al. Transcriptome profiling using pyrosequencing shows genes associated with bast fiber development in ramie (*Boehmeria nivea* L.). *BMC Genom.* **2014**, *15*, 919. DOI:10.1186/1471-2164-15-919
14. She W, Zhu S, Jie Y, Xing H, Cui G. Expression Profiling of Cadmium Response Genes in Ramie (*Boehmeria nivea* L.) Root. *Bull. Environ. Contam. Toxicol.* **2015**, *94*, 453–459. DOI:10.1007/s00128-015-1502-z
15. Shang L, Zhou Y, Wen S, Wang K, Li Y, Zhang M, et al. Construction of heat stress regulation networks based on Illumina and SMRT sequencing data in potato. *Front. Plant Sci.* **2023**, *14*, 1271084. DOI:10.3389/fpls.2023.1271084
16. Jian H, Sun H, Liu R, Zhang W, Shang L, Wang J, et al. Construction of drought stress regulation networks in potato based on SMRT and RNA sequencing data. *BMC Plant Biol.* **2022**, *22*, 381. DOI:10.1186/s12870-022-03758-8
17. Yang F, Lv G. Characterization of the gene expression profile response to drought stress in Haloxylon using PacBio single-molecule real-time and Illumina sequencing. *Front. Plant Sci.* **2022**, *13*, 981029. DOI:10.3389/fpls.2022.981029
18. Zhu Z, Chen H, Xie K, Liu C, Li L, Liu L, et al. Characterization of Drought-Responsive Transcriptome During Seed Germination in Adzuki Bean (*Vigna angularis* L.) by PacBio SMRT and Illumina Sequencing. *Front. Genet.* **2020**, *11*, 996. DOI:10.3389/fgene.2020.00996
19. Wei M, Wang B, Li C, Li X, He C, Li Y. Integrated PacBio SMRT and Illumina sequencing uncovers transcriptional and physiological responses to drought stress in whole-plant *Nitraria tangutorum*. *Front. Genet.* **2024**, *15*, 1474259. DOI:10.3389/fgene.2024.1474259
20. Xiong Y, Xiang X, Xiao C, Zhang N, Cheng H, Rao S, et al. Illumina RNA and SMRT Sequencing Reveals the Mechanism of Uptake and Transformation of Selenium Nanoparticles in Soybean Seedlings. *Plants* **2023**, *12*, 789. DOI:10.3390/plants12040789
21. Rhoads A, Au KF. PacBio Sequencing and its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. DOI:10.1016/j.gpb.2015.08.002

22. Au KF, Underwood JG, Lee L, Wong WH. Improving PacBio Long Read Accuracy by Short Read Alignment. *PLoS ONE* **2012**, *7*, e46679. DOI:10.1371/journal.pone.0046679
23. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. DOI:10.1093/bioinformatics/btl158
24. Liaquat F, Munis MFH, Arif S, Haroon U, Shi J, Saqib S, et al. PacBio Single-Molecule Long-Read Sequencing Reveals Genes Tolerating Manganese Stress in *Schima superba* Saplings. *Front. Genet.* **2021**, *12*, 635043. DOI:10.3389/fgene.2021.635043
25. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, et al. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Mol. Plant* **2016**, *9*, 1667–1670. DOI:10.1016/j.molp.2016.09.014
26. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. DOI:10.1093/nar/gkt646
27. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **2007**, *35*, W345–W349. DOI:10.1093/nar/gkm391
28. Zhao N, Zhou E, Miao Y, Xue D, Wang Y, Wang K, et al. High-quality faba bean reference transcripts generated using PacBio and Illumina RNA-Seq data. *Sci. Data* **2024**, *11*, 359. DOI:10.1038/s41597-024-03204-4
29. Li A, Zhang J, Zhou Z. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinform.* **2014**, *15*, 311. DOI:10.1186/1471-2105-15-311
30. Li W, Fu Y, Lv W, Zhao S, Feng H, Shao L, et al. Characterization of the early gene expression profile in *Populus ussuriensis* under cold stress using PacBio SMRT sequencing integrated with RNA-Seq reads. *Tree Physiol.* **2022**, *42*, 646–663. DOI:10.1093/treephys/tpab130
31. Weng X, Zhou X, Xie S, Gu J, Wang ZY. Identification of cassava alternative splicing-related genes and functional characterization of MeSCL30 involvement in drought stress. *Plant Physiol. Biochem.* **2021**, *160*, 130–142. DOI:10.1016/j.plaphy.2021.01.016
32. Foissac S, Sammeth M. ASTALAVISTA: Dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* **2007**, *35*, W297–W299. DOI:10.1093/nar/gkm311
33. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. DOI:10.1093/bioinformatics/btp616
34. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. DOI:10.1093/bioinformatics/bti610
35. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. DOI:10.1093/nar/28.1.27
36. Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, et al. SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **2018**, *19*, 40. DOI:10.1186/s13059-018-1417-1
37. Figueiredo MRAD, Strader LC. Intrinsic and extrinsic regulators of Aux/IAA protein degradation dynamics. *Trends Biochem. Sci.* **2022**, *47*, 865–874. DOI:10.1016/j.tibs.2022.06.004
38. Chen Q, Li N, Cui X, Ge F. AP2/ERF transcription factors regulate the biosynthesis of terpenoids, phenolics, and alkaloids in plants. *Hortic. Res.* **2026**, *13*, uhaf280. DOI:10.1093/hr/uhaf280
39. Song H, Duan Z, Zhang J. WRKY transcription factors modulate flowering time and response to environmental changes. *Plant Physiol. Biochem.* **2024**, *210*, 108630. DOI:10.1016/j.plaphy.2024.108630
40. Pannell JR. Plant Sex Determination. *Curr. Biol.* **2017**, *27*, R191–R197. DOI:10.1016/j.cub.2017.01.052
41. Andreuzza S. Y Keep Your X? Insights into the Genetic Basis of Plant Sex Chromosome Evolution. *Plant Cell* **2020**, *32*, 1769–1770. DOI:10.1105/tpc.20.00265
42. Li N, Meng Z, Tao M, Wang Y, Zhang Y, Li S, et al. Comparative transcriptome analysis of male and female flowers in *Spinacia oleracea* L. *BMC Genom.* **2020**, *21*, 850. DOI:10.1186/s12864-020-07277-4
43. Zhang H, Li W, Zhang X, Pan R, Tang M, Peng Y, et al. Integrating transcriptome and metabolome analyses to characterize flower development in *Salix variegata* franch., a typical dioecious plant. *Sci. Rep.* **2025**, *15*, 8010. DOI:10.1038/s41598-025-91317-0
44. Zhang K, Cao Y, Guo X, Kong F, Sun H, Jing T, et al. Comparative transcriptome analysis of differentially expressed genes and pathways in male and female flowers of *Fraxinus mandshurica*. *PLoS ONE* **2024**, *19*, e0308013. DOI:10.1371/journal.pone.0308013