frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 February 2025
DOI 10.3389/fpls.2025.1500654

Check for updates

OPEN ACCESS

EDITED BY
Yongfeng Guo,
Chinese Academy of Agricultural Sciences, China

REVIEWED BY
Xiaoxu Li,
Chinese Academy of Agricultural Sciences (CAAS), China
Wenbin Guo,
Higentec Breeding Innovation (ZheJiang) Co., Ltd., China

*CORRESPONDENCE
Zhixiao Yang
linyingxian2006@126.com

RECEIVED 23 September 2024
ACCEPTED 20 January 2025
PUBLISHED 07 February 2025

CITATION
Yu S, Wan J, Xu T, Zhang J, Cao L, Liu J, Liu H, Ren X and Yang Z (2025) A gene expression atlas of *Nicotiana tabacum* across various tissues at transcript resolution. *Front. Plant Sci.* 16:1500654.
doi: 10.3389/fpls.2025.1500654

COPYRIGHT
© 2025 Yu, Wan, Xu, Zhang, Cao, Liu, Liu, Ren and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A gene expression atlas of *Nicotiana tabacum* across various tissues at transcript resolution

Shizhou Yu[1], Jufen Wan[1,2], Tenghang Xu[1,3], Jie Zhang[1], Linggai Cao[1], Jie Liu[1], Hongfeng Liu[4], Xueliang Ren[1] and Zhixiao Yang[1*]

[1]Molecular Genetics Key Laboratory of China Tobacco, Guizhou Academy of Tobacco Science, Guiyang, China, [2]State Key Laboratory for Crop Stress Resistance and High-Efficiency Production, Northwest A&F University, Yangling, China, [3]College of Agronomy and Biotechnology, Hebei Normal University of Science and Technology, Qinhuangdao, China, [4]Guiyang Branch Company of Guizhou Tobacco Company, Guiyang, China

Alternative splicing (AS) expands the transcriptome diversity by selectively splicing exons and introns from pre-mRNAs to generate different protein isoforms. This mechanism is widespread in eukaryotes and plays a crucial role in development, environmental adaptation, and stress resistance. In this study, we collected 599 tobacco RNA-seq datasets from 35 projects. 207,689 transcripts were identified in this study, of which 35,519 were annotated in the reference genome, while 172,170 transcripts were newly annotated. Additionally, tissue-specific analysis revealed 4,585 transcripts that were uniquely expressed in different tissues, highlighting the complexity and specialization of tobacco gene expression. The analysis of AS events (ASEs) across different tissues showed significant variability in the expression levels of ASE-derived transcripts, with some of these transcripts being associated with stress resistance, such as the *geranyl diphosphate synthase* (*GGPPS*). Moreover, we identified 21,763 splicing quantitative trait locus (sQTLs), which were enriched in genes involved in biological processes such as histone acetylation. Furthermore, sQTLs involved genes related to plant hormone signal transduction, terpenoid backbone biosynthesis, and other resistance pathways. These findings not only reveal the diversity of gene expression in tobacco but also provide new insights and strategies for improving tobacco quality and resistance.

KEYWORDS

***Nicotiana tabacum*, ASE, tissue-specific transcript, stress response, SQTL**

Introduction

AS is a crucial mechanism in gene expression regulation, where selective splicing of exons and introns in pre-mRNA generates multiple distinct mRNA isoforms, thereby expanding the coding capacity of the genome. This process is widespread among eukaryotes and plays a key role in human diseases, plant and animal development, and stress responses. Studies have shown that various gene mutations affecting the global regulation of AS or altering AS of specific genes are associated with human brain diseases (Licatalosi and Darnell, 2006; Raj and Blencowe, 2015). Comparative analysis of AS in adipose tissue from different sheep breeds revealed that genes related to ASE are closely related to adipose tissue development (Miao et al., 2023). Temperature changes regulate intron retention and affect starch accumulation in *Arabidopsis* (Seo et al., 2011). Numerous studies indicate that AS plays a significant role in stress responses. Analysis of AS in *Arabidopsis* under normal physiological conditions and various abiotic stress treatments revealed that isoforms generated by the AS play roles in responding to abiotic stress (Filichkin et al., 2010). Significant changes in AS were observed in wheat under stress conditions such as high temperature and drought (Liu et al., 2018). The *OsCYP19-4* gene produces various transcripts through AS, which have different subcellular localizations, and these isoforms play distinct roles in responding to cold stress (Lee et al., 2016).

With the advancement of AS research, it has become evident that genetic variations also play an integral role in regulating AS patterns. sQTL is genetic variation sites associated with ASE that influence the composition of gene expression isoforms by regulating pre-mRNA splicing patterns (Dwivedi et al., 2023). For instance, sQTL in different *Arabidopsis* ecotypes are enriched in genes related to circadian rhythm, flowering, and stress responses (Khokhar et al., 2019), indicating that single nucleotide polymorphisms (SNPs) regulate gene expression by altering transcript isoforms. The identification of sQTL in barley (*Hordeum vulgare*) core accessions under Cd exposure has revealed potential mechanisms for Cd accumulation in plants (Deng et al., 2022). sQTL studies provide new insights into the complexity of gene regulation, uncovering how genetic variation influences splicing mechanisms to regulate gene expression and ultimately affect biological phenotypes.

Tobacco (*Nicotiana tabacum*) is a widely cultivated plant with significant economic and agricultural importance, serving as the primary source of tobacco used in the global cigarette industry (Warner, 2000). Beyond its economic value, tobacco has also become a model plant in research due to its well-characterized genome and ease of genetic manipulation (Ganapathi et al., 2004). However, like many other crops, tobacco is affected by various environmental stresses, including drought, salinity, extreme temperatures, and pathogen attacks, which can severely impact its growth, yield, and quality. Understanding the mechanisms by which tobacco responds to these abiotic and biotic stresses is crucial for developing more resilient varieties.

Currently, research on AS is primarily focused on crops like *Arabidopsis* and rice. However, the regulatory mechanisms and functional significance of AS in tobacco. A deeper understanding of AS mechanisms not only helps to uncover the diversity of gene expression but also provides new insights and strategies for crop improvement. In this study, we performed transcriptomic analysis based on RNA-seq expression profiles of 599 tobacco samples from 13 tissues, including transcript identification, tissue-specific transcripts identification, expression characteristics analysis of transcripts produced by different ASEs, and sQTL identification. These findings greatly enrich the tobacco transcriptome library and offer valuable insights for tobacco breeding.

Materials and methods

Data sources

RNA-seq data was retrieved from 35 BioProjects available in NCBI, covering 599 samples across 13 tissues (Supplementary Table S1). Tobacco genome data was obtained from the Sol Genomics Network database (https://solgenomics.net/), specifically using the genome version published by Edwards et al. in 2017 (Edwards et al., 2017).

Transcript assembly, processing and expression

The SRA Toolkit (https://github.com/ncbi/sra-tools/wiki) was used with default parameters to convert SRA files into FASTQ format. The FASTQ data underwent quality control using fastp (Chen et al., 2018) with default settings, resulting in 599 clean datasets. Post-QC FASTQ files were aligned to the tobacco genome using STAR (v2.7.10a) (Dobin et al., 2013) with default parameters. Transcript assembly and merging were performed with StringTie (v2.1.7) (Pertea et al., 2015) using the '-merge' option on the aligned data. GffCompare (0.11.2) (Pertea and Pertea, 2020) was employed to compare the merged transcript assemblies with the reference genome annotations. Transcript quantification was conducted using RSEM (Li and Dewey, 2011) with default parameters. Transcripts not matched to genes or those aligned in the opposite direction to gene transcription were considered potential non-coding RNAs and excluded from further analysis. Additionally, transcripts with coordinates overlapping two or more genes were identified as fusion transcripts and excluded from subsequent analysis to ensure the accuracy of the results. Gene quantification was carried out using featureCounts (Liao et al., 2014) with default settings. tSNE clustering analysis of transcript expression levels was performed using the R package Rtsne (Krijthe and van der Maaten,

Abbreviations: AS, Alternative splicing; ASEs, AS events; sQTLs, Splicing quantitative trait locus; SNPs, Single nucleotide polymorphisms; TPM, Transcripts Per Million; PSI, The percent spliced in; PCA, Principal component analysis; RI, Intron retention; SE, Exon skipping; A3, Alternative 3'-acceptor; A5, Alternative 5'-donor; AF, Alternative first exon; AL, Alternative last exon.

2015), and GO, KEGG, and PFAM enrichment analyses were conducted using the clusterProfiler v4.0 package (Wu et al., 2021).

Tissue-specific expressed transcripts

Following the methodology of Kryuchkova-Mostacci et al (Kryuchkova-Mostacci and Robinson-Rechavi, 2017), Tau index (τ) was used as a measure of tissue specificity for each transcript (Yanai et al., 2005). In brief, we first calculated the median TPM (Transcripts Per Million) for each transcript across all tissues. Transcripts with a median TPM below 0.1 were considered not expressed (TPM = 0). The corrected TPM values were then log2-transformed using the formula log2(TPM + 1). For each transcript, only those with a sum of log2(TPM + 1) across all tissues greater than 0.1 were retained; all other transcripts were discarded. The filtered results were subsequently used to calculate Tau (García-Pérez et al., 2021). The calculation formula is:

$$\tau = \frac{\sum_{i=1}^{n}\left(1 - \frac{x_i}{\max\limits_{1\le i\le n} x_i}\right)}{n-1}$$

n represents the total number of tissues, and Xi denotes the expression level of transcript X in tissue i. Transcripts with a Tau value greater than 0.85 and a maximum expression level exceeding 5 TPM were considered tissue-specific. Based on the correspondence between transcripts and genes, GO enrichment analysis was performed on the host genes of these tissue-specific transcripts. Host genes are those from which the corresponding transcripts are transcribed.

Correlation between transcripts and host genes

we calculated the Pearson correlation coefficient between the expression levels of transcripts and their host genes. Under the condition that the p-value is less than 0.05, a Pearson correlation coefficient greater than 0.3 was considered indicative of a positive correlation between transcript and gene expression, while a coefficient less than -0.3 was considered indicative of a negative correlation. Otherwise, the expression of the transcript was deemed uncorrelated with that of its host gene (Wu et al., 2023).

AS analysis

SUPPA2 (v2.3) (Trincado et al., 2018) was used with default parameters to identify alternative splicing events (ASEs), followed by filtering of the identified events. Briefly, we first inferred ASEs based on the transcript structure and expression levels of genes using the "–generateEvents" parameter. Then, the Percent Spliced In (PSI) value for each ASE in each sample was calculated using the "–psiPerEvent" parameter. We retained events with a PSI value greater than 0.1 in at least 5% of the samples (n = 14) to generate a set of high-confidence events.

Tobacco leaves SNP calling

This study conducted SNP calling for 375 tobacco leaf samples. The specific steps were as follows: First, BAM files generated from the mapping of tobacco leaf samples to the reference genome were sorted using SAMtools (v1.21) (Li et al., 2009). Next, PCR duplicates were marked using the MarkDuplicates function of Picard (https://github.com/broadinstitute/picard). Subsequently, the SplitNCigarReads function in GATKwas applied to split reads into exon segments (removing Ns while maintaining grouping information) and hard-clip any sequences overhanging into intronic regions (Van Auwera and O'Connor, 2020). SNP calling was then performed with HaplotypeCaller function of GATK. Detected SNPs were filtered using the VariantFiltration tool of GATK, retaining only those with a Fisher Strand (FS) < 30.0 and a Quality by Depth (QD) > 2.0. Further filtering was conducted with SelectVariants tool of GATK to retain only biallelic variants. Finally, VCFtools (v0.1.16) (Danecek et al., 2011) was used to filter the SNP dataset with parameters set to –max-missing 0.9 and –maf 0.05, retaining SNP sites with a missing rate below 10% and a minor allele frequency above 5% in the population.

Identification of sQTL

ASEs were identified using LeafCutter (Li et al., 2018), and the official script prepare_phenotype_table.py was used to clusters the identified introns, filters out introns present in less than 40% of the population or those with minimal variation, and normalizes the PSI values. We employed PEER (v1.0) to infer hidden determinants from the normalized PSI values, selecting the top 10 PEER factors as covariates for sQTL analysis to correct for batch effects and other influencing factors. Additionally, we used the top five PCA results from the prepare_phenotype_table.py output as covariates. Principal component analysis (PCA) was performed using PLINK (v1.90b6.21) (Purcell et al., 2007), and the top 10 principal components were selected to correct for population structure. The PEER factors, PSI PCA results, and VCF PCA results were combined to create a covariate file for sQTL identification. The normalized PSI value for each intron splicing event was used as the phenotype, and sQTLs were identified using FastQTL with the '-normal' parameter. sQTLs located within 100 kb of an ASE were defined as cis-sQTLs. An FDR threshold of <0.05 was applied to identify all significant cis-sQTLs.

Results

Extensive transcriptome diversity and complexity in tobacco

In this study, a total of 207,689 transcripts were identified from 599 transcriptome datasets derived from 13 different tissues. Based on whether these transcripts were annotated in the reference genome, they were classified into newly annotated transcripts and

those already annotated in the reference genome. A total of 172,170 newly annotated transcripts were identified, accounting for 83% of all transcripts. This indicates that the tobacco transcriptome contains a large number of transcripts that were not annotated in the reference genome, demonstrating the complexity and diversity of tobacco gene expression. According to the reference genome's GFF file, 113,553 transcripts overlap with known genes, while 58,617 transcripts are located in intergenic regions. These transcripts were considered potential non-coding RNAs and were excluded from downstream analysis (Figure 1A). In terms of exon count, the number of transcripts decreases as the number of exons increases. Most transcripts have fewer than 20 exons, with transcripts containing 2 exons being the most common, totaling over 50,000. Compared to the annotated transcripts in the reference genome, the newly annotated transcripts include more transcripts with over 20 exons (Figure 1B). Regarding transcript length distribution, the number of transcripts shows a negative correlation with length. The majority of transcripts are shorter than 25 kb, while transcripts longer than 25 kb are mostly newly annotated (Figure 1C). These results suggest that the newly annotated transcripts not only dominate in quantity but also exhibit more complex structural characteristics, such as longer lengths and more exons. The transcriptional intensity across the population varies among different chromosomes. The number of expressed transcripts across the 599 samples typically clusters within a peak range for each chromosome. For example, chromosomes 17 and 19 of tobacco have the highest number of expressed transcripts, ranging between 6,000 to 8,000. Chr02,

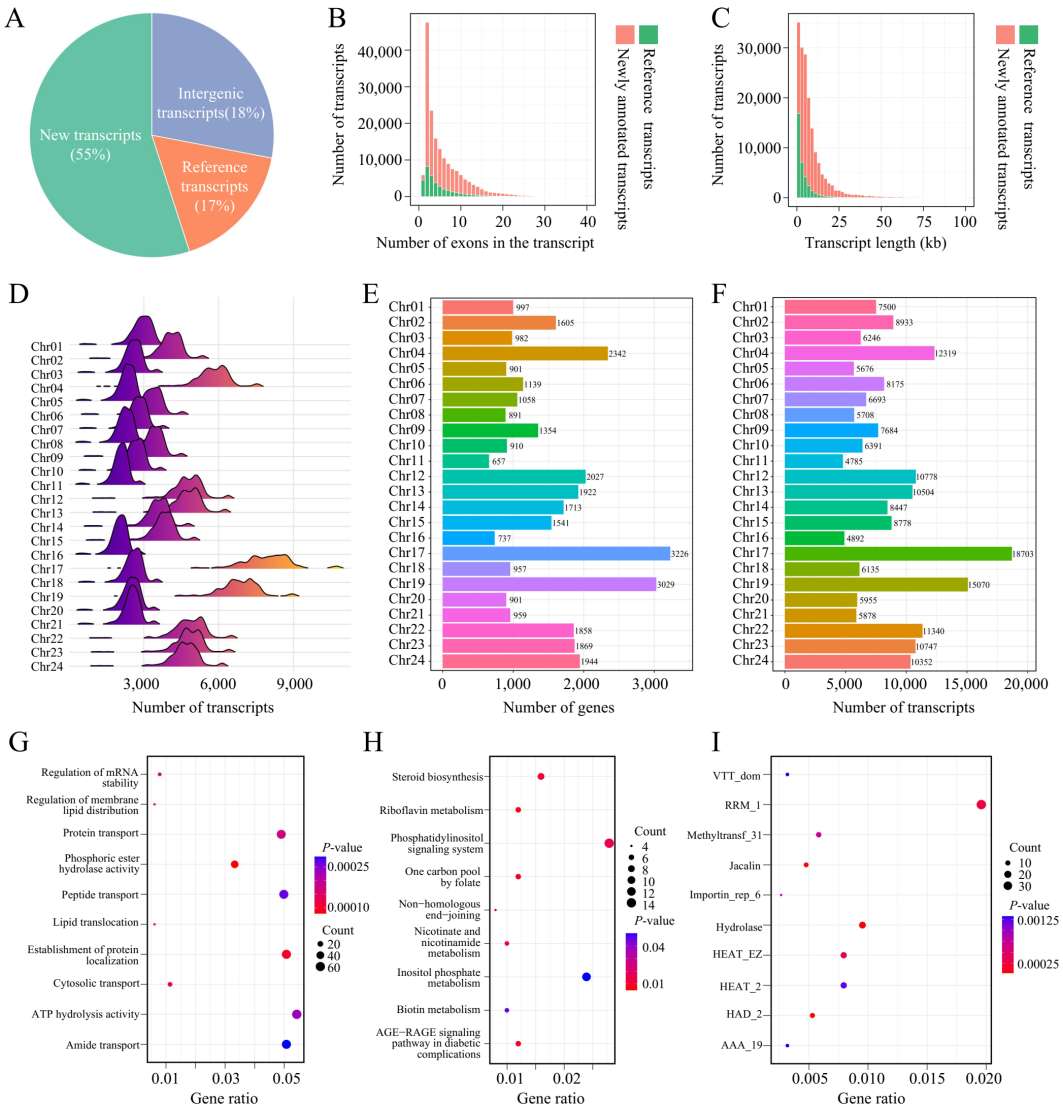

FIGURE 1
Transcriptome profiling and functional analysis in tobacco. **(A)** Transcripts identified in 599 tobacco RNA-seq datasets. "New Transcripts," "Reference Transcripts," and "Intergenic Transcripts" correspond to newly annotated transcripts, transcripts annotated in the reference and transcripts that are located in intergenic regions, respectively. **(B)** Histogram of exon counts per transcript. **(C)** Histogram of transcript length distribution. **(D)** Distribution of expressed transcripts across chromosomes in different samples. **(E)** Gene counts across different chromosomes. **(F)** Gene counts across different chromosomes. **(G)** GO enrichment analysis of the top 5% genes with the highest isoform counts. **(H)** KEGG enrichment analysis of the top 5% genes with the highest isoform counts. **(I)** PFAM enrichment analysis of the top 5% genes with the highest isoform counts.

Chr12, Chr13, Chr14, Chr15, Chr21, Chr22, and Chr23 have a number of expressed transcripts ranging between 3,000 to 6,000, while the remaining chromosomes generally have fewer than 3,000 expressed transcripts (Figure 1D). The differences in the number of expressed transcripts across chromosomes are related to the number of genes annotated on each chromosome. Generally, the more genes annotated on a chromosome, the more transcripts are expressed. However, although chromosome Chr06 has fewer annotated genes than Chr09, Chr06 shows more expressed transcripts than Chr09 (Figures 1E, F), which may be due to differences in transcriptional regulation on Chr06. These findings underscore the complexity of the tobacco transcriptome and provide new insights into the spatial distribution of gene expression across chromosomes.

To explore the relationship between the number of transcript isoforms and gene function, we performed functional enrichment analysis on the top 5% of genes with the highest isoform numbers. The results showed that genes with a large number of transcript isoforms play key roles in important biological processes, such as the regulation of mRNA stability, regulation of membrane lipid distribution, phosphoric ester hydrolase activity, and protein transport (Figure 1G). The transcript diversity of these genes may enable cells to flexibly adjust functions under different environmental conditions, thus adapting to physiological demands. Additionally, KEGG pathway analysis revealed that these genes are enriched in critical pathways such as the phosphatidylinositol signaling system, inositol phosphate metabolism, and steroid biosynthesis (Figure 1H), further suggesting that these genes play essential roles in maintaining cellular metabolic balance and signal regulation. PFAM analysis showed that these genes are enriched in protein families such as RRM_1, Methyltransf_31, and Hydrolase (Figure 1I), indicating that transcript diversity may contribute to gene expression regulation through mechanisms involving RNA processing, metabolic regulation, and material transport. Overall, these findings reveal that genes with a large number of transcript isoforms potentially have important functions in various biological processes, providing new insights into the relationship between gene diversity and gene function.

Transcript-specific expression patterns across different tissues

The distinct functions of plant tissues lead to tissue-specific gene expression patterns (Shi et al., 2021). However, AS can cause certain transcripts to be expressed specifically in certain tissues. To identify the tissue-specific transcripts across 10 tobacco issues, we first excluded non-coding transcripts and conducted tSNE dimensionality reduction clustering on 137,568 transcripts. The results showed significant differences in transcript expression between different tissues. Notably, several tobacco leaf tissues (leaf, blade, midrib, shoot) exhibited low variability, clustering together in the tSNE plot, except for the leaf of seedlings (Figure 2A). Therefore, we combined blade, lamina, and midrib with leaf for tissue-specific analysis. We also compared the distribution of Tau values across tissues (Figure 2B), finding that trichomes showed a peak near 1, indicating a higher number of tissue-specific transcripts in this tissue. Subsequently, we categorized transcripts into four classes based on their expression levels (null, weak, broad, and tissue-specific, Supplementary Table S2). The results showed that more than half (51%) were low-expressed null transcripts (TPM <1 in all tissues; n=70,211), 22.4% were weak transcripts (TPM <5 in all tissues; n=30,868), and 23.1% were broad transcripts (TPM >=5, Tau<0.85; n=31,763), with only 3.5% (4,585) being tissue-specific transcripts (TPM >=5, Tau >=0.85; n=4,585) (Figure 2C). Among these, trichomes had the highest number of tissue-specific transcripts (Figure 2D). Notably, 27.3% (1,251) of these tissue-specific transcripts were derived from the reference genome annotation, while newly annotated transcripts accounted for 72.7% (3,334), highlighting the significance of AS-derived novel transcripts in tissue-specific analysis (Figure 2E). Furthermore, expression clustering analysis of these transcripts revealed similar expression patterns between the transcripts annotated in the reference and newly annotated transcripts (Figure 2F). This study used tSNE clustering to uncover transcript expression differences between different tobacco tissues, identifying numerous tissue-specific transcripts, particularly in trichomes, and demonstrating the importance of novel transcript assembly for a deeper understanding of tissue-specific expression.

Tissue-specific transcripts showed inconsistent expression patterns with tissue-specific genes

To investigate the differences in expression patterns between transcripts and genes, we compared the number of tissue-specific transcripts and genes across different tissues, as well as the expression correlations between transcripts and their corresponding host genes. When comparing tissue-specific genes and transcripts, we found that the number of tissue-specific expressions was similarly distributed across tissues for both genes and transcripts. However, in trichomes, the number of tissue-specific transcripts far exceeded that of genes (Figure 3A). Trichomes are specialized epidermal cells in the aerial parts of plants and are associated with the plants response to biotic and abiotic stresses (Pattanaik et al., 2014). To explore the functions of the tissue-specific transcripts in trichomes, we analyzed the functions (GO terms) of the host genes of these transcripts and found that they are mainly involved in biological processes and cellular components, such as primary metabolic processes, nitrogen compound metabolic processes, and response to stimuli (Figure 3B). Furthermore, GO enrichment analysis revealed that these host genes are significantly enriched in endosome-related genes (Figure 3C). Endosomes are a group of heterogeneous organelles responsible for sorting and delivering internalized materials from the cell surface and transporting materials from the Golgi apparatus to lysosomes or vacuoles, playing crucial roles in plant hormone and defense signaling (Huotari and Helenius, 2011; Contento and Bassham, 2012). This suggests that the abundant transcripts generated by AS in trichomes may be an important mechanism by which plants respond to biotic and abiotic stresses. Enrichment analysis of host genes for tissue-

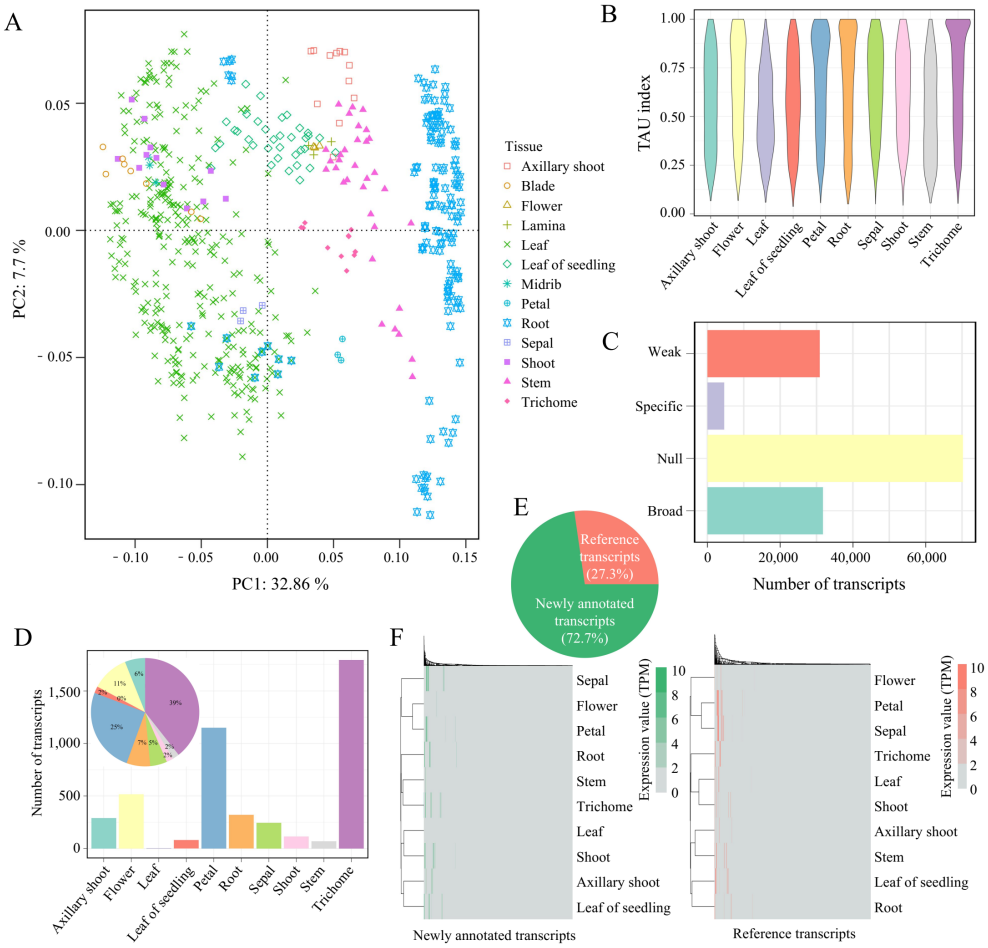

FIGURE 2
Tissue-specific transcripts across different tissues. **(A)** tSNE clustering of transcripts. **(B)** Distribution of Tau for transcripts across different tissues. **(C)** Transcripts number of different transcript types. **(D)** Number and proportion of tissue-specific transcripts in different tissues. **(E)** Proportion of tissue-specific transcripts in transcripts annotated in the reference and newly annotated transcripts. **(F)** Expression clustering of tissue-specific transcripts in transcripts annotated in the reference (reference transcripts) and newly annotated transcripts. Each column represents the expression level of different transcripts in the samples.

specific transcripts across all tissues showed that tissues such as flowers, petals, and roots had the most enrichment results (Supplementary Figure S1, Supplementary Table S3). In the flower tissue, GO terms related to floral organ development were found, indicating that AS is also necessary for producing various transcripts to meet the growth requirements during flower morphogenesis.

Next, we calculated the expression correlations between transcripts and their host genes. A total of 130,567 transcript-gene pairs were analyzed, of which 56.44% (73,694 transcripts) showed a positive correlation with their host genes, while 43.5% (56797) of the transcripts were not correlated with their host genes (Figure 3D). As the number of transcript types per gene increased, the number of transcripts were irrelevant to gene expression also increased (Figure 3E). For instance, among the 54 genes enriched in the endosome GO term, 61.1% (294 transcripts) of the transcripts were not correlated with the expression of their host genes, which is significantly higher than the number of positively correlated transcripts. This suggests that tobacco may generate more transcripts through AS to adapt to external stresses. To explore the functions of these transcripts that are not correlated with their host genes, we conducted enrichment analysis, which revealed that these transcripts are mainly enriched in GO terms such as purine nucleotide binding, protein transport, and intracellular protein transport (Figure 3F). Genes related to purine nucleotide binding are essential for repairing certain DNA damage during plant vegetative growth and play roles in plant development and defense (Islam et al., 2020), while intracellular protein transport is also involved in various stress responses in plants (Kamal et al., 2020; Wolff et al., 2021). These findings further emphasize that transcript-level resolution is higher than gene-level resolution in transcriptome analysis, particularly in the study of stress-related genes.

Analysis of ASE types

To further investigate transcript diversity, we compared the transcripts annotated in the reference genome with newly

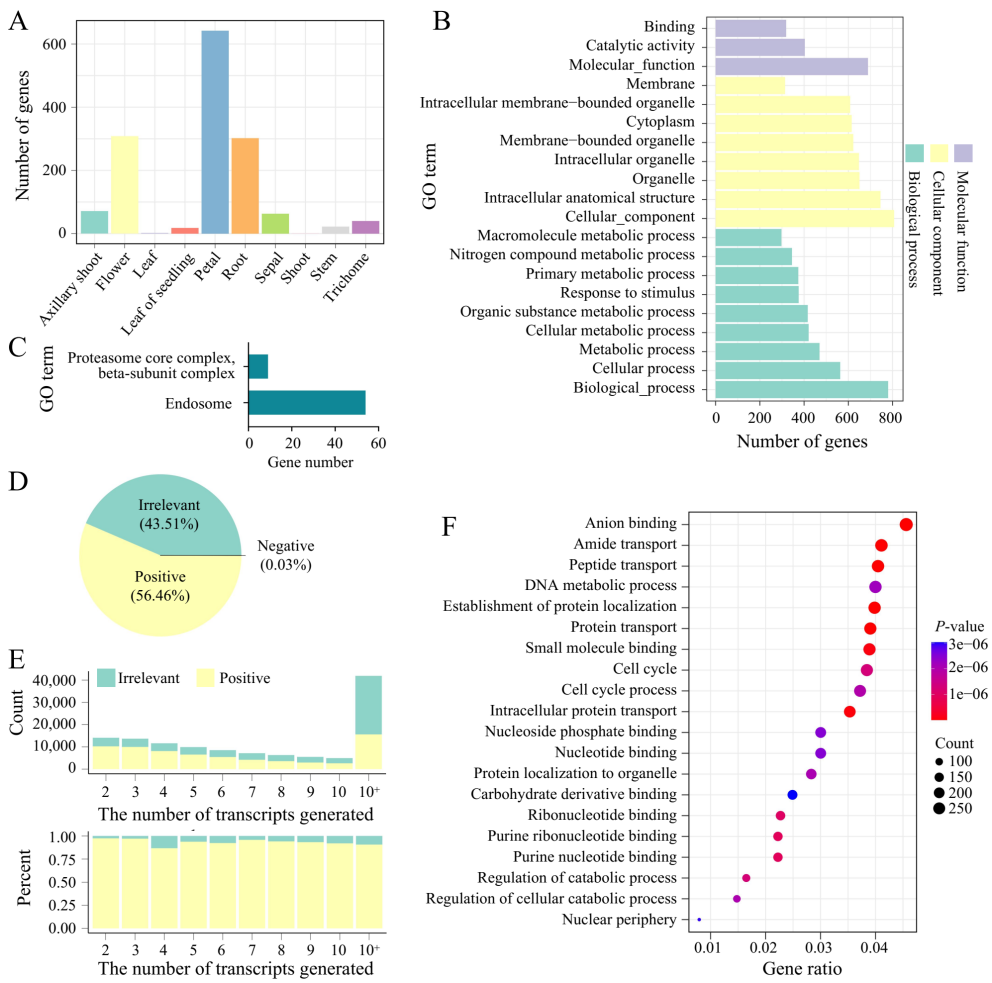

FIGURE 3
Transcript-host gene expression correlation and tissue-specific enrichment analysis. **(A)** The number of tissue-specific genes in different tissues. **(B)** GO enrichment analysis of tissue-specific genes in trichomes. **(C)** GO enrichment analysis of tissue-specific transcripts in trichomes. **(D)** Proportions of transcripts that are positively, negatively, or irrelevantly correlated with their host genes. **(E)** The count (top) and expression ratios (bottom) of transcripts that are positively correlated or irrelevant to their host genes. The "10+" on the x-axis represents genes that generated more than 10 transcripts. **(F)** GO enrichment analysis of transcripts that are irrelevant to the expression of their host genes.

annotated transcripts across different types of ASEs. We quantified levels of intron retention (RI), exon skipping (SE), alternative 3'-acceptor (A3), alternative 5'-donor (A5), alternative first exon (AF), alternative last exon (AL), and mutually exclusive exon (MX) splicing events. In this study, we identified a total of 107,140 ASEs occurring in 17,758 genes. Notably, A3 events accounted for the largest proportion of ASEs in both transcripts annotated in the reference genome and newly annotated transcripts (Figure 4A). Previous studies have reported similar results in tobacco using different tools to predict ASEs. However, it is well-established that RI are the most frequent type of AS event in plants, while SE dominate in humans (Wang and Brendel, 2006; Barbazuk et al., 2008; Filichkin et al., 2010; Zhou et al., 2024). Since the reference genome used in this study has only 64% of the genome anchored to pseudomolecules (Edwards et al., 2017), further identification of ASEs based on a higher-quality genome is required to determine whether this phenomenon is a species-specific characteristic of tobacco. Additionally, we used SUPPA2 to calculate the relative expression levels of different AS types, represented by "percent spliced in" (PSI). The expression levels of newly annotated transcripts in each splicing class did not show significant differences compared to those annotated in the reference genome. However, newly annotated transcripts in AF and AL classes exhibited lower relative expression levels compared to those in the reference genome (Figure 4B). The expression of newly annotated transcripts tended to be more variable, with isoforms produced by AF, AL, and MX showing greater variance among all splicing types (Figure 4C). Consequently, we conducted KEGG gene enrichment analysis on the top 1,000 transcripts with the greatest expression variance within the same ASE, finding that these transcripts were mainly enriched in pathways such as valine, leucine, and isoleucine degradation, terpenoid backbone biosynthesis, and steroid biosynthesis (Figure 4D). Terpenoids are the largest class of plant metabolites, playing crucial roles in essential plant processes such as respiration, photosynthesis, growth, development, reproduction, and environmental

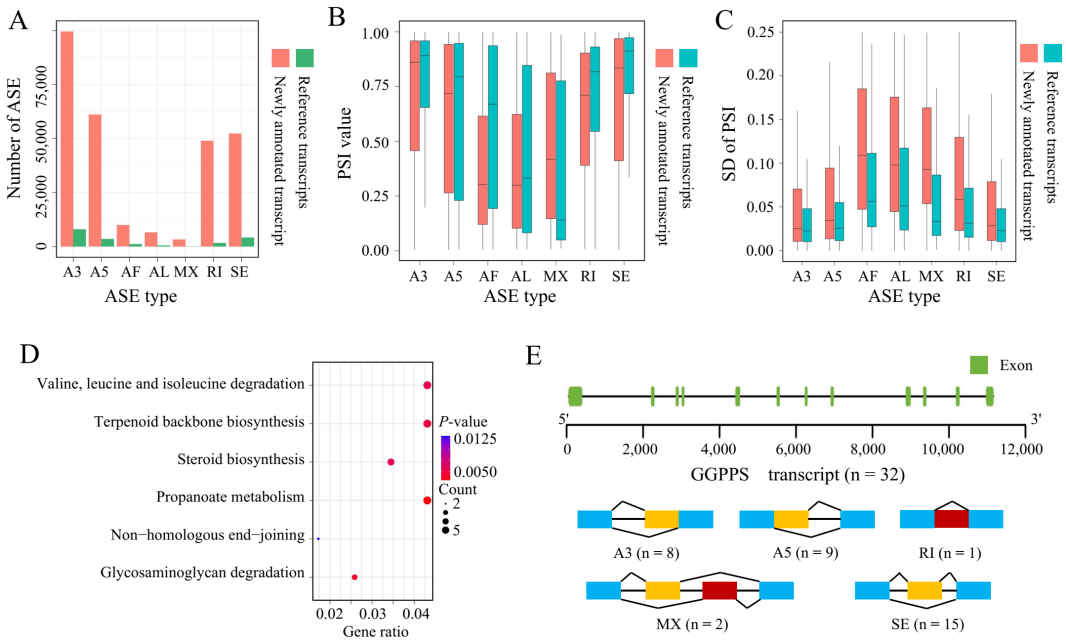

FIGURE 4
Analysis of AS types. **(A)** Comparison of different ASE types between transcripts annotated in the reference genome and newly annotated transcripts. **(B)** Relative expression levels (PSI) of different ASE types in transcripts annotated in the reference genome and newly annotated transcripts. **(C)** Variability in relative expression levels (PSI) of different ASE types between transcripts annotated in the reference genome and newly annotated transcripts. **(D)** KEGG enrichment analysis of the top 1,000 transcripts with the highest variance in PSI. **(E)** Schematic of splicing event types identified in the *GGPPS*.

adaptation (Rodríguez-Concepción and Boronat, 2002; Gershenzon and Kreis, 2018). Our results showed that the host genes of these transcripts include genes involved in the terpenoid backbone biosynthesis pathway, such as geranyl diphosphate synthase (GGPPS) and farnesol kinase (FOLK). We analyzed the number and types of transcripts produced by GGPPS and found that GGPPS generated a total of 32 transcripts, encompassing 5 types of ASEs, indicating a high degree of transcript diversity. Geranyl diphosphate (GPP), synthesized by GGPPS, is the entry point for the synthesis of many monoterpenoid end products, and these monoterpenoids play important roles in plant-insect interactions by directly repelling or deterring herbivorous insects (Yin et al., 2017). Furthermore, we analyzed the proportion of tissue-specific expression among these 1,000 transcripts. We found that the highest number of tissue-specific transcripts were expressed in 10 tissues (blade, lamina, and midrib were combined with leaf), suggesting that the high variability of terpenoid biosynthesis-related transcripts in tobacco during polyploidization may be a key mechanism for responding to environmental changes (Figure 4E, Supplementary Figure S2).

Regulatory impact of sQTLs on functional genes through ASEs

Although this study has revealed the presence of widespread and diverse ASEs in tobacco, the regulatory mechanisms of these ASEs remain to be studied further. To analyze the impact of SNPs on ASEs, we used the PSI of 178,676 ASEs from 375 leaves as the phenotype and 598,293 SNPs as the genotype for sQTL analysis. There were 21,763 SNPs significantly associated with 9,068 ASEs with FDR<0.05 as the significant threshold (Supplementary Table S4. These SNPs significantly associated with ASEs are unevenly distributed on the chromosomes, with certain regions on chr11 and chr12 showing higher densities (Figure 5A). Comparing the distances between these SNPs and ASEs, we found that the number of cis-sVariants is inversely proportional to the distance from the ASE, with 80% of the cis-sVariants located within 10 kb of the ASE (Figure 5B). Most SNPs regulate only 1 to 2 ASEs, but a small number of SNPs regulate more than 10 ASEs. These SNPs may be key regulatory factors, exerting significant influence on gene expression regulation by affecting multiple ASEs simultaneously (Figure 5C). By mapping the ASE coordinates to genes, we identified sVariants that regulate genes through the modulation of ASEs in the sQTL analysis. A total of 3,086 genes were involved in sQTL (Supplementary Table S5), with an average of 2.4 genes regulated by each sVariant. Functional enrichment analysis of these genes revealed that these sGenes are mainly enriched in histone modification-related terms, such as histone acetyltransferase activity, peptide-lysine-N-acetyltransferase activity, and N-acyltransferase activity (Figure 5D). These results suggest that these sQTLs may influence chromatin structure and gene expression through histone acetylation and other modifications, further contributing to epigenetic regulation. Additionally, a large number of sGenes were annotated to resistance-related pathways in the KEGG database, including metabolic pathways (372 genes), plant hormone signal transduction (33 genes), and terpenoid backbone

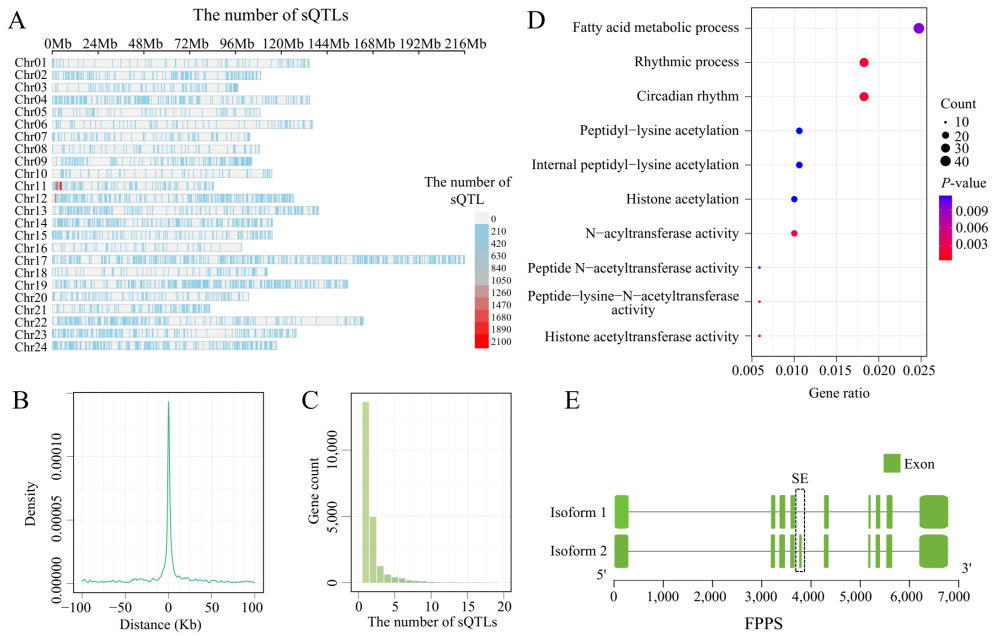

FIGURE 5
The identification of sQTLs in tobacco leaves. **(A)** The density distribution of sQTLs across chromosomes. **(B)** Position of sQTL variants in relation to the splice junction. **(C)** Statistics on the number of ASEs regulated by sQTLs. **(D)** GO enrichment analysis of the genes corresponding to ASEs regulated by sQTLs. **(E)** Schematic representation of the ASE occurring in the *FPPS* (*Nitab4.5_0000368g0370*).

biosynthesis (7 genes). Although these pathways did not reach significant enrichment levels in the enrichment analysis, their associations suggest that these sQTLs may have potential importance in plant resistance responses. For example, in the terpenoid backbone biosynthesis pathway, the gene *farnesyl diphosphate synthase* (*FPPS*, Nitab4.5_*0005478g0020*) has one ASE, and this ASE is significantly regulated by two sQTLs (Figure 5E). FPPS and GGPP are two major isoprenoid intermediates. Similar to GGPP, FPP is mainly synthesized by FPPS, playing an important role in the terpenoid compound pathway, which is crucial in plant secondary metabolism and stress response (Ueoka et al., 2020; Conart et al., 2023). These results suggest that sQTLs may regulate FPPS through ASEs, potentially leading to changes in FPP expression levels, which could directly affect the synthesis and accumulation of terpenoid compounds, thereby influencing plant resistance. These findings not only enrich our understanding of tobacco gene regulatory mechanisms but also provide new insights into the intricate structure of gene regulatory networks.

Discussion

This study analyzed transcript expression in 599 RNA-seq samples from 13 different tobacco tissues, creating the first transcript expression atlas for tobacco. Compared to the 35,519 transcripts annotated from protein-coding genes in the reference genome, this research identified a total of 149,072 transcripts derived from protein-coding genes. Notably, 76% of these are novel transcripts, originating from 70% of the genes, indicating the widespread occurrence of AS in tobacco. While these transcripts may include potential false positives, such as those with non-canonical splice junctions, fragmented transcripts, or redundant transcripts, this study provides a valuable resource for molecular biologists and breeders as a reference for future research and applications. AS is a crucial post-transcriptional regulatory process in which a single gene can produce multiple transcripts, thereby increasing the complexity of the transcriptome, gene regulation, and proteome diversity in multicellular eukaryotes (Mazzucotelli et al., 2008; Chaudhary et al., 2019). In plants, 60-70% of intron-containing genes undergo AS (Pan et al., 2008; Marquez et al., 2012). For example, in the potato genome, an important economic crop, approximately 33,000 genes have been identified (Phytozome v.13), with over 44,000 predicted transcripts, and more than 7,000 genes are known to produce multiple transcripts (Betz et al., 2024). The proportion of multi-transcript genes in the potato genome is much lower than what we observed in tobacco. However, this does not imply that potato genes do not generate a large number of transcripts. Rather, it highlights the effectiveness of transcript expression analysis on large and diverse RNA-seq datasets to uncover the AS patterns of a species like this study.

Polyploid plants often exhibit an increase in genome size, which can lead to enhanced hybrid vigor and improved stress resistance compared to their ancestors. This may influence the size of the transcriptome or the abundance of transcripts (Yang et al., 2014; Han et al., 2016; Song et al., 2020). The novel transcripts annotated in this study could be a result of tetraploid tobacco adapting to environmental changes and enhancing its immune responses. Similar observations have been made in other polyploid plants. For instance, Li et al. reported that in the allotetraploid rapeseed,

there is an increase in the number of splicing isoforms, pre-mRNAs undergoing ASE, pre-mRNAs undergoing APA, and transcription factors (TFs) (Li et al., 2022). Yu et al. also found that polyploidization in wheat led to widespread ASE (Yu et al., 2020). Additionally, some studies suggest that genes may exhibit changes in transcript proportions in response to biotic interactions without altering overall mRNA expression levels (Rigo et al., 2019). Therefore, investigating the tissue-specific transcripts could offer new insights into the evolution, environmental adaptation, and immune responses of tetraploid tobacco. Specifically, our study reveals extensive tissue-specific expression of tobacco transcripts, with notable differences in expression patterns between transcripts and genes. For example, in key functional organs or tissues, trichome-specific transcripts account for 39% of the total, whereas trichome-specific genes constitute only 2.8% of the total. Trichomes are considered the first line of defense in plants, serving as physical and chemical barriers against herbivores and pathogens (Weinhold and Baldwin, 2011). They possess the ability to secrete and store a variety of secondary metabolites, such as terpenoids, phenylpropanoids, and acyl sugars. The presence of trichome-specific promoters and other regulatory elements drives the tissue-specific genes involved in secondary metabolite biosynthesis, leading to the accumulation of these metabolites (Wang et al., 2002). In this study, we identified a large number of trichome-specific transcripts in tobacco, particularly those derived from 92 genes related to secondary metabolites (https://www.genome.jp/pathway/map01110), providing a foundation for further research into the ecological role of trichomes. These results suggest that the changes in AS within specific tissues may be potentially linked to the environmental adaptability of tetraploid tobacco.

AS plays a crucial role in plant adaptation to abiotic stress and environmental constraints (Dubrovina et al., 2013; Staiger and Brown, 2013). For instance, Filichkin et al. discovered that temperature stress can alter transcript abundance and splicing patterns, thereby affecting both transcript and gene expression (Filichkin et al., 2010). In our study, the top 1,000 transcripts with the most significant changes in expression were enriched in pathways related to stress response and metabolism, such as the Terpenoid backbone biosynthesis pathway. Among these, the transcripts produced by the tobacco GGPPS gene exhibited particularly high variability in expression. GGPPS, located in plastids, can interact with GPPS to modify product specificity, thereby efficiently producing GPP. To prevent this process from becoming uncontrolled, plants typically regulate it at the transcriptional level (Tholl et al., 2004; Pateraki and Kanellis, 2008; Wang and Dixon, 2009). Our findings show that GGPPS generates a total of 32 transcripts through 35 ASEs across 5 different splicing types, suggesting that the generation of transcript diversity and differential expression via ASEs may be a key mechanism for the regulation of the GGPPS gene in tobacco.

AS is regulated through both co-transcriptional and post-transcriptional mechanisms, with sQTL explaining part of the variation in transcript isoform ratios (Jia et al., 2020; Li et al., 2020; Wilkinson et al., 2020). In this study, 3,086 sQTL-regulated genes were identified in tobacco leaves, with significant enrichment in terms such as histone acetyltransferase activity, reflecting the complexity of regulatory mechanisms. Previous research revealed a dynamic interplay between DNA methylation, histone modifications, and AS. For example, pre-messenger RNA can precisely regulate AS by recruiting histone-modifying enzymes (Zhou et al., 2014). Histone acetyltransferases (HATs), as key epigenetic regulators, transfer acetyl groups to lysine residues on histones, leading to histone acetylation—a process directly linked to chromatin remodeling and the regulation of gene transcription (Roth et al., 2001). Although we cannot yet assess the direct impact of sQTLs on histone acetylation, exploring the potential connections between these sQTL-regulated genes and histone acetylation could advance our understanding of the regulatory mechanisms underlying gene expression in tobacco.

We also identified 7 sQTL-regulated genes within the terpenoid backbone biosynthesis pathway, which undergo 9 ASEs and are influenced by 19 sQTLs. Terpenoid biosynthesis plays a crucial role in plant secondary metabolism, with many secondary metabolites involved in the plants response to biotic and abiotic stresses (Nishida, 2014). The synthesis of these metabolites is regulated at multiple molecular levels, including AS. For instance, in rice, the WRKY transcription factor regulates diterpenoid biosynthesis through AS, contributing to the plant defense response (Liu et al., 2016). The research of AS in relation to these secondary metabolites enhances our understanding of its regulatory role in plant metabolism and offers insights for crop improvement. In our study, we found that the FPPS, involved in tobacco terpenoid synthesis, contains 1 ASE that is significantly regulated by two sQTLs. This suggests that sQTLs may play a role in regulating the synthesis and accumulation of terpenoid compounds. Overall, this research provides a comprehensive view of the transcriptomic features of polyploid tobacco. Notably, the high transcript diversity in tobacco may significantly impact the production of secondary metabolites and the response to biotic and abiotic stresses of plants, offering new insights into the transcriptional regulation of tobacco.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author/s.

Author contributions

SY: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. JW: Data curation, Visualization, Writing – original draft. TX: Data curation, Visualization, Writing – review & editing. JZ: Data curation, Visualization, Writing – original draft. LC: Data curation, Supervision, Writing – original draft. JL: Data curation, Supervision, Writing – original draft. HL: Data curation, Visualization, Writing – original draft. XR: Supervision, Writing – review & editing. ZY: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Guizhou Provincial Basic Research Program (Natural Science) ((2024) 648), the Program of China National Tobacco Corporation (110202101032(JY-09), 110202201003(JY-03)), and the Program of Guizhou Branch of China National Tobacco Corporation (2023XM02, 2024XM01).

Conflict of interest

Authors HL was/were employed by Guiyang Branch Company of Guizhou Tobacco Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1500654/full#supplementary-material

SUPPLEMENTARY FIGURE 1
GO enrichment results of tissue-specific transcripts.

SUPPLEMENTARY FIGURE 2
Transcripts of GGPPS.

References

Barbazuk, W. B., Fu, Y., and McGinnis, K. M. (2008). Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Res.* 18, 1381–1392. doi: 10.1101/gr.053678.106

Betz, R., Heidt, S., Figueira-Galán, D., Hartmann, M., Langner, T., and Requena, N. (2024). Alternative splicing regulation in plants by SP7-like effectors from symbiotic arbuscular mycorrhizal fungi. *Nat. Commun.* 15, 7107. doi: 10.1038/s41467-024-51512-5

Chaudhary, S., Jabre, I., Reddy, A. S. N., Staiger, D., and Syed, N. H. (2019). Perspective on alternative splicing and proteome complexity in plants. *Trends Plant Sci.* 24, 496–506. doi: 10.1016/j.tplants.2019.02.006

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Conart, C., Bomzan, D. P., Huang, X.-Q., Bassard, J.-E., Paramita, S. N., Saint-Marcoux, D., et al. (2023). A cytosolic bifunctional geranyl/farnesyl diphosphate synthase provides MVA-derived GPP for geraniol biosynthesis in rose flowers. *Proc. Natl. Acad. Sci.* 120, e2221440120. doi: 10.1073/pnas.2221440120

Contento, A. L., and Bassham, D. C. (2012). Structure and function of endosomes in plant cells. *J. Cell Sci.* 125, 3511–3518. doi: 10.1242/jcs.093559

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Deng, P., Yan, T., Ji, W., Zhang, G., Wu, L., and Wu, D. (2022). Population-level transcriptomes reveal gene expression and splicing underlying cadmium accumulation in barley. *Plant J.* 112, 847–859. doi: 10.1111/tpj.15986

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dubrovina, A. S., Kiselev, K. V., and Zhuravlev, Y. N. (2013). The role of canonical and noncanonical pre-mRNA splicing in plant stress responses. *BioMed. Res. Int.* 2013, 264314. doi: 10.1155/2013/264314

Dwivedi, S. L., Quiroz, L. F., Reddy, A. S. N., Spillane, C., and Ortiz, R. (2023). Alternative splicing variation: accessing and exploiting in crop improvement programs. *Int. J. Mol. Sci.* 24, 15205. doi: 10.3390/ijms242015205

Edwards, K. D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A. D., Bombarely, A., et al. (2017). A reference genome for Nicotiana tabacum enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18, 448. doi: 10.1186/s12864-017-3791-6

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109

Ganapathi, T. R., Suprasanna, P., Rao, P. S., and Bapat, V. A. (2004). Tobacco (Nicotiana tabacum L.)-A model system for tissue culture interventions and genetic engineering. *Indian J. Biotechnol.* 3, 171–184.

García-Pérez, R., Esteller-Cucala, P., Mas, G., Lobón, I., Di Carlo, V., Riera, M., et al. (2021). Epigenomic profiling of primate lymphoblastoid cell lines reveals the evolutionary patterns of epigenetic activities in gene regulatory architectures. *Nat. Commun.* 12, 3116. doi: 10.1038/s41467-021-23397-1

Gershenzon, J., and Kreis, W. (2018). Biochemistry of terpenoids: monoterpenes, sesquiterpenes, diterpenes, sterols, cardiac glycosides and steroid saponins. *Annu. Plant Rev. Online* 2, 218–294. doi: 10.1002/9781119312994.apr0016

Han, C., Zhang, P., Ryan, P. R., Rathjen, T. M., Yan, Z., and Delhaize, E. (2016). Introgression of genes from bread wheat enhances the aluminium tolerance of durum wheat. *Theor. Appl. Genet.* 129, 729–739. doi: 10.1007/s00122-015-2661-3

Huotari, J., and Helenius, A. (2011). Endosome maturation. *EMBO J.* 30, 3481–3500. doi: 10.1038/emboj.2011.286

Islam, M. S., Coronejo, S., and Subudhi, P. K. (2020). Whole-genome sequencing reveals uniqueness of black-hulled and straw-hulled weedy rice genomes. *Theor. Appl. Genet.* 133, 2461–2475. doi: 10.1007/s00122-020-03611-2

Jia, J., Long, Y., Zhang, H., Li, Z., Liu, Z., Zhao, Y., et al. (2020). Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat. Plants* 6, 780–788. doi: 10.1038/s41477-020-0688-1

Kamal, M. M., Ishikawa, S., Takahashi, F., Suzuki, K., Kamo, M., Umezawa, T., et al. (2020). Large-scale phosphoproteomic study of arabidopsis membrane proteins reveals early signaling events in response to cold. *Int. J. Mol. Sci.* 21, 8631. doi: 10.3390/ijms21228631

Khokhar, W., Hassan, M. A., Reddy, A. S. N., Chaudhary, S., Jabre, I., Byrne, L. J., et al. (2019). Genome-wide identification of splicing quantitative trait loci (sQTLs) in diverse ecotypes of arabidopsis thaliana. *Front. Plant Sci.* 10, 1160. doi: 10.3389/fpls.2019.01160

Krijthe, J. H., and van der Maaten, L. (2015). Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation. *R Packag. version 0.13*. Available at: https://github. com/jkrijthe/Rtsne.

Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* 18, 205–214. doi: 10.1093/bib/bbw008

Lee, A., Lee, S. S., Jung, W. Y., Park, H. J., Lim, B. R., Kim, H.-S., et al. (2016). The osCYP19-4 gene is expressed as multiple alternatively spliced transcripts encoding isoforms with distinct cellular localizations and PPIase activities under cold stress. *Int. J. Mol. Sci.* 17, 1154. doi: 10.3390/ijms17071154

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323. doi: 10.1186/1471-2105-12-323

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, M., Hu, M., Xiao, Y., Wu, X., and Wang, J. (2022). The activation of gene expression and alternative splicing in the formation and evolution of allopolyploid Brassica napus. *Hortic. Res.* 9, uhab075. doi: 10.1093/hr/uhab075

Li, Y. I., Knowles, D. A., Humphrey, J., Barbeira, A. N., Dickinson, S. P., Im, H. K., et al. (2018). Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* 50, 151–158. doi: 10.1038/s41588-017-0004-9

Li, S., Wang, Y., Zhao, Y., Zhao, X., Chen, X., and Gong, Z. (2020). Global co-transcriptional splicing in arabidopsis and the correlation with splicing regulation in mature RNAs. *Mol. Plant* 13, 266–277. doi: 10.1016/j.molp.2019.11.003

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Licatalosi, D. D., and Darnell, R. B. (2006). Splicing regulation in neurologic disease. *Neuron* 52, 93–101. doi: 10.1016/j.neuron.2006.09.017

Liu, J., Chen, X., Liang, X., Zhou, X., Yang, F., Liu, J., et al. (2016). Alternative splicing of rice WRKY62 and WRKY76 transcription factor genes in pathogen defense. *Plant Physiol.* 171, 1427–1442. doi: 10.1104/pp.15.01921

Liu, Z., Qin, J., Tian, X., Xu, S., Wang, Y., Li, H., et al. (2018). Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (Triticum aestivum L.). *Plant Biotechnol. J.* 16, 714–726. doi: 10.1111/pbi.12822

Marquez, Y., Brown, J. W. S., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* 22, 1184–1195. doi: 10.1101/gr.134106.111

Mazzucotelli, E., Mastrangelo, A. M., Crosatti, C., Guerra, D., Stanca, A. M., and Cattivelli, L. (2008). Abiotic stress response in plants: When post-transcriptional and post-translational regulations control transcription. *Plant Sci.* 174, 420–431. doi: 10.1016/j.plantsci.2008.02.005

Miao, X., Luo, Q., Zhao, H., and Qin, X. (2023). Comparison of alternative splicing (AS) events in adipose tissue of polled dorset versus small tail han sheep. *Heliyon* 9. doi: 10.1016/j.heliyon.2023.e14938

Nishida, R. (2014). Chemical ecology of insect–plant interactions: ecological significance of plant secondary metabolites. *Biosci. Biotechnol. Biochem.* 78, 1–13. doi: 10.1080/09168451.2014.877836

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415. doi: 10.1038/ng.259

Pateraki, I., and Kanellis, A. K. (2008). Isolation and functional analysis of two Cistus creticus cDNAs encoding geranylgeranyl diphosphate synthase. *Phytochemistry* 69, 1641–1652. doi: 10.1016/j.phytochem.2008.02.005

Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. *Front. Plant Sci.* 5, 259. doi: 10.3389/fpls.2014.00259

Pertea, G., and Pertea, M. (2020). GFF utilities: gffRead and gffCompare. *F1000Research* 9, 1–19. doi: 10.12688/f1000research.23297.2

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Raj, B., and Blencowe, B. J. (2015). Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron* 87, 14–27. doi: 10.1016/j.neuron.2015.05.004

Rigo, R., Bazin, J., Crespi, M., and Charon, C. (2019). Alternative splicing in the regulation of plant–microbe interactions. *Plant Cell Physiol.* 60, 1906–1916. doi: 10.1093/pcp/pcz086

Rodríguez-Concepción, M., and Boronat, A. (2002). Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol.* 130, 1079–1089. doi: 10.1104/pp.007138

Roth, S. Y., Denu, J. M., and Allis, C. D. (2001). Histone acetyltransferases. *Annu. Rev. Biochem.* 70, 81–120. doi: 10.1146/annurev.biochem.70.1.81

Seo, P. J., Kim, M. J., Ryu, J.-Y., Jeong, E.-Y., and Park, C.-M. (2011). Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nat. Commun.* 2, 303. doi: 10.1038/ncomms1303

Shi, D., Jouannet, V., Agustí, J., Kaul, V., Levitsky, V., Sanchez, P., et al. (2021). Tissue-specific transcriptome profiling of the Arabidopsis inflorescence stem reveals local cellular signatures. *Plant Cell* 33, 200–223. doi: 10.1093/plcell/koaa019

Song, Q., Ando, A., Jiang, N., Ikeda, Y., and Chen, Z. J. (2020). Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. *Genome Biol.* 21, 178. doi: 10.1186/s13059-020-02094-0

Staiger, D., and Brown, J. W. S. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803

Tholl, D., Kish, C. M., Orlova, I., Sherman, D., Gershenzon, J., Pichersky, E., et al. (2004). Formation of Monoterpenes in Antirrhinum majus and Clarkia breweri Flowers Involves Heterodimeric Geranyl Diphosphate Synthases. *Plant Cell* 16, 977–992. doi: 10.1105/tpc.020156

Trincado, J. L., Entizne, J. C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D. J., et al. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* 19, 40. doi: 10.1186/s13059-018-1417-1

Ueoka, H., Sasaki, K., Miyawaki, T., Ichino, T., Tatsumi, K., Suzuki, S., et al. (2020). A cytosol-localized geranyl diphosphate synthase from lithospermum erythrorhizon and its molecular evolution1 [OPEN. *Plant Physiol.* 182, 1933–1945. doi: 10.1104/pp.19.00999

Van der Auwera, G. A., and O'Connor, B. D. (2020). Genomics in the cloud: using Docker, Gatk, and WDL in Terra. Sebastopol, CA: O'Reilly Media, Inc.

Wang, B.-B., and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci.* 103, 7175–7180. doi: 10.1073/pnas.0602039103

Wang, G., and Dixon, R. A. (2009). Heterodimeric geranyl(geranyl)diphosphate synthase from hop (Humulus lupulus) and the evolution of monoterpene biosynthesis. *Proc. Natl. Acad. Sci.* 106, 9914–9919. doi: 10.1073/pnas.0904069106

Wang, E., Gan, S., and Wagner, G. J. (2002). Isolation and characterization of the CYP71D16 trichome-specific promoter from Nicotiana tabacum L. *J. Exp. Bot.* 53, 1891–1897. doi: 10.1093/jxb/erf054

Warner, K. E. (2000). The economics of tobacco: myths and realities. *Tob. Control* 9, 78–89. doi: 10.1136/tc.9.1.78

Weinhold, A., and Baldwin, I. T. (2011). Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proc. Natl. Acad. Sci.* 108, 7855–7859. doi: 10.1073/pnas.1101306108

Wilkinson, M. E., Charenton, C., and Nagai, K. (2020). RNA splicing by the spliceosome. *Annu. Rev. Biochem.* 89, 359–388. doi: 10.1146/annurev-biochem-091719-064225

Wolff, H., Jakoby, M., Stephan, L., Koebke, E., and Hülskamp, M. (2021). Heat stress-dependent association of membrane trafficking proteins with mRNPs is selective. *Front. Plant Sci.* 12, 670499. doi: 10.3389/fpls.2021.670499

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2, 100141. doi: 10.1016/j.xinn.2021.100141

Wu, H., Wang, J., Hu, X., Zhuang, C., Zhou, J., Wu, P., et al. (2023). Comprehensive transcript-level analysis reveals transcriptional reprogramming during the progression of Alzheimer's disease. *Front. Aging Neurosci.* 15, 1191680. doi: 10.3389/fnagi.2023.1191680

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21, 650–659. doi: 10.1093/bioinformatics/bti042

Yang, C., Zhao, L., Zhang, H., Yang, Z., Wang, H., Wen, S., et al. (2014). Evolution of physiological responses to salt stress in hexaploid wheat. *Proc. Natl. Acad. Sci.* 111, 11882–11887. doi: 10.1073/pnas.1412839111

Yin, J.-L., Wong, W.-S., Jang, I.-C., and Chua, N.-H. (2017). Co-expression of peppermint geranyl diphosphate synthase small subunit enhances monoterpene production in transgenic tobacco plants. *New Phytol.* 213, 1133–1144. doi: 10.1111/nph.14280

Yu, K., Feng, M., Yang, G., Sun, L., Qin, Z., Cao, J., et al. (2020). Changes in alternative splicing in response to domestication and polyploidization in wheat. *Plant Physiol.* 184, 1955–1968. doi: 10.1104/pp.20.00773

Zhou, H.-L., Luo, G., Wise, J. A., and Lou, H. (2014). Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res.* 42, 701–713. doi: 10.1093/nar/gkt875

Zhou, Y., Rashad, S., and Niizuma, K. (2024). Transcriptome-wide alternative mRNA splicing analysis reveals post-transcriptional regulation of neuronal differentiation. *bioRxiv 2024.07.16.603656*. doi: 10.1101/2024.07.16.603656