Article

Time-Series Gene Expression Analysis of Terpene Synthase Gene Family Reveals Its Evolutionary Dynamics in a Tobacco Plant

Zhijun Tong , Ke Zhang , Dunhuang Fang , Sanjie Jiang , Haiyan Li , Shancen Zhao , Chunqiong Wang [*] , Bingguang Xiao [*]

Article

Time-Series Gene Expression Analysis of Terpene Synthase Gene Family Reveals Its Evolutionary Dynamics in a Tobacco Plant

Zhijun Tong [1],[#], Ke Zhang [2],[#], Dunhuang Fang [1], Sanjie Jiang [3], Haiyan, Li [2], Shancen Zhao [3], Chunqiong Wang [2],[*] and Bingguang Xiao [1],[*]

[1] Key Laboratory of Tobacco Biotechnological Breeding, National Tobacco Genetic Engineering Research Center, Yunnan Academy of Tobacco Agricultural Sciences, Kunming 650021, China; Zhijun Tong, tzj861@163.com; Ke Zhang, swukirk@126.com; Dunhuang Fang, sxh861@163.com

[2] Yunnan Tobacco Quality Inspection & Supervision Station, Kunming 650106, China; Haiyan Li, lihaiyan19750503@163.com

[3] BGI-Shenzhen, Shenzhen 518083, China; Sanjie Jiang, jiangsanjie@genomics.cn; Shancen Zhao, zhaoshancen@genomics.cn

[*] Correspondence: Chunqiong Wang, wangchunqiong.ytqs@outlook.com; Bingguang Xiao, frankxiao658@163.com

[#] contributed equally.

Abstract: Several specialized terpenoids that are used as phytoalexins are produced in the plant of tobacco. They play a key role in plant defense against pests and pathogens. Here, we performed genome-wide identification, classification, and expression profiling of the terpene synthase (TPS) genes in the tobacco genome. A total of 186 TPS genes were identified, which is significantly higher than that in wild *Nicotiana*, *Capsicum* and *Solanum* species. Phylogenetic analysis of these TPS genes in tobacco and *Solanaceous* relatives as well as *Arabidopsis* showed that a few tobacco-exclusive TPS clades occurred recently through gene duplication events. These clades are likely to be responsible for a range of compounds that separate *N. tabacum* cultivars from its wild relatives. Transcriptome sequencing of two cultivars of *N. tabacum* at several developmental stages revealed that some TPS genes from the TPS-a and TPS-b lineages are constitutively expressed and a few duplicated TPS genes are only expressed post anthesis at the time of harvest. Additionally, some TPS genes underwent alternative splicing, further diversifying the gene products of the TPS genes. Our findings provide an important candidate genes for further understanding of the diversity and evolution as well as functionality of TPS genes in the plant of tobacco.

Keywords: TPS; tobacco; RNAseq; gene expression analysis; gene duplication

Introduction

Tobacco, *Nicotiana tabacum*, is native to Mesoamerica and South America and is used worldwide for smoking. *N. tabacum* contains a wide range of terpenoids that act as herbivore deterrents *in situ*, and several of these terpenoids contribute to the complex flavor and odor of tobacco smoke. Terpenoids are the largest group of plant natural products [1], and for all mono-, sesqui- and diterpenoids, the first specific enzymatic step is performed by a terpene synthase (TPS).

The diverse structures of terpenoids have made this class of specialized metabolites extremely versatile. Some terpenoids have key cellular functions (e.g.as part of the electron transport chain, membrane components, and serve as hormones). However, the majority serve as specialized metabolites, which influence the fitness of the plant [2]. Plant terpenoids are known to serve as insect and herbivore deterrents, allochemicals, and as attractants for pollinators and beneficial insects in tritrophic interactions [3]. Scented terpenoids in essential oils exist in the majority of plant lineages,

where they contribute to the flavour and fragrance of e.g. cinnamon and cloves, the scents of eucalyptus and ginger, and the odours of tobacco and cannabis [4].

Derived from the general terpenoid biosynthesis, one or two molecules of isopentenyl diphosphate (IPP) and one dimethylallyl diphosphate (DMAPP) are condensed by geranyl diphosphate synthase (GPPS) or farnesyl diphosphate synthase (FPPS) into geranyl diphosphate (GPP) and farnesyl diphosphate (FPP, farnesyl pyrophosphate), respectively [5,6]. Geranyl diphosphate is the precursor for C-10 monoterpenoids and farnesyl diphosphate is the precursor for C15 sesquiterpenoids. Following this the terpene synthase (TPS) will utilize GPPS or FPPS as part of the first enzyme in the specific terpenoid biosynthesis [7]. The TPS's has been extensively studied due to the highlighted importance of terpenes in plants. The TPS gene family has been identified, annotated and characterized in numerous plant species [8]. TPS class I and II are mainly monofunctional; however, the diterpene synthases in bryophytes are bifunctional, while both class I and II of TPS genes are functioning in one enzyme [9,10]. TPSs in the plant genome are divided into 6 subfamilies, TPS-a to TPS-f, which correspond to their functionality [11].

Here, we identified and characterized the TPS gene family in *N. tabacum*. To understand the evolution of the TPS gene family in context, we further constructed a phylogeny of tobacco TPS genes alongside TPS from other solanaceous relatives and Arabidopsis. Transcriptome analysis of two cultivars at three distinct growth stages was performed.

Materials and methods

Sample collection and transcriptome sequencing

The tobacco K326 reference genome and gene set (version No.: gca_000715075.1, https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_000715075.1/) were obtained from NCBI. Based on the importance of developmental stages related to TPS synthesize, fresh tobacco leaf samples were collected at three developmental stages: (a) flowering stage, (b) seven days before harvest and baking, and (c) harvest and baking. Three biological replicates were prepared for each stage. Total RNA were extracted by Trizol method and were subsequentially sequenced on MGI-seq platform at BGI-Shenzhen Co. Ltd, Shenzhen, China, following the manufacturer's standard protocol.

TPS identification and classification analysis

The known TPS protein sequences from previous studies were retrieved and aligned to the tobacco K326 reference genome using tblastn, and then the complete protein structures of the aligned regions along with 2 kb flanking regions were identified by Genewise [12]. The overlapping gene set with the tobacco genome reference was removed. The remaining genes were considered to be newly identified *TPS* genes. Finally, by integrating the tobacco reference gene set and these possible new TPS genes, the domains of protein-coding genes were identified based on the results of Pfam and InterproScan. The genes containing PF01397 and/or PF03936 domains were defined as candidate TPS genes. To classify the candidate *TPS* genes, we compared them with the above known *TPS* genes from the references and kept the optimal ratio pair result. MEME (version 5.3.3) was used to analyze the motif characteristics of the tobacco TPS domain [13]. The same pipeline of analysis were performed for other species used in this study.

Evolutionary analysis of TPS

The candidate TPS protein sequences from different species were used for global alignment. Here, MuSiC (version 3.8.31) was used for multiple sequence alignment analysis [14], and then Gblocks (version 0.91b) was used to screen the multiple sequence alignment results and to retain the conservative sites [15]. Sequence regions that cannot be properly aligned such as large deletion/insertion were removed. Finally, approximate maximum likelihood phylogenetic trees were constructed using FastTree (version 2.1.10) based those conserved loci. iTOL (https://itol.embl.de) was used for visual display [16].

Gene expression analysis

The clean data of the transcriptome were aligned to the tobacco reference genome using HISAT (version 2.0) [17]. Then, based on BAM and GFF files, the reads were counted using featureCounts (version 1.6.5) [18]. The gene expression amount in FPKM level was calculated using the customized Perl script . Gene isoforms were assembled using StringTie (version 2.1.5) [19]. Deseq2 (version 1.6.5) of the R platform was used to calculate the differentially expressed genes [20]. Genes with at least twofold differential expression and an FDR < 0.05 were considered as differentially expressed genes. In order to study the alternative splicing of genes, we used rMATS (version 4.1.0) for differential alternative splicing analysis [21]. FDR < 0.05 was considered to be differential alternative splicing at a significant level.

Results

Genome-wide identification of TPS genes in tobacco

To understand the critical role of *TPS* genes during the process of terpene production, we identified the whole set of *TPS* genes in the tobacco species *Nicotiana tabacum* using homolog-based method. TBLASTN was used to map known TPS protein sequences against the reference genome of *N. tabacdum*. Retrieval within the *N. tabacum* genome database was iterated for each matched sequence until no new sequences were found. After filtering out incomplete and redundant matched sequences, 186 *TPS* genes were identified in *Nicotiana tabacum* (Supplementary File 1 and 2). The number of *TPS* genes was significantly higher than that in its known Solanaceae relatives (59 ~ 84 TPS genes) and Arabidopsis (46 TPS genes) and much higher than that in the simple moss *Physcomitrium patens* (1 TPS gene)[9]. This suggests that *N. tabacum* has undergone substantial TPS gene expansion through multiple gene duplication events. Human selection for specific fragrances and flavours in tabacco has been going on for over thousands of years, the complex terpene profile revealed by our analysis may reflect the genomic basis of tabacco breeding. We further discovered that many TPS genes are in tandem duplication along chromosome, suggesting that tandem duplications events may play a role in the expansion of TPS genes in tobacco.

Evolutionary analysis of TPS genes in Solanaceae

To further investigate the evolutionary relationships within the *TPS* gene family, we constructed a NJ tree using *TPS* genes identified in *N. tabacum*, wild tobacco relative *N. attenuata*, chili pepper (*Capsicum annuum*), tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*) and the model plant Arabidopsis (Figure 1, Supplementary Figure S1). A total of six clades, including TPS-a, TPS-b, TPS-c, TPS-e, TPS-f, and TPS-g were discovered. The clades TPS-a and b align well with the overall molecular phylogeny of Solanaceae while the clade of *Nicotiana* is more basal to *Capsicum* and *Solanum* [11,22]. For the clades c-g, the Solanaceae genes were interspersed, suggesting a higher homology among these clades of the Solanaceae *TPS* gene family (Figures 1 and 2). The few conserved elements of TPS needed for functionality means that the evolutionary relationship over time and in families has a greater impact on the sequence than functionality (within the a, b, c clades). Our analysis also suggested *TPS* genes of Arabidopsis are the evolutionary youngest and have little overlap with Solanaceae clades.
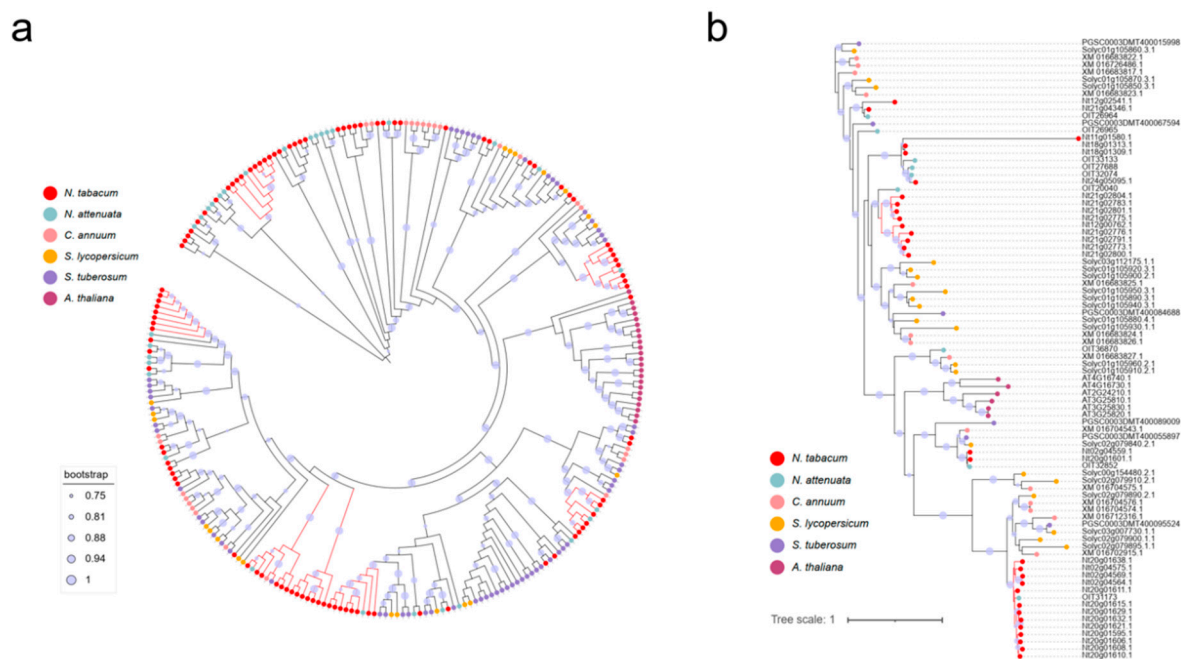
Figure 1. (a) Phylogeny of *TPS* genes identified in *N. tabacum, N. attenuata, Capsicum annuum, Solanum lycopersicum, Solanum tuberosum* and Arabidopsis. TPS-a, TPS-b, TPS-c, TPS-e, TPS-f, and TPS-g were found in all plants. **(b)** Detailed phylogeny of TPS subgroups, including TPS-a and TPS-b.

Analysis of recently occurring duplicated genes

Tandemly duplicated genes (TDGs) result from gene duplication events produce identical adjacent gene sequences locally. This event usually generates a novel copy of a single gene, but the occurrence of multiple gene-scale tandem duplications is also possible depending on the position of the breakage on chromosomes [23]. As shown in the phylogenic tree (Figure 2), there are several tabacum-exclusive sub-clades of *TPS* genes, suggesting possible neofunctionalization events happened during the evolutionary process. Synonymous substitution rates (Ks) were calculated for each pairwise comparison of *TPS* genes in tabacum-exclusive clades. Ks values under the assumption of a stable molecular clock were analyzed to provide insights regarding the timing and selective pressures from *TPS* duplicates. Out of 20 pairs of recently occurring TDGs we characterized, 13 pairs showed a Ks value less than 0.2, indicating a duplication event less than 7 MYA (Supplementary Figure 2). Three pairs of *TPS* genes possess a Ks score of 0, suggesting extremely recent duplication events. There was a Ks peak at 3.11 MYA of TDGs, these recently derived *TPS* genes might account for the complex terpene profile exclusively found in tobacco.
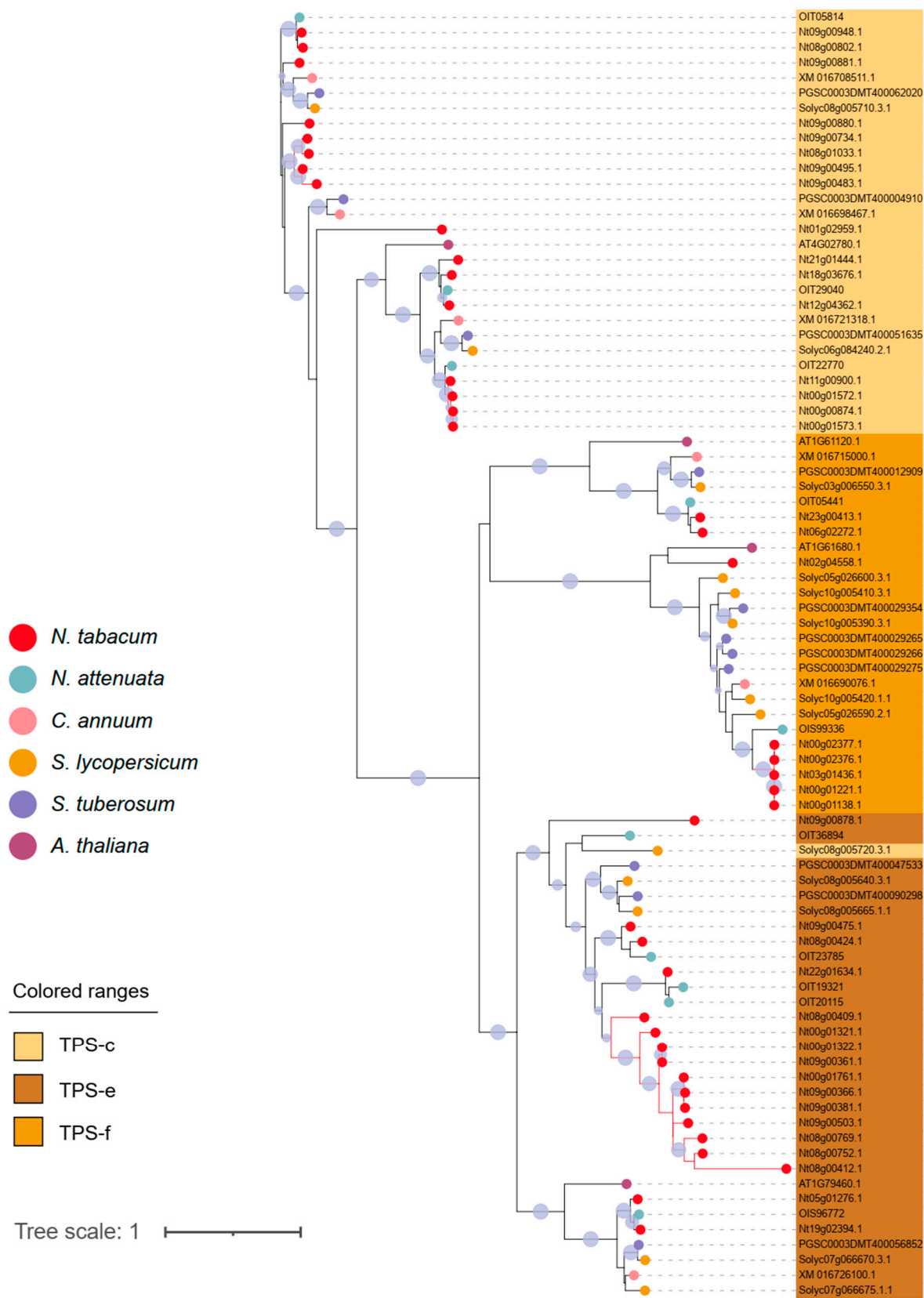
5

Figure 2. Detailed phylogeny of TPS subgroups including TPS-c, TPS-e and TPS-f. *N. tabacum*, *N. attenuata*, *Capsicum annuum*, *Solanum lycopersicum*, *Solanum tuberosum* and Arabidopsis were used for analysis.

Motif analysis of TPS protein

To explore the motif characteristics of the tobacco *TPS* gene family, MEME was used to analyze the tobacco *TPS* gene sequence in two key domains (PF01397 and PF03936). Motif analysis of the tobacco *TPS* gene showed that 6 main motifs exist in the two domains (Figure 3a,c). Most *TPS* genes contains these six major motifs, and the order of these motifs in the *TPS* gene tends to be consistent (Figure 3b,d). Additionally, the six main motifs of the two Pfam domains showed presence-absence variations in tobacco *TPS* genes. Some TPS genes harboured fewer motifs while the motifs of each Pfam domain were mainly arranged in the same order on the corresponding domain. Our analysis showed a landscape of both the diversity and the conservation among the TPS protein sequence domains (Figure 3b,d).
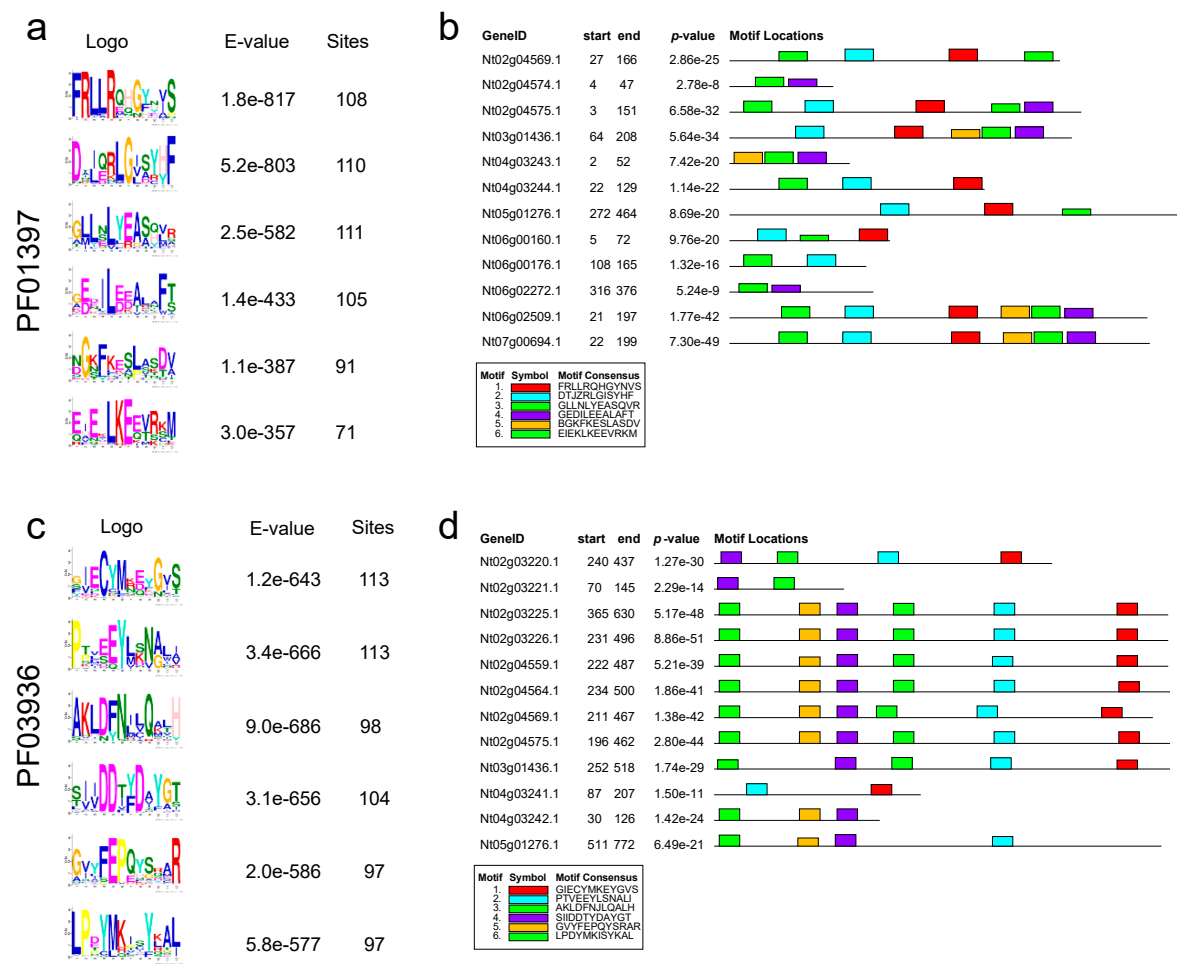

Figure 3. (a, c) MEME motifs identified and **(b, d)** locations of these motifs in two domains (PF01397 and PF03936) of TPS genes in tobacco. The order of these six motifs in the *TPS* gene tends to be consistent, and the motifs of each Pfam domain were arranged in the same order on the corresponding domain, indicating the conservation of their domains.

Alternative splicing analysis of TPS genes

Alternative splicing is widespread in eukaryotes and significantly increases the versatility of gene products. It is conferring great adaptability to a species [24]. By comparing transcriptomes of the two tobacco cultivars including the widely spread accession Y3 and the narrowly distributed accession SK6 at different developmental stages (stage A, B, and C), five alternative splicing types were found on the above identified tobacco TPS genes. Those splicing events included skipped exons (SE), retained introns (RI), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS) and mutually exclusive exons (MXE) (Table 1). TPS transcripts of the same cultivar always differ in

splicing by SE or RI between stages, while the same genes can result in different gene products via alternative splicing sites (A5SS, A3SS) or MXE across different cultivars. This implies that alternative splicing events possibly play a role in the diversification of TPS genes.

Table 1. Alternative splicing events of TPS genes identified from pairwise comparison of tobacco transcriptomes. The number of AS events was listed.

Groups of comparisons	Number of AS
TPS_SK6-A-**vs.**-SK6-B_SE	2
TPS_SK6-B-**vs.**-SK6-C_RI	4
TPS_SK6-B-**vs.**-SK6-C_SE	2
TPS_YY3-A-**vs.**-SK6-A_A5SS	2
TPS_YY3-A-**vs.**-SK6-A_MXE	2
TPS_YY3-A-**vs.**-SK6-A_RI	7
TPS_YY3-A-**vs.**-SK6-A_SE	3
TPS_YY3-B-**vs.**-SK6-B_A3SS	4
TPS_YY3-B-**vs.**-SK6-B_RI	4
TPS_YY3-B-**vs.**-SK6-B_SE	3
TPS_YY3-B-**vs.**-YY3-C_RI	2
TPS_YY3-C-**vs.**-SK6-C_A5SS	1
TPS_YY3-C-**vs.**-SK6-C_RI	6
TPS_YY3-C-**vs.**-SK6-C_SE	4

Expression dynamics of TPS genes at different growth stages

To investigate gene expression activities of TPS genes during development from a temporal perspective, fresh mature leaves of two *N. tabacum* cultivars (Y3 and SK6) were harvested and further sequenced at three different growth stages: (A) during anthesis; (B) seven days before harvest and curing; and (C) during harvesting and curing. Total RNA sequencing generated 18 high-quality transcriptomic datasets. Over 95% of reads remained after quality control. After gene expression quantification, we found that six of the *TPS-a* and *TPS-b* as well as *TPS-e* genes were constitutively expressed at high levels regardless of developmental stages and different cultivars, while some of the *TPS-a* and *TPS-b* as well as *TPS-e* genes expressed differently between stages or cultivars (Figure 4). Meanwhile, the majority of *TPS* genes was expressed at relatively low levels regardless of genetic background and developmental stage (Figure 4). These results suggested that the six constitutively expressed *TPS* genes might serve as "housekeeping" to synthesize the common terpenes/terpenoids. Also, some *TPS* genes might participated in crosstalk with pathways involved in leave development. The non-expressed *TPS* genes may play roles in specialized terpene/terpenoid biosynthesis or under special environment.
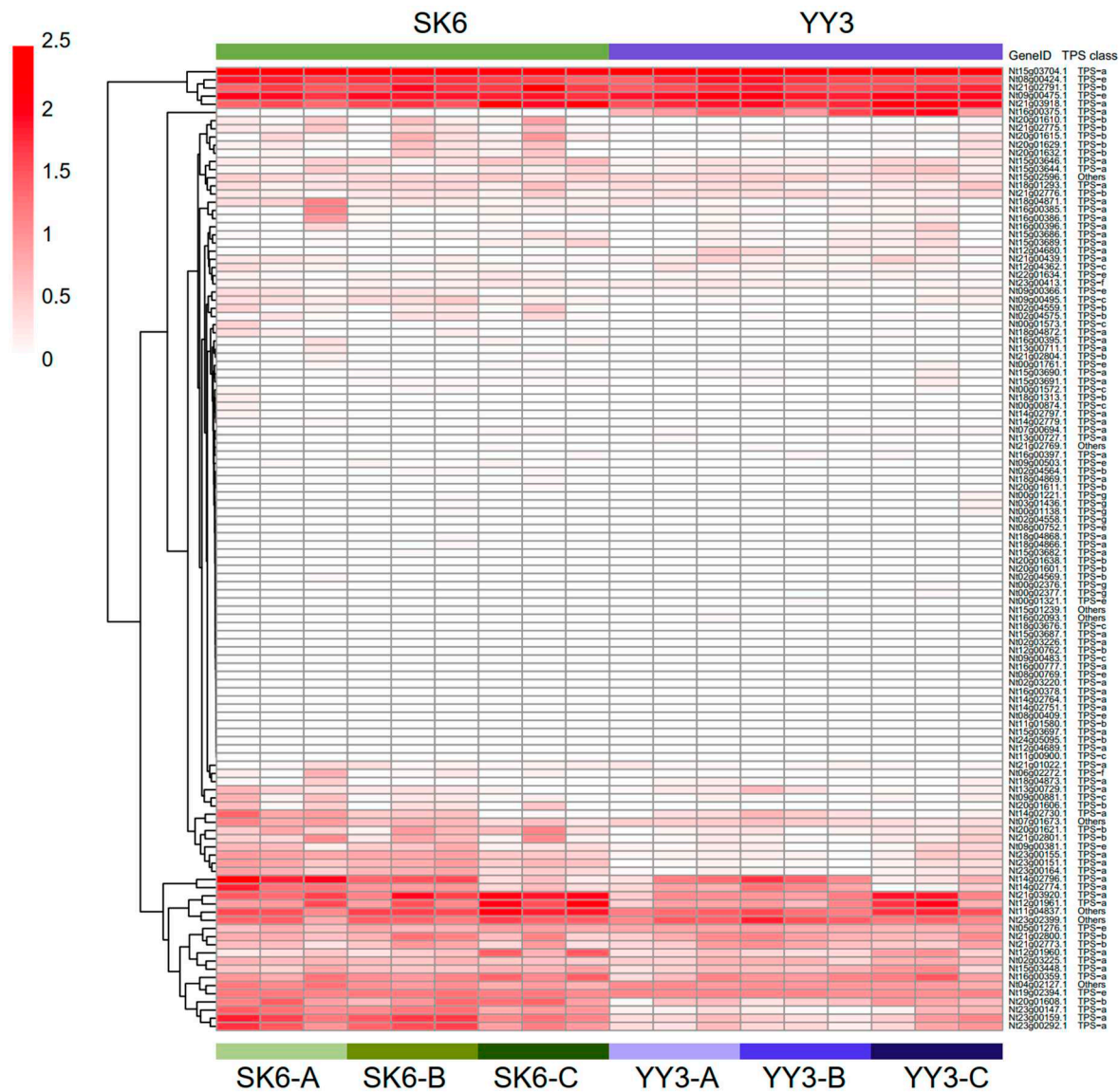
Figure 4. Comparative heatmap of TPS gene expression levels of the same cultivar in different growth stages; A: during anthesis; B: 7 days before harvest and curing; C: at harvest and curing. Several genes showed various expression pattern.

Differentially expressed genes (DEGs) in the TPS gene family

SK6 and Y3 cultivars were compared between growth stages A (during anthesis) and B (7 days before harvest and curing, post anthesis). A low number of DEGs were detected (3 and 0, respectively), which suggests conserved expression of TPS genes during development (Figure 5). For the comparison of both cultivars between growth stages B and C (at harvesting and curing), the transcriptomic landscape for the *TPS* gene family shifted considerably, as 19 *TPS* genes showed differential expression. Specifically, seven and 14 *TPS* genes were upregulated while 12 and five were downregulated. Those results indicated a substantial change in the terpene profile of tobacco post anthesis.
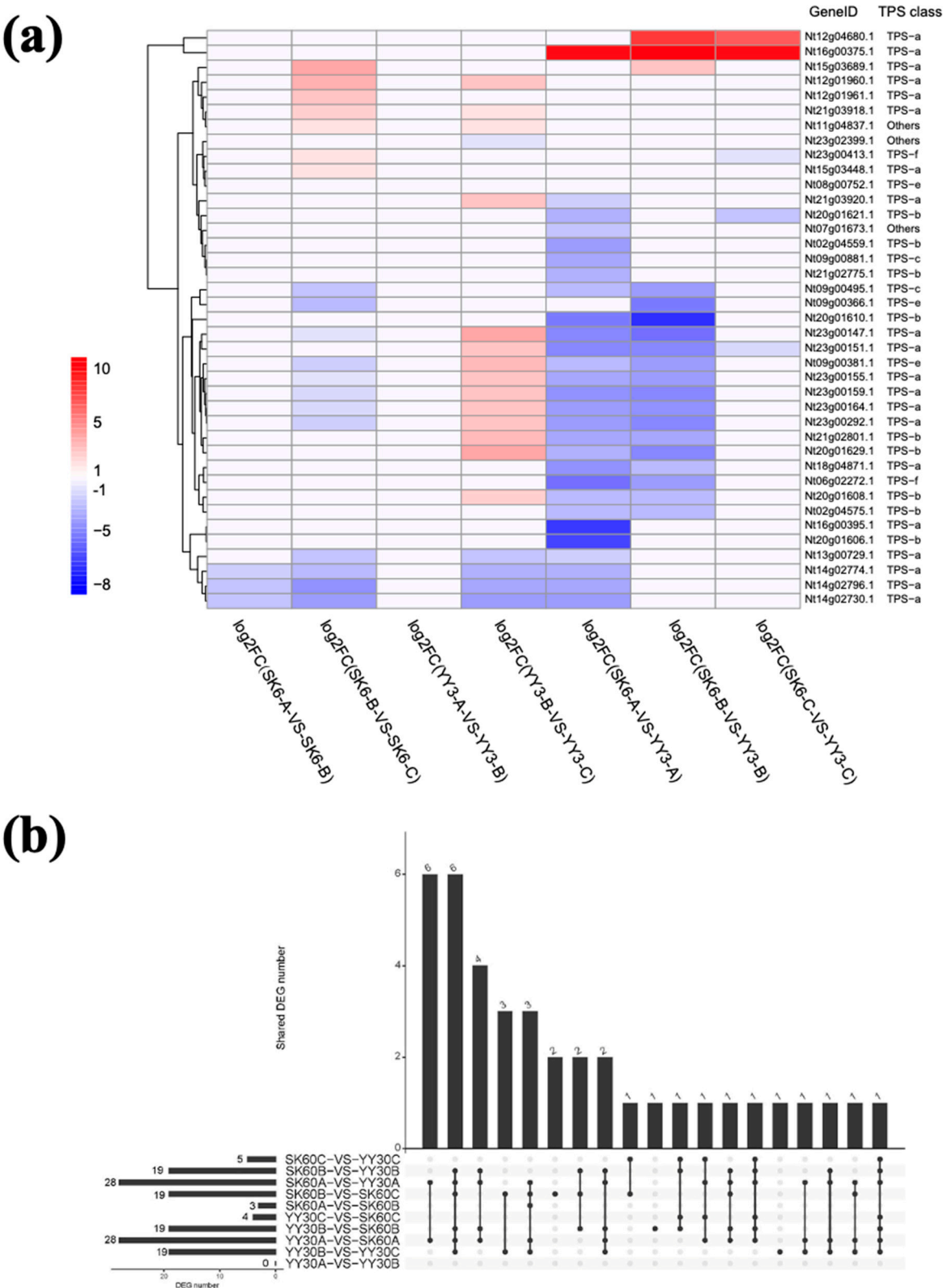
Figure 5. (a) Pairwise comparative DEG heatmap of N. tabacum TPS gene family expression. Four groups on same cultivar, different growth stages (SK6A vs. SK6B; SK6B vs. SK6C; YY3A vs. YY3B; YY3B vs. YY3C) and three groups on same growth stages, different cultivars (SK6A vs. YY3A; SK6B vs. YY3B; SK6C vs. YY3C). **(b)** Overall statistics of the numbers of DEGs found in pairwise comparisons in the aforementioned groups.

In terms of the comparison between the two cultivars, two *TPS* genes (Nt12g04680.1 and Nt16g00375.1) were found to have considerably higher expression levels in SK6 than in Y3. These

TPS genes shared the same origin with wild tobacco relatives such as *N. attentuata*, suggesting that they might be functionally related to herbivory or stress resistance traits. On the other hand, more derived *TPS* lineages from SK6 were found to have lower expression levels in growth stages A and B, with 27 and 16 downregulated *TPS* DEGs, respectively. However, the expression of the aforementioned downregulated genes in SK6 mostly increased to the same level as that in Y3 during harvest. The aforementioned derived *TPS* genes are likely the result of tandem duplications, contributing to the flavour of tobacco smoke. In addition, it is suspected that those flavour terpenes present highly distinct expression patterns in the Y3 and SK6 cultivars. Specifically, they were constitutively produced in Y3 but can only be generated post anthesis in SK6.

Discussion

As an economically important worldwide crop, *N. tabacum* has long been known for serving as a significant raw material for tabacco production. In addition to its economic roles, *N. tabacum* contains a wide range of terpenes and terpenoids, which previously were considered to be deterrents to herbivory *in situ* [25]. Fresh leaves induce vomiting and narcosis to other animals. Many essential oil terpenes contribute to the complex odour of tobacco smoke, which can vary between different cultivars. Terpene synthase is regarded as the key enzyme in the biosynthesis of terpene and has been extensively highlighted to have important functions in plants, therefore we did sequence analyses for identification and characterization of its TPS gene family in this study. A phylogeny of tobacco TPS genes alongside TPS from other solanaceous relatives and *Arabidopsis thaliana* was subsequently investigated to understand the evolution of the TPS gene family. Transcriptomes of two cultivars at three distinct growth stages were sequenced and compared.

The number of TPS genes found in this study is relatively high. With 186 identified TPS genes in *N. tabacum* 'Y3', the number of identified TPS is significantly higher compared with its known Solanaceae relatives and Arabidopsis. This suggested that *N. tabacum* possesses a highly different and complex terpene profile pattern. As reported earlier, the recruitment of a multitude of genetic materials over millennia of human selection may play the major role [26].

The evolutionary patterns of the TPS family in *N. tabacum* cultivar 'Y3', wild tobacco relative *N. attenuata*, chili pepper (*Capsicum annuum*), tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*) and the model plant *Arabidopsis thaliana* showed some clades contain TPS genes only from *N. tabacum* and *N. attenuate*. We found this pattern concurs with the overall molecular phylogeny of Solanaceae, where the genus *Nicotiana* is at a more basal level than *Capsicum* and *Solanum* on the phylogenetic tree [22]. It is observed that tandem duplication played an important role in the occurrence of tabacum-exclusive clades [27]. The tandem duplication accounts for the formation of gene clusters of the same gene family and confers new phenotypic traits [28]. Soybean (*Glycine max*) mutants, whose quantitative trait locus *Rhg1* gene has been duplicated, gained resistance to cyst nematodes [29]. A previous study indicated that tandem duplication significantly contributed to gene family expansion and expression divergence in the survival of these expanded genes [30]. Therefore, tandem duplication of genomic regions explains the occurrence of tabacum-exclusive clades on the TPS phylogeny and thus the complex terpene profile of *N. tabacum*.

Our analysis of the expression profiling for the two tobacco cultivars (Y3 and SK6) at different growth stages discovered five alternative splicing mechanisms [31]. Alternative splicing is a transcriptional modification to pre-mRNA that allows one single gene to code multiple versions of proteins [32]. Certain omitted exons and introns may be retained. All possible candidates of mature mRNA are called splice variants [24]. In our study, transcripts of the same cultivar always differed in splicing by SE or RI, whereas the same genes can result in different gene products via alternative splicing sites (A5SS, A3SS) or MXE across different cultivars. The *TPS-a* and *TPS-b* genes and the derived TPS were clearly expressed in a different pattern.

Differential gene expression analyses between the two cultivars at different growth stages also indicated significantly different terpene profiles before and post anthesis. Two TPS genes (Nt12g04680.1 and Nt16g00375.1) that had higher expression levels in SK6 shared the same ancestry with *N. attentuata*, which is a wild tobacco relative. This suggested a functional linkage to herbivory

or stress resistance traits. Interestingly, it was observed that more derived TPS lineages from SK6 have lower expression levels in growth stages A and B before harvest, while their expression nearly increased to the same level as those in Y3 during harvest. The aforementioned derived TPS genes might be the result of tandem duplications, which resulted in the flavour of tobacco smoke [33]. Generally, our results suggested that those flavour terpenes displayed highly distinct expression patterns in the Y3 and SK6 cultivars.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Supplementary File 1. The genomic coordinates of 186 *TPS* genes in GFF3 format. Supplementary File 2. The protein sequence alignment of TPS genes. Supplementary Figure S1. The full NJ tree constructed using TPS genes. Supplementary Figure S2. The distribution of divergence time between TPS genes.

Data Availability statement: The transcriptome short-read sequence data of the 18 fresh tobacco leaves in this study have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn/search/) with Bioproject ID PRJCA017998. The sequencing data accessions for the 18 tobacco samples can be found at (https://ngdc.cncb.ac.cn/biosample/browse/). The sample accessions for tobacco cultivar SK6 are SAMC2870118 to SAMC2870126 (9 samples), and for tobacco cultivar Y3, they are SAMC2870127 to SAMC2870135 (9 samples). All the materials in this study are available upon request.

Conflicts of Interest: There are no conflicts of interest.

Plants Availability Statement: The plants or plant components in the study adhere to international, national, and institutional protocols. Necessary permissions and licenses have been acquired to collect materials from tobacco plants in the study.

References

1. Croteau, R., Kutchan, T.M. and Lewis, N. G. *Natural Products (Secondary Metabolites). Second Edi. Biochemistry and Molecular Biology of Plants*. (2000).
2. Ashour, M., Wink, M. & Gershenzon, J. Biochemistry of Terpenoids: Monoterpenes, Sesquiterpenes and Diterpenes. in *Annual Plant Reviews Volume 40: Biochemistry of Plant Secondary Metabolism* 258–303 (2010). https://doi.org/10.1002/9781444320503.ch5.
3. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414 (2007).
4. Webb, H., Foley, W. J. & Kühlheim, C. The genetic basis of foliar terpene yield: Implications for breeding and profitability of Australian essential oil crops. *Plant Biotechnol.* **31**, 363–376 (2015).
5. Chatzivasileiou, A. O., Ward, V., Edgar, S. M. & Stephanopoulos, G. Two-step pathway for isoprenoid synthesis. *Proc. Natl. Acad. Sci.* **116**, 506–511 (2019).
6. Kuzuyama, T. & Seto, H. Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **88**, 41–52 (2012).
7. Oldfield, E. & Lin, F.-Y. Terpene biosynthesis: modularity rules. *Angew. Chem. Int. Ed. Engl.* **51**, 1124–1137 (2012).
8. Chen, X. *et al.* Characterization and evolution of gene clusters for terpenoid phytoalexin biosynthesis in tobacco. *Planta* **250**, 1687–1702 (2019).
9. Hayashi, K. *et al.* Identification and functional analysis of bifunctional ent-kaurene synthase from the moss Physcomitrella patens. *FEBS Lett.* **580**, 6175–6181 (2006).
10. Zhan, X., Bach, S. S., Hansen, N. L., Lunde, C. & Simonsen, H. T. Additional diterpenes from Physcomitrella patens synthesized by copalyl diphosphate/kaurene synthase (PpCPS/KS). *Plant Physiol. Biochem.* **96**, 110–114 (2015).
11. Bohlmann, J., Meyer-Gauen, G. & Croteau, R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci.* **95**, 4126–4133 (1998).
12. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
13. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202-8 (2009).
14. Tsai, Y. Te, Huang, Y. P., Yu, C. T. & Lu, C. L. MuSiC: a tool for multiple sequence alignment with constraints. *Bioinformatics* **20**, 2309–2311 (2004).
15. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).

16. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–W478 (2011).

17. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–60 (2015).

18. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

19. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

20. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

21. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5593-601 (2014).

22. Olmstead, R. G. *et al.* A molecular phylogeny of the Solanaceae. *Taxon* **57**, 1159–1181 (2008).

23. Zhang, L., Ma, B., Wang, L. & Xu, Y. Greedy method for inferring tandem duplication history. *Bioinformatics* **19**, 1497–1504 (2003).

24. Singh, P. & Ahi, E. P. The importance of alternative splicing in adaptive evolution. *Mol. Ecol.* **31**, 1928–1938 (2022).

25. Rabara, R. C., Kudithipudi, C. & Timko, M. P. Identification of Terpene-Related Biosynthetic Gene Clusters in Tobacco through Computational-Based Genomic, Transcriptomic, and Metabolic Analyses. *Agronomy* vol. 13 (2023).

26. Lücker, J. *et al.* Increased and altered fragrance of tobacco plants after metabolic engineering using three monoterpene synthases from lemon. *Plant Physiol.* **134**, 510–519 (2004).

27. Cui, A., Jin, Y., Li, Y., Nie, T. & Sun, L. Systematic identification of TPS genes in Gossypium and their characteristics in response to flooding stress. *Front. Plant Sci.* **14**, 1126884 (2023).

28. Panchy, N., Lehti-Shiu, M. & Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **171**, 2294–2316 (2016).

29. Cook, D. E. *et al.* Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* **338**, 1206–1209 (2012).

30. Jiang, Z., Dong, X., Li, Z.-G., He, F. & Zhang, Z. Differential coexpression analysis reveals extensive rewiring of arabidopsis gene coexpression in response to pseudomonas syringae infection. *Sci. Rep.* **6**, 35064 (2016).

31. Shang, X., Cao, Y. & Ma, L. Alternative Splicing in Plant Genes: A Means of Regulating the Environmental Fitness of Plants. *Int. J. Mol. Sci.* **18**, (2017).

32. Lopez, A. J. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* **32**, 279–305 (1998).

33. Chen, H. *et al.* Combinatorial Evolution of a Terpene Synthase Gene Cluster Explains Terpene Variations in Oryza. *Plant Physiol.* **182**, 480–492 (2020).