Plant Communications, Volume *3*

Supplemental information

Chromosome-level assembly and analysis of the *Thymus* genome provide insights into glandular secretory trichome formation and monoterpenoid biosynthesis in thyme

Meiyu Sun, Yanan Zhang, Li Zhu, Ningning Liu, Hongtong Bai, Guofeng Sun, Jinzheng Zhang, and Lei Shi

Chromosome-level assembly and analysis of the *Thymus* genome provide insights on glandular secretory trichome formation and monoterpenoid biosynthesis in thyme

Meiyu Sun[1#], Yanan Zhang[1,2#], Li Zhu[1,2], Ningning Liu[1,2], Hongtong Bai[1], Guofeng Sun[3], Jinzheng Zhang[1*] and Lei Shi[1*]

[1] Key Laboratory of Plant Resources and Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.

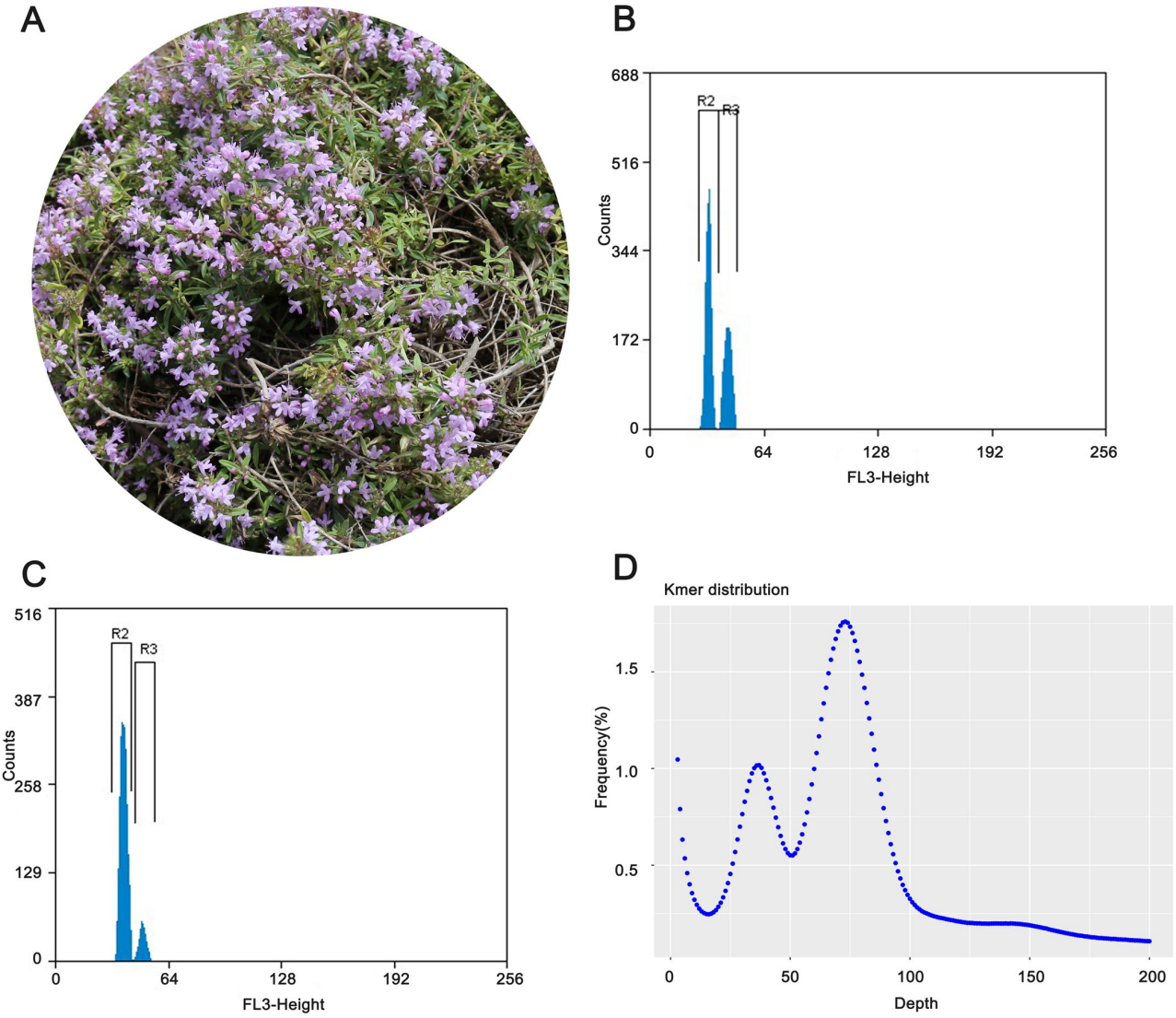
[2] University of Chinese Academy of Sciences, Beijing 100049, China.

[3] Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.

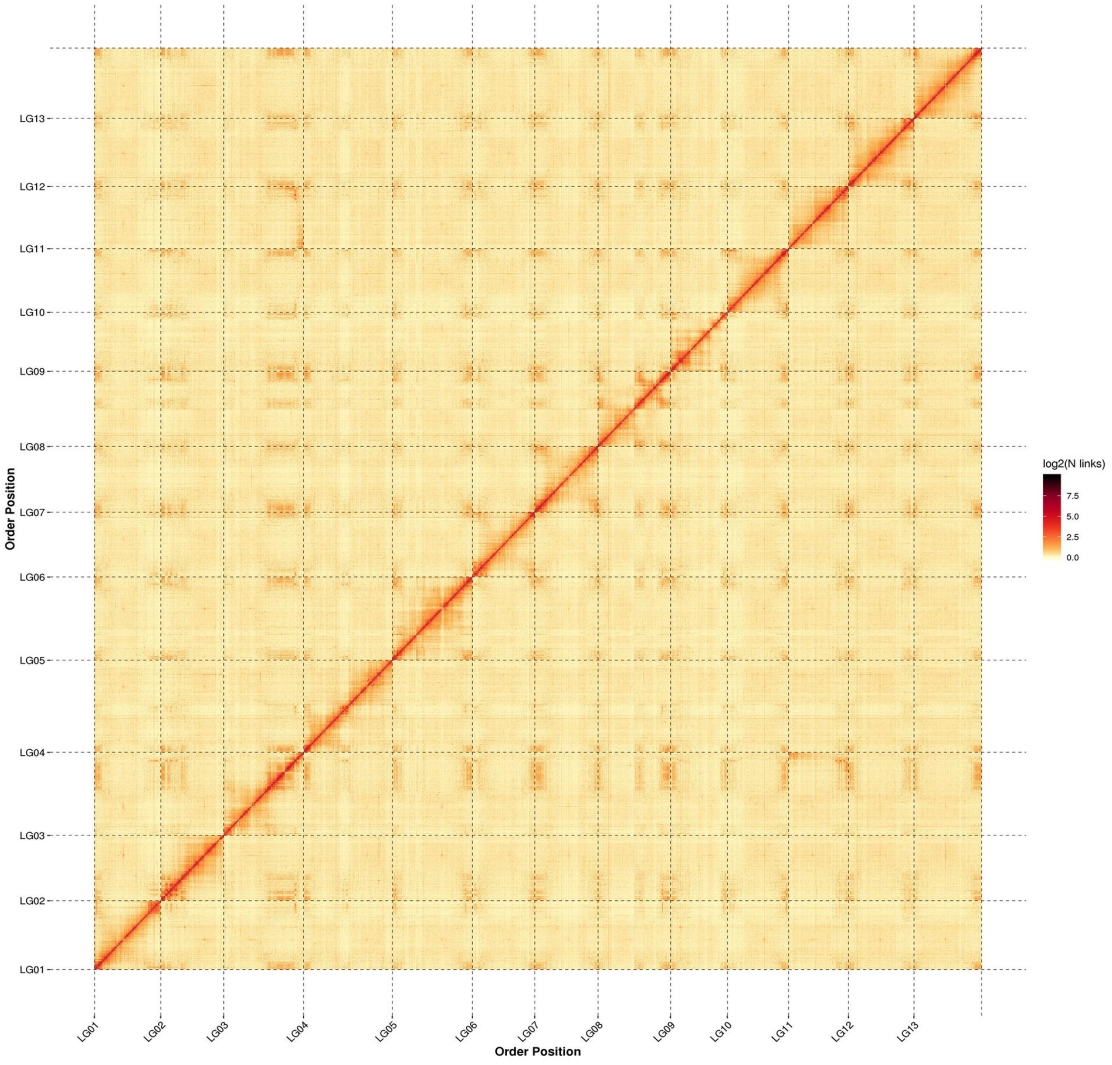
[#] These authors contributed equally to this work.

[*]Correspondence: Lei Shi (shilei_67@126.com), Jinzheng Zhang (caohua@ibcas.ac.cn)

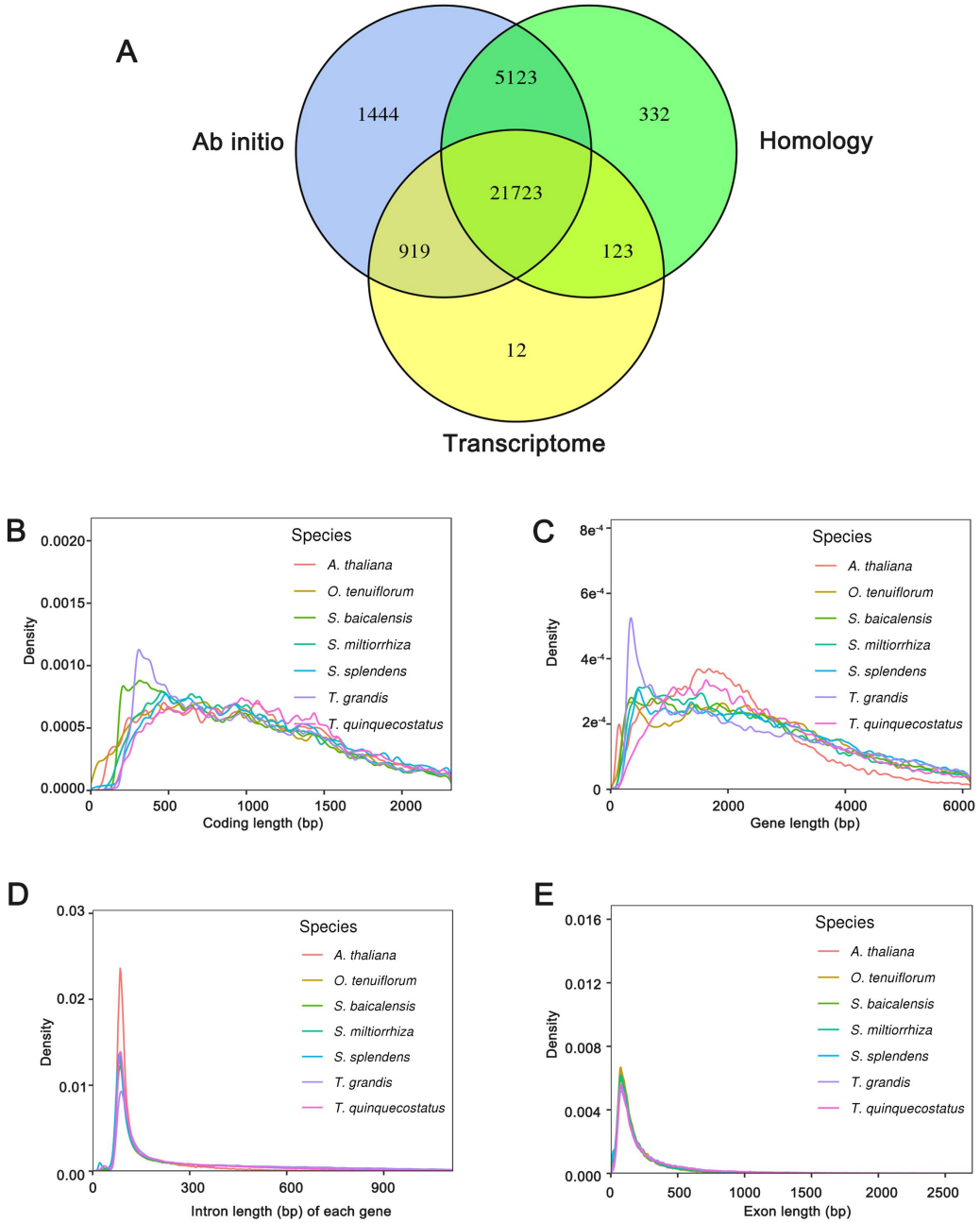
Supplementary Figures



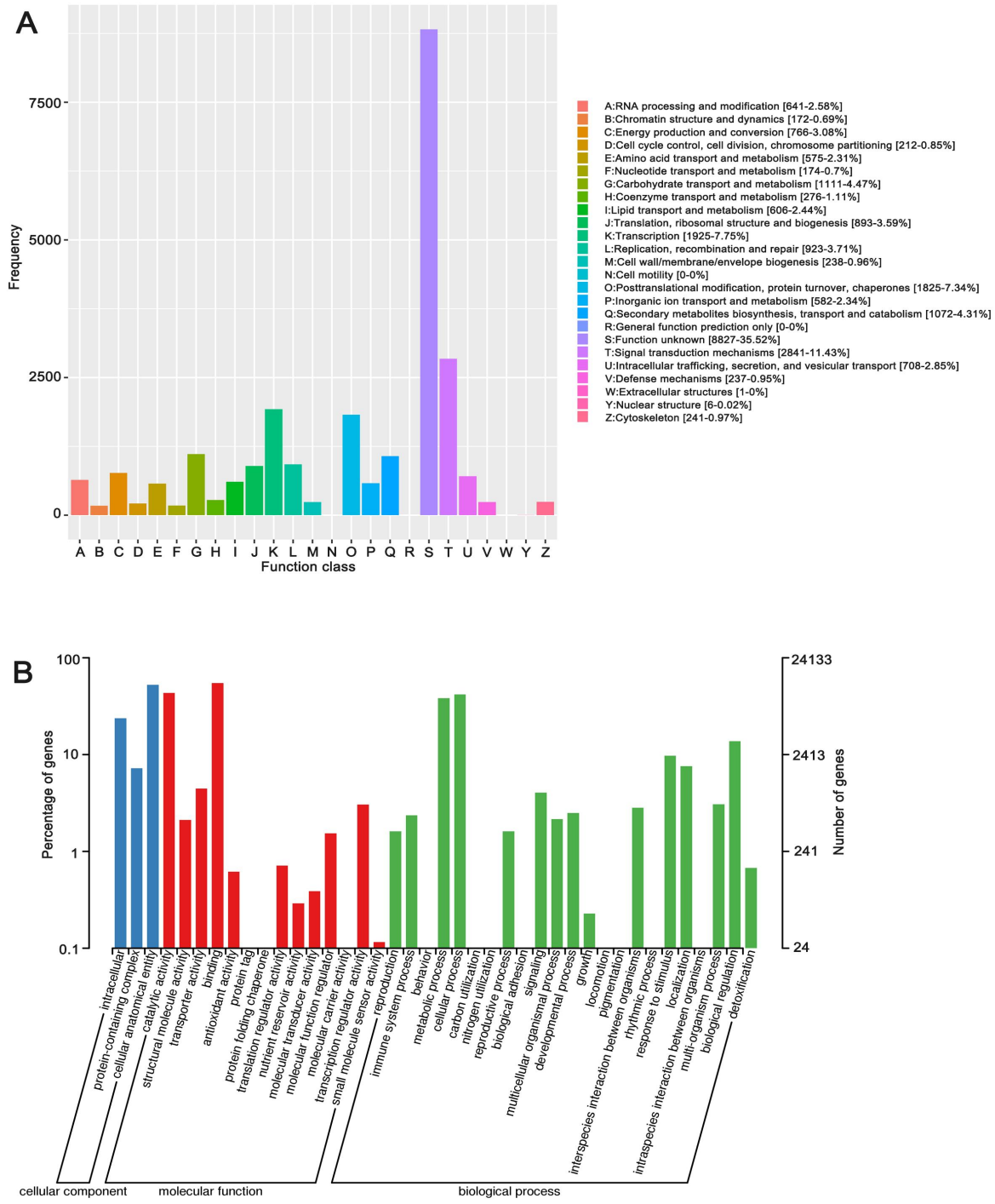
Supplemental Figure 1. Image exhibition, flow cytometry estimation and k-mer distribution analysis of *T. quinquecostatus*. (A) The flower image of *T. quinquecostatus* (CNT39, NCBI: TaxID 228974). **(B)** The genome size of *T. quinquecostatus* was estimated using *Oryza sativa* ssp. *indica* (93-11, 393.5 Mb). **(C)** The genome size of *T. quinquecostatus* was estimated using *Sorghum bicolor* L. (Tx430, 732.2Mb). **(D)** The genome size of *T. quinquecostatus* was estimated based on k-mer counting.



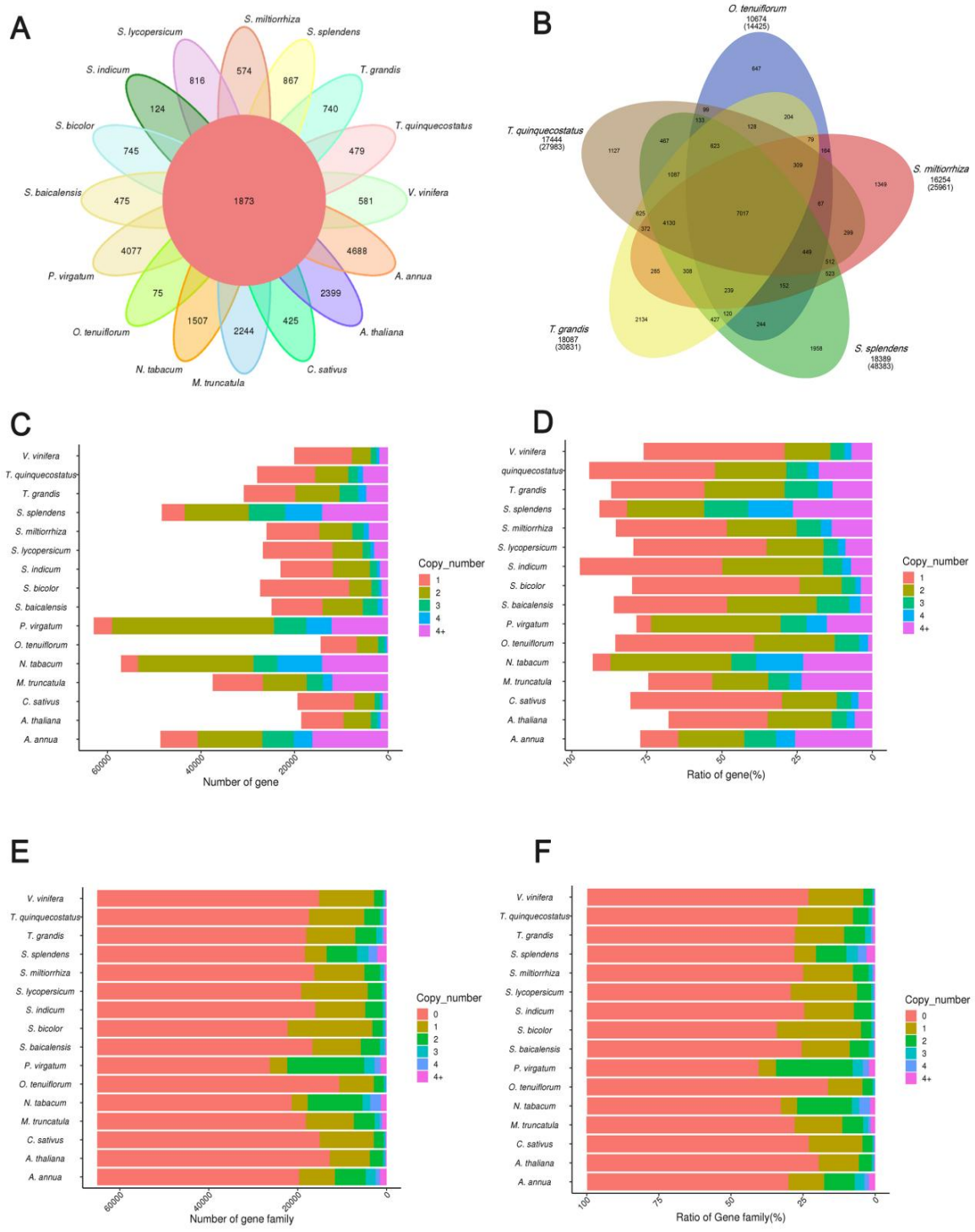
Supplemental Figure 2. The genome-wide Hi-C heat map of chromosome interactions in *T. quinquecostatus*. The intensity of pixels represents the count of Hi-C links between 100 kb windows on 13 pseudochromosomes on a logarithmic scale.



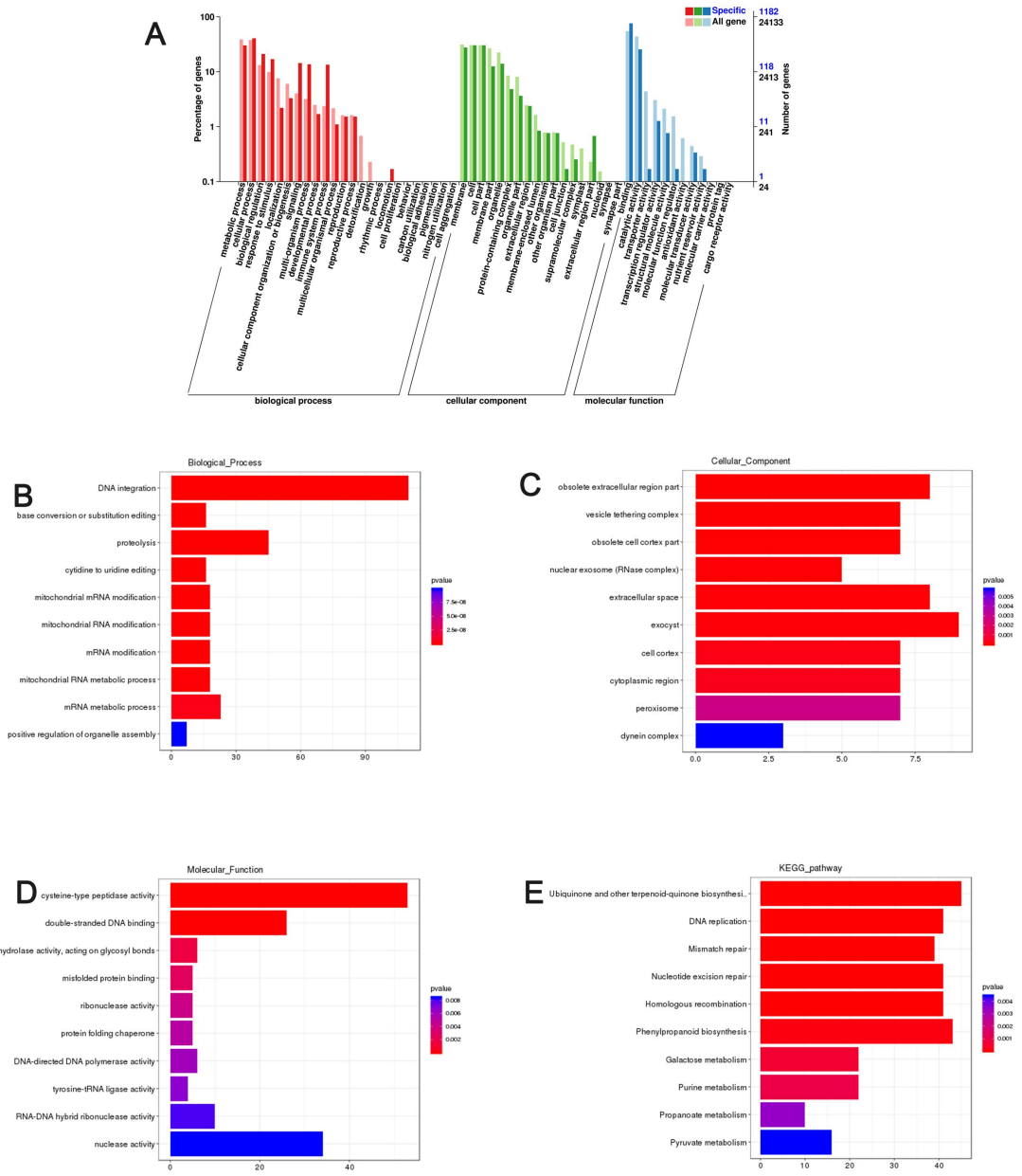
Supplemental Figure 3. Predicted protein-coding genes were predicted in *T. quinquecostatus*. (A) Venn diagram showing protein-coding genes were predicted on the basis of 3 different strategies. **(B)** The coding length of protein-coding genes in *T. quinquecostatus* and related species. **(C)** The gene length of protein-coding genes in *T. quinquecostatus* and related species. **(D)** The intron length of protein-coding genes in *T. quinquecostatus* and related species. **(E)** The exon length of protein-coding genes in *T. quinquecostatus* and related species.



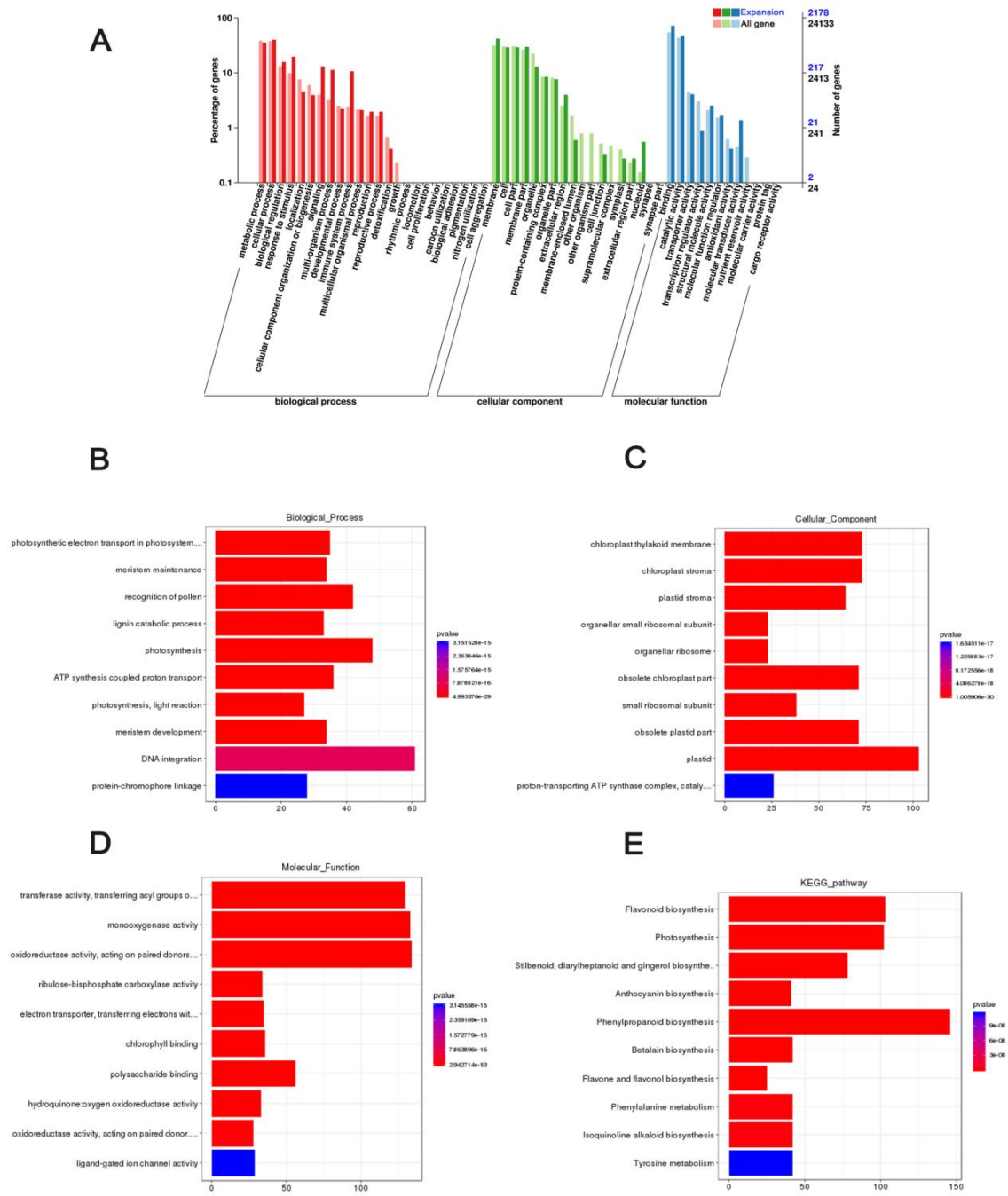
Supplemental Figure 4. Enrichment analysis of protein-coding genes in *T. quinquecostatus*. (A) eggNOG enrichment analysis of protein-coding genes in *T. quinquecostatus*. **(B)** GO enrichment analysis of protein-coding genes in *T. quinquecostatus*.



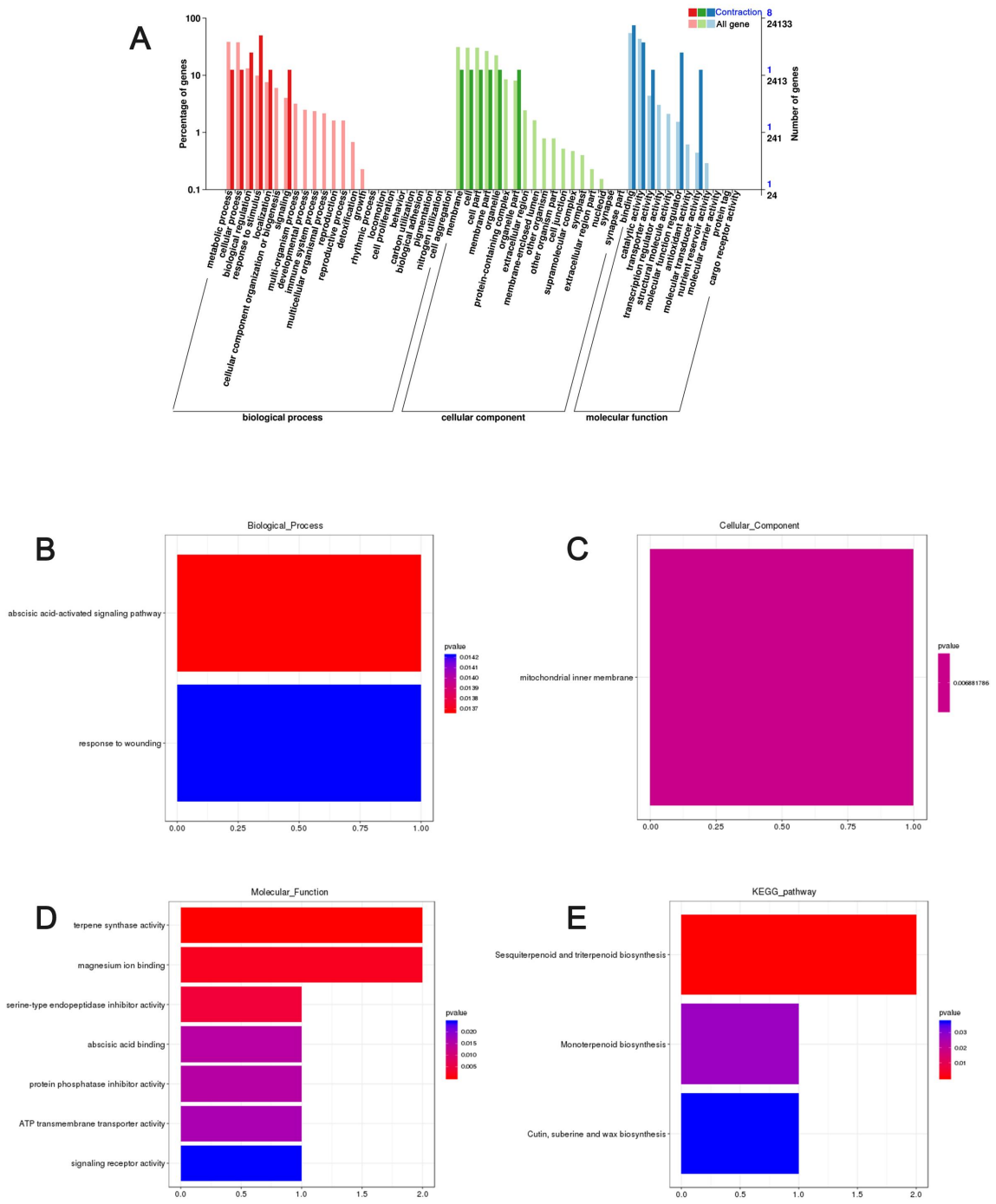
Supplemental Figure 5. Cluster analysis of gene and gene families in *T. quinquecostatus* and other 15 related species. (A) Petal diagram of gene families cluster in 16 species. **(B)** Venn diagram of gene families cluster in 5 Labiaceae species. **(C)** The number of copies distribution of all genes in 16 species. **(D)** The ratio of copies distribution of all genes in 16 species. **(E)** The number of copies distribution of all gene families in 16 species. **(F)** The ratio of copies distribution of all gene families in 16 species.



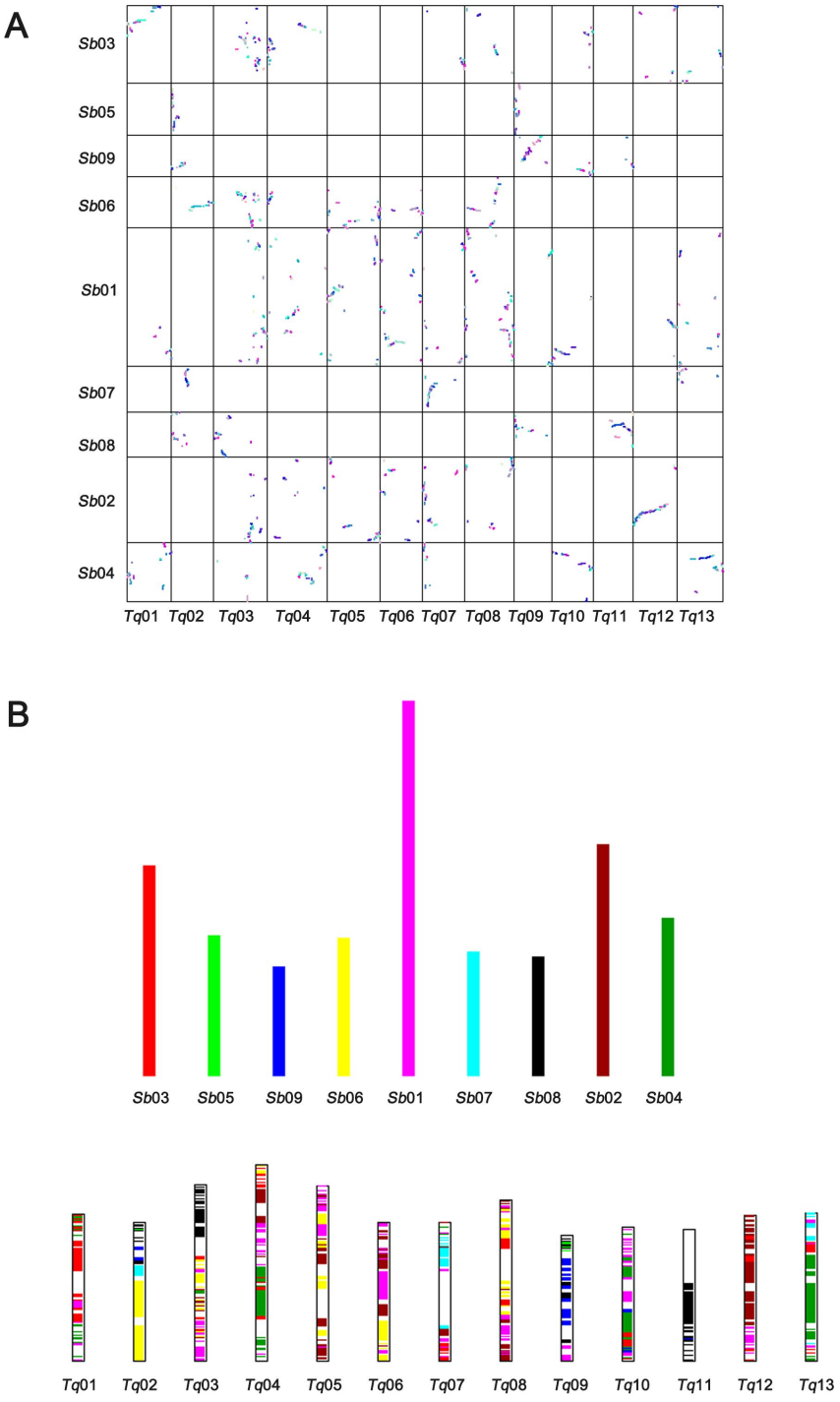
Supplemental Figure 6. Cluster analysis of specific genes in *T. quinquecostatus*. (A) GO enrichment analysis of specific genes in *T. quinquecostatus*. **(B)** Biological process enrichment analysis of specific genes in *T. quinquecostatus*. **(C)** Cellular component enrichment analysis of specific genes in *T. quinquecostatus*. **(D)** Molecular function enrichment analysis of specific genes in *T. quinquecostatus*. **(E)** KEGG enrichment analysis of specific genes in *T. quinquecostatus*.



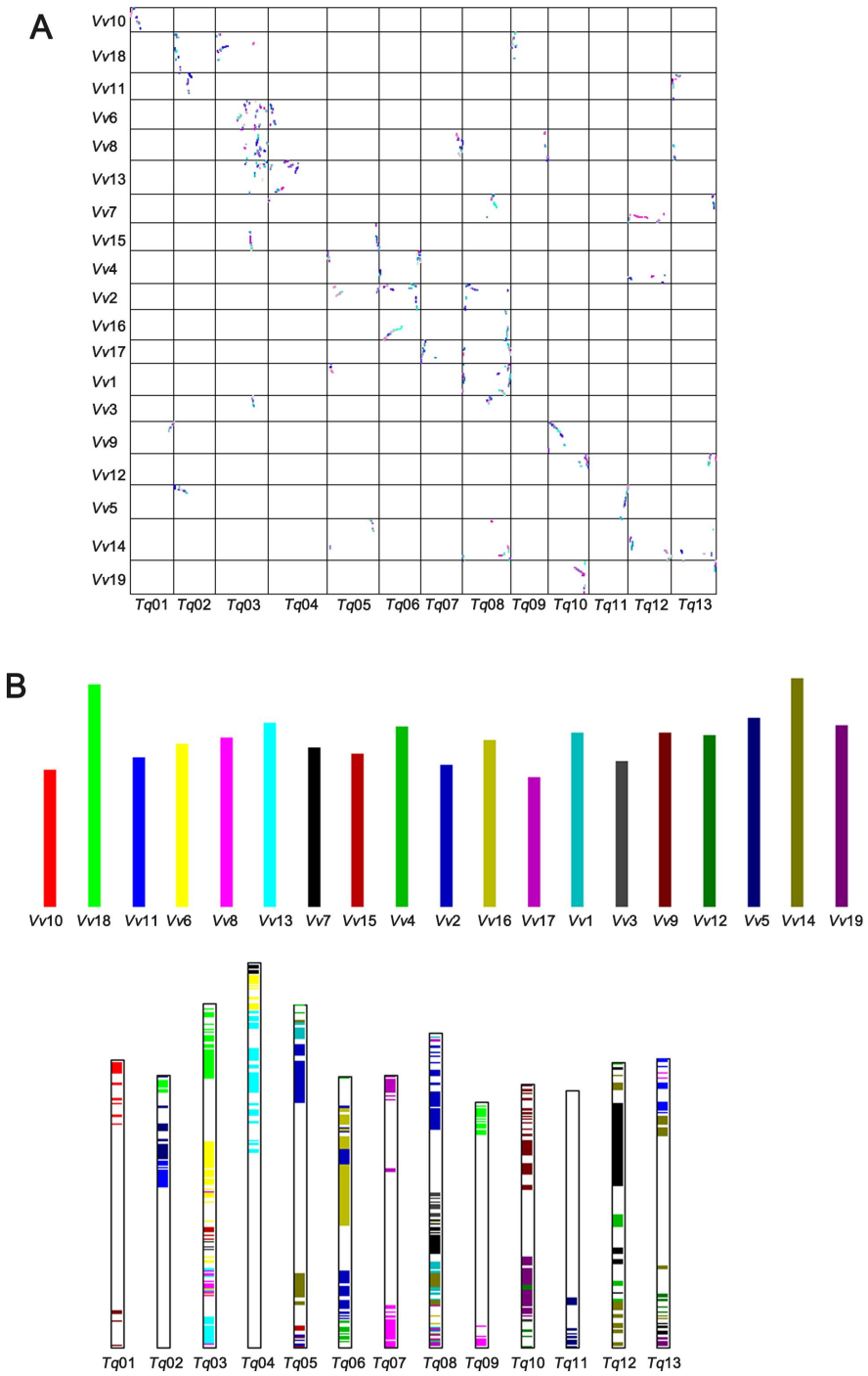
Supplemental Figure 7. Cluster analysis of expansion genes in *T. quinquecostatus*. (A) GO enrichment analysis of expansion genes in *T. quinquecostatus*. **(B)** Biological process enrichment analysis of expansion genes in *T. quinquecostatus*. **(C)** Cellular component enrichment analysis of expansion genes in *T. quinquecostatus*. **(D)** Molecular function enrichment analysis of expansion genes in *T. quinquecostatus*. **(E)** KEGG enrichment analysis of expansion genes in *T. quinquecostatus*.



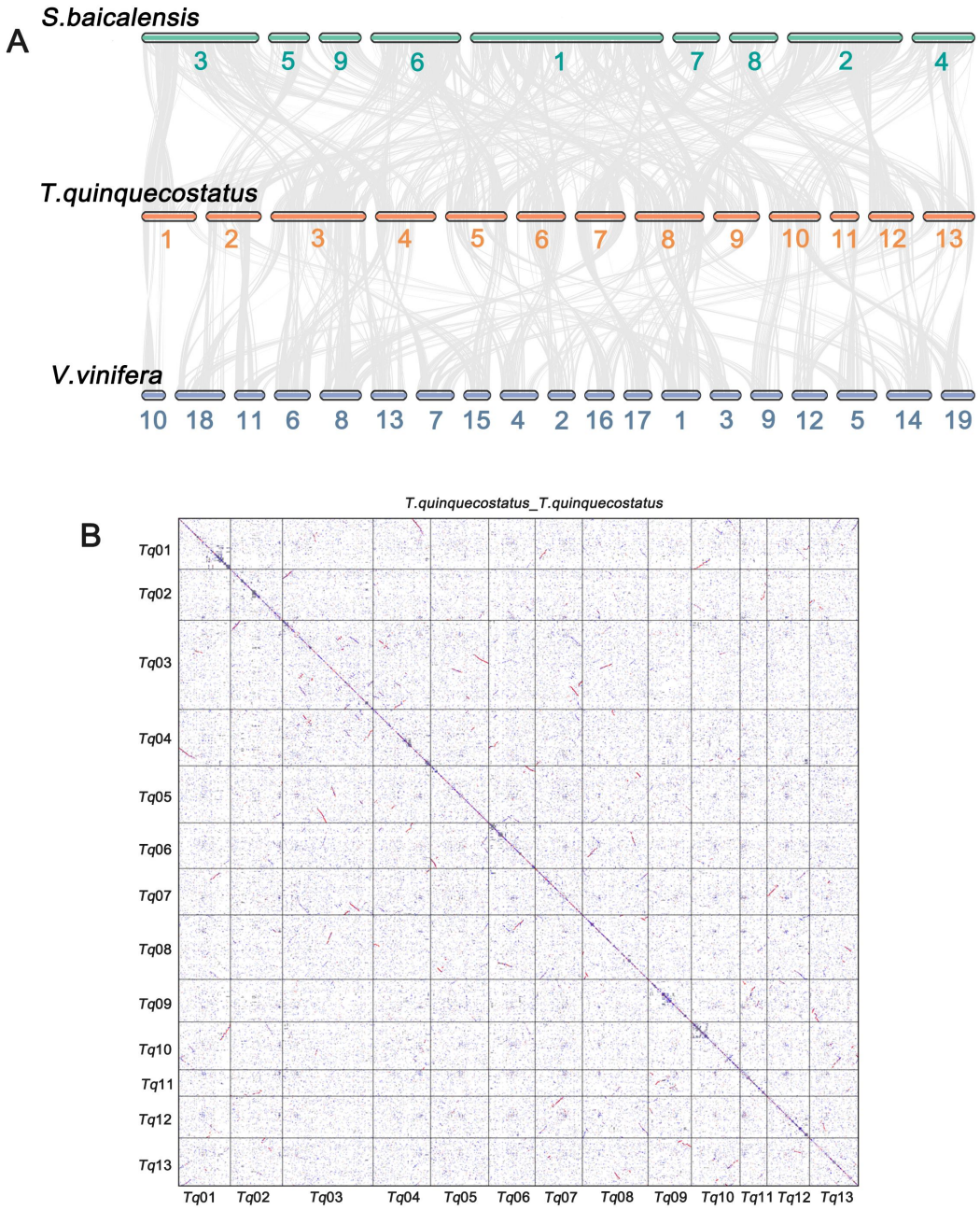
Supplemental Figure 8. Cluster analysis of contraction genes in *T. quinquecostatus*. (A) GO enrichment analysis of contraction genes in *T. quinquecostatus*. **(B)** Biological process enrichment analysis of contraction genes in *T. quinquecostatus*. **(C)** Cellular component enrichment analysis of contraction genes in *T. quinquecostatus*. **(D)** Molecular function enrichment analysis of contraction genes in *T. quinquecostatus*. **(E)** KEGG enrichment analysis of contraction genes in *T. quinquecostatus*.



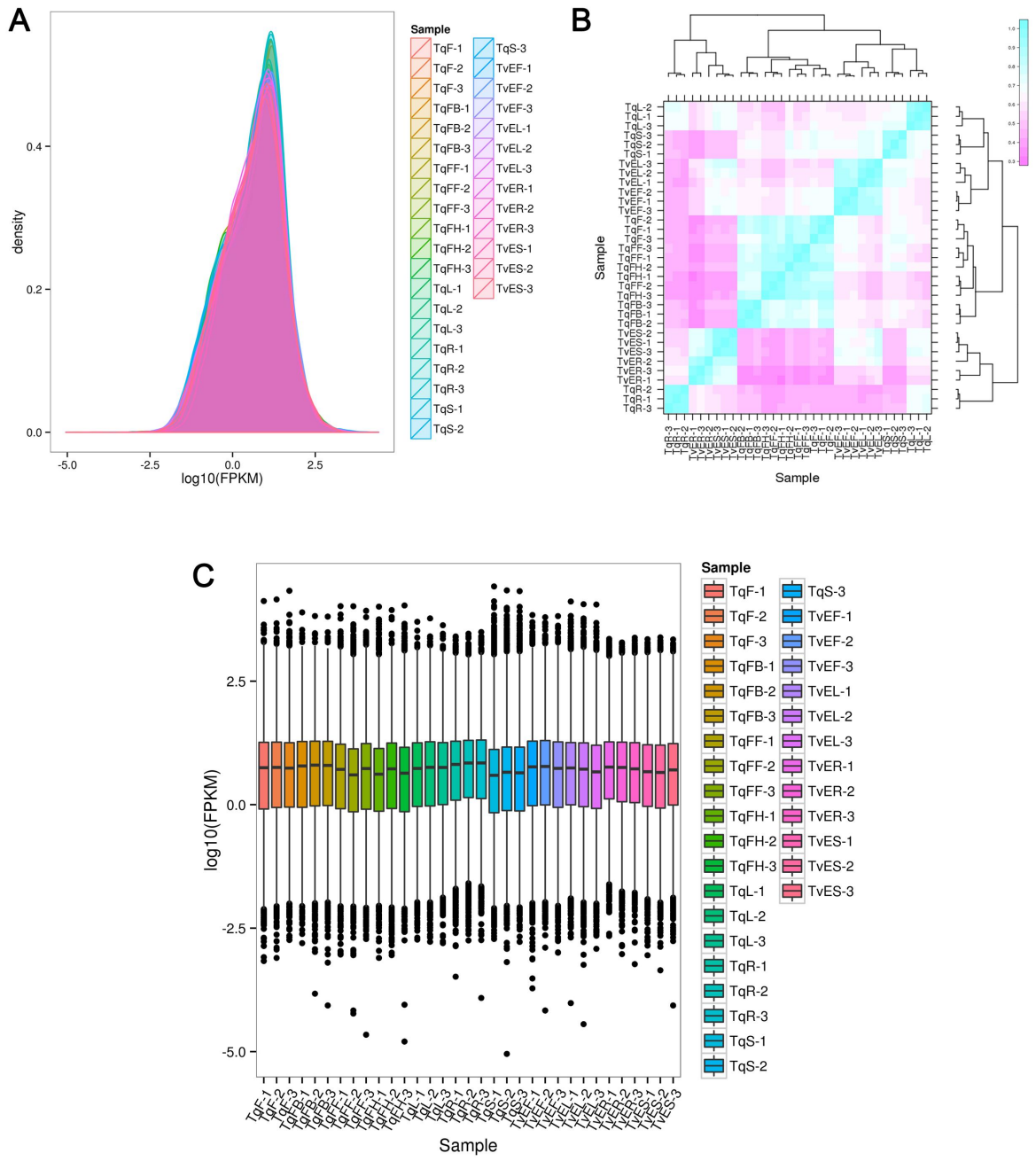
Supplemental Figure 9. Visualization of collinear blocks between *T. quinquecostatus* and *S. baicalensis*. (A) The dot graphs of synteny blocks between *T. quinquecostatus* and *S. baicalensis*. Dots closest to the diagonal line represent collinearity between the two genomes with < 10 Kb fragments filtered out. **(B)** The bar graphs of synteny blocks between *T. quinquecostatus* and *S. baicalensis*.



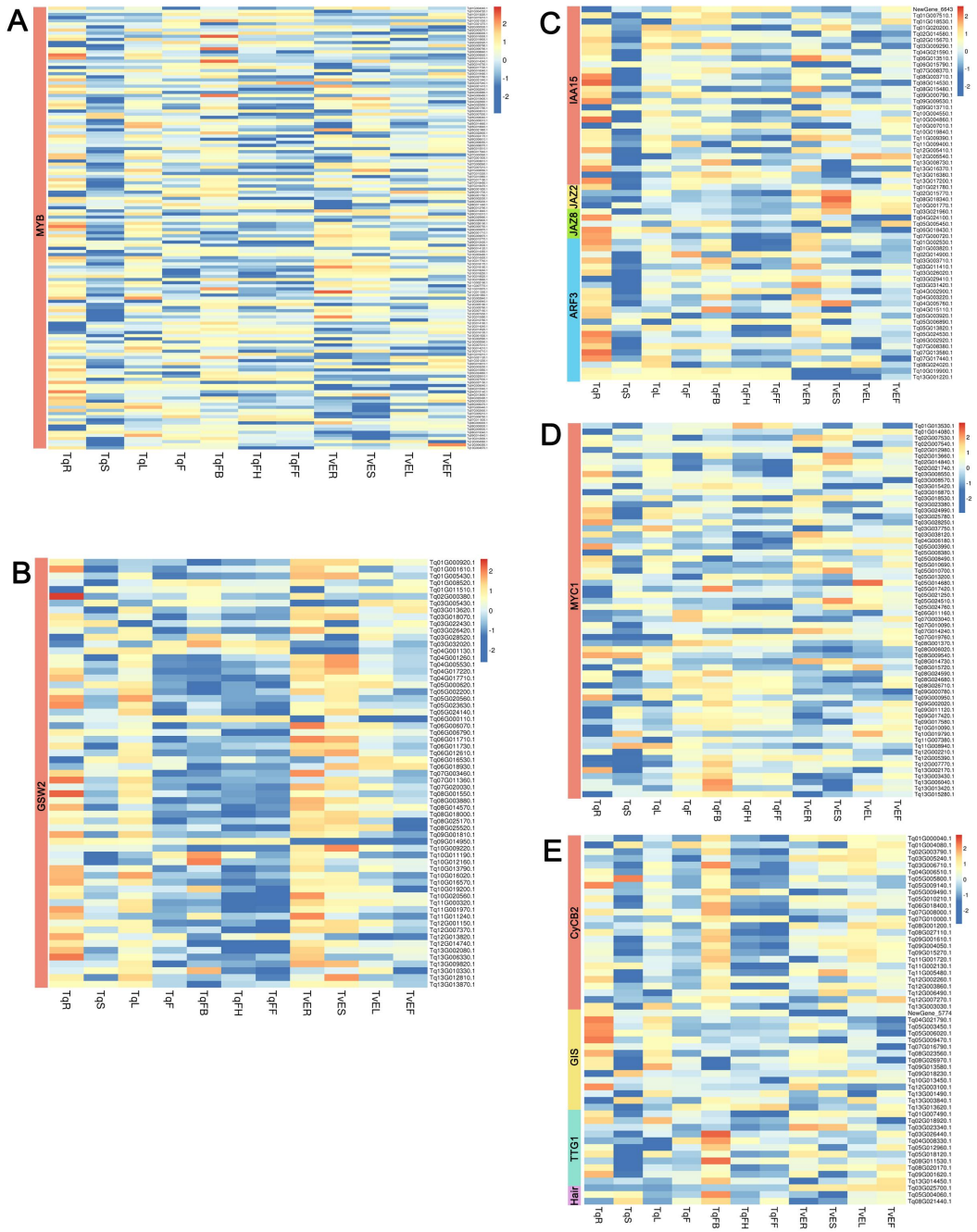
Supplemental Figure 10. Visualization of collinear blocks between *T. quinquecostatus* and *V. vinifera*. (A) The dot graphs of synteny blocks between *T. quinquecostatus* and *V. vinifera*. Dots closest to the diagonal line represent collinearity between the two genomes with < 10 Kb fragments filtered out. **(B)** The bar graphs of synteny blocks between *T. quinquecostatus* and *V. vinifera*.



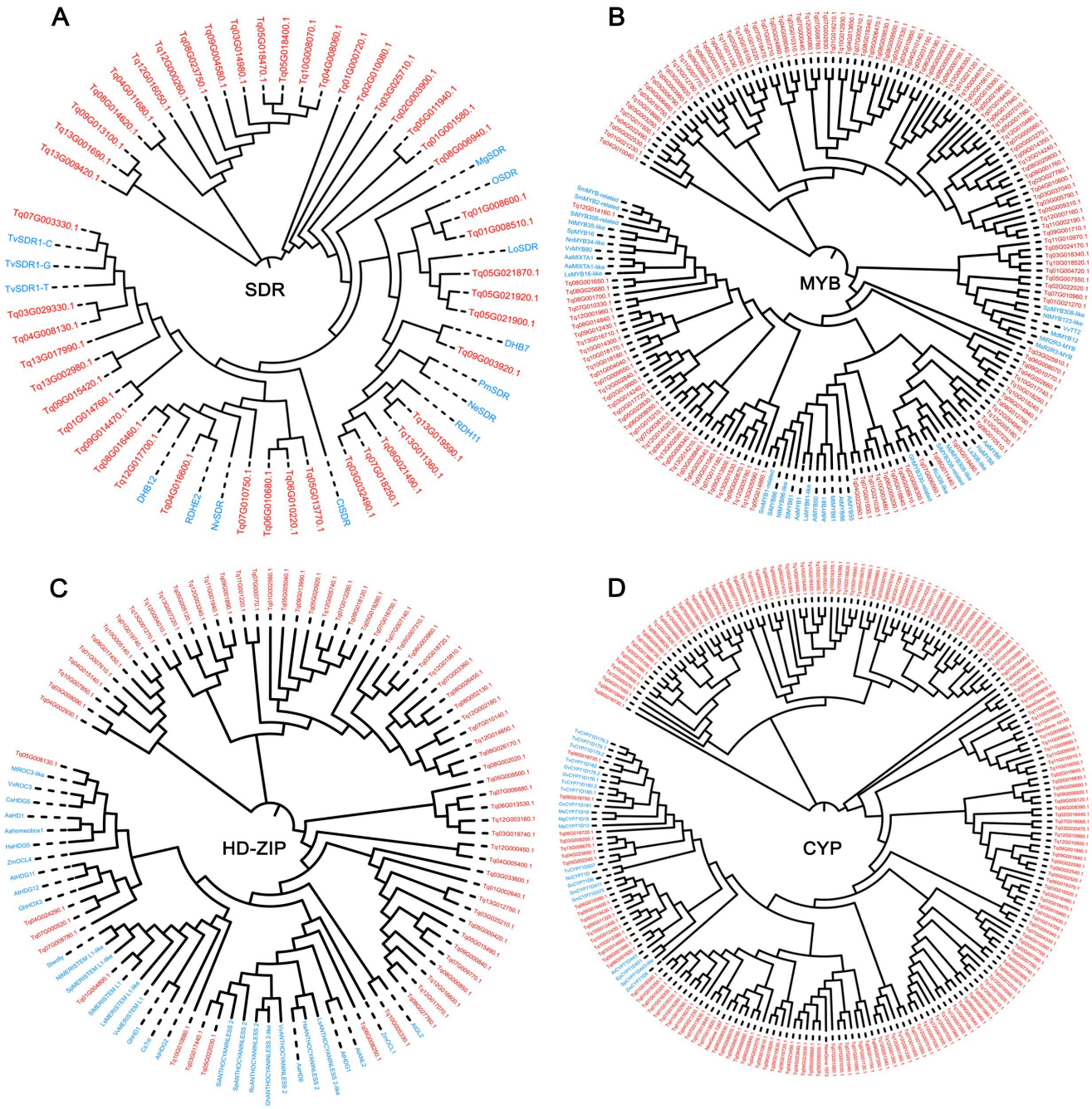
Supplemental Figure 11. Visualization of collinear blocks between *T. quinquecostatus*, *S. baicalensis* and *V. vinifera*. (A) Comparison of the karyotype between *T. quinquecostatus*, *S. baicalensis* and *V. vinifera*. **(B)** The dot graphs of synteny blocks of *T. quinquecostatus* genome. The gene set of *T. quinquecostatus* was self-aligned using blast program. The alignment result was then fed to MCScanX package for synteny blocks identification. The X- and Y-axis represent the 13 chromosomes of *T. quinquecostatus*, respectively.



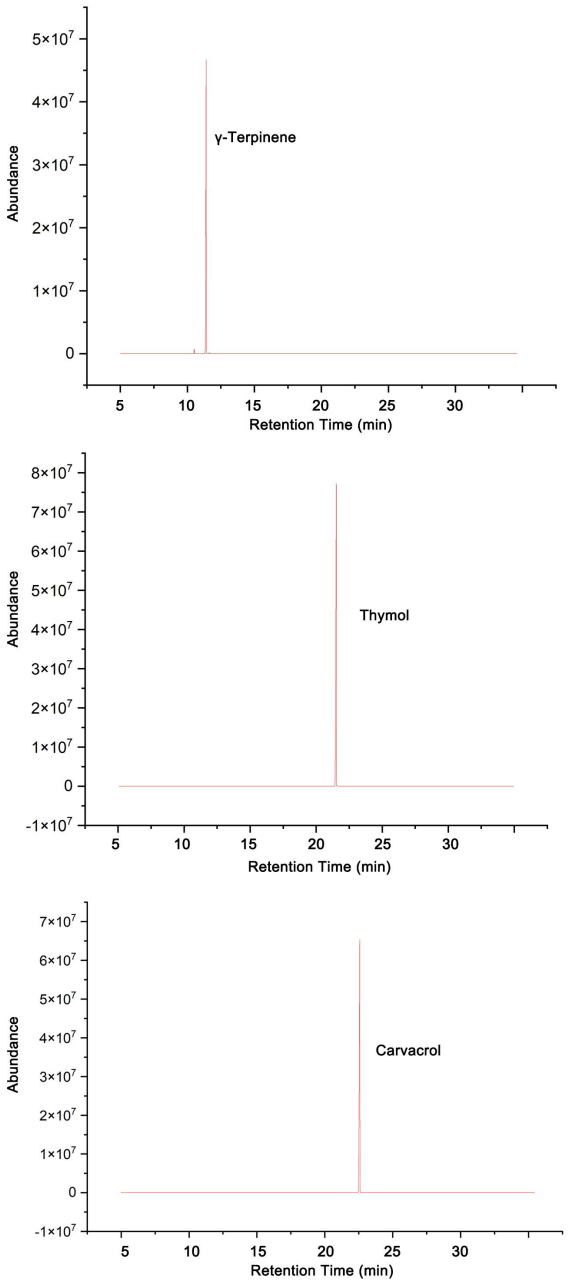
Supplemental Figure 12. Results of thyme transcriptome sequencing. (A) Density of FPKM in 33 samples. **(B)** Heatmap of 33 samples. **(C)** Box of FPKM in 33 samples.



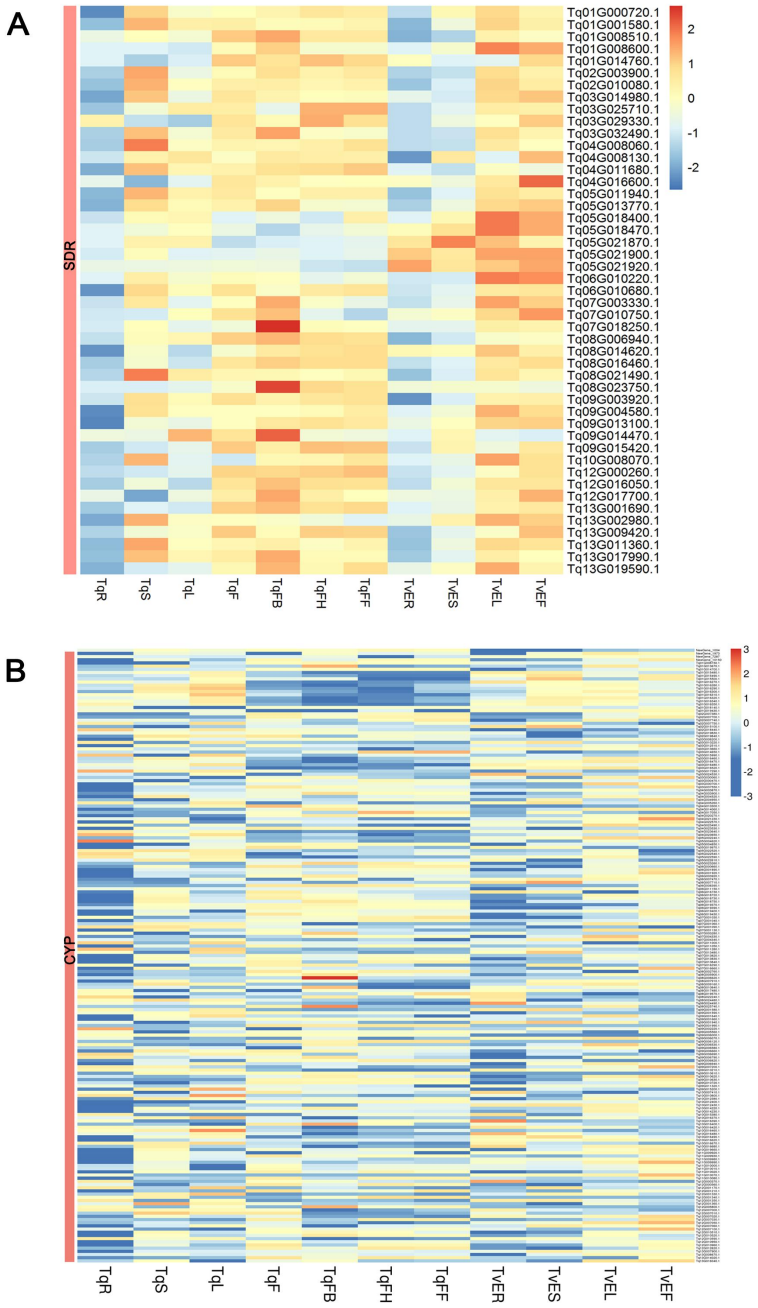
Supplemental Figure 13. Expression heatmap of other key DEGs involved in glandular secretory trichomes (GSTs) formation mechanism. (A) Expression heatmap of MYB-encoding genes in GSTs formation mechanism. **(B)** Expression heatmap of *GSW2s* in GSTs formation mechanism. **(C)** Expression heatmap of hormone-related genes in GSTs formation mechanism. **(D)** Expression heatmap of *MYC1s* in GSTs formation mechanism. **(E)** Expression heatmap of *CyCB2s*, *GISs*, *TTG1s*, and *Hairs* in GSTs formation mechanism. TqS, *T. quinquecostatus* stem; TqL, *T. quinquecostatus* leaf; TqF, *T. quinquecostatus* flower; TqFB, *T. quinquecostatus* flower bud stage; TqFH, *T. quinquecostatus* flower half-bloom stage; TqFF, *T. quinquecostatus* flower full-bloom stage; TvER, *T. vulgaris* 'Elsbeth' root; TvES, *T. vulgaris* 'Elsbeth' stem; TvEL, *T. vulgaris* 'Elsbeth' leaf; TvEF, *T. vulgaris* 'Elsbeth' flower.The bar at the right top marked the normalized FPKM values for both in *T. quinquecostatus* and *T. vulgaris* 'Elsbeth' (red denoted high expression and blue indicated low expression, and expression data were Z-score standardized to -3 to 3 per gene.). All of expression genes showing minimum two-fold changes at $p < 0.05$ (Student's two tailed t-test).



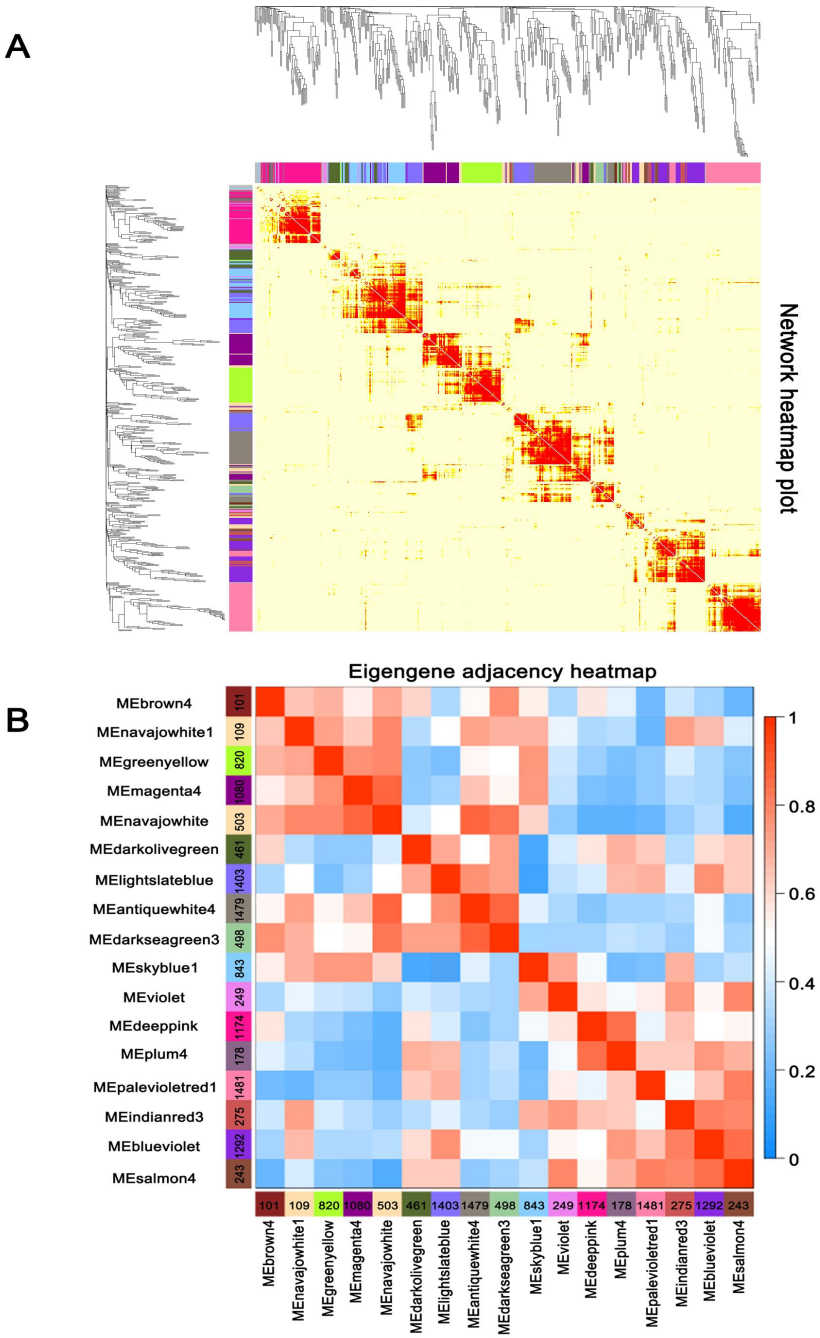
Supplemental Figure 14. Phylogenetic analysis of SDR-, MYB-, HD-ZIP-, and CYP-encoding genes between the *T. quinquecostatus* and related species. (A) SDR-encoding genes phylogenetic analysis in *T. quinquecostatus* and other related species. **(B)** MYB-encoding genes phylogenetic analysis in *T. quinquecostatus* and other related species. **(C)** HD-ZIP-encoding genes phylogenetic analysis in *T. quinquecostatus* and other related species. **(D)** CYP-encoding genes phylogenetic analysis in *T. quinquecostatus* and other related species.



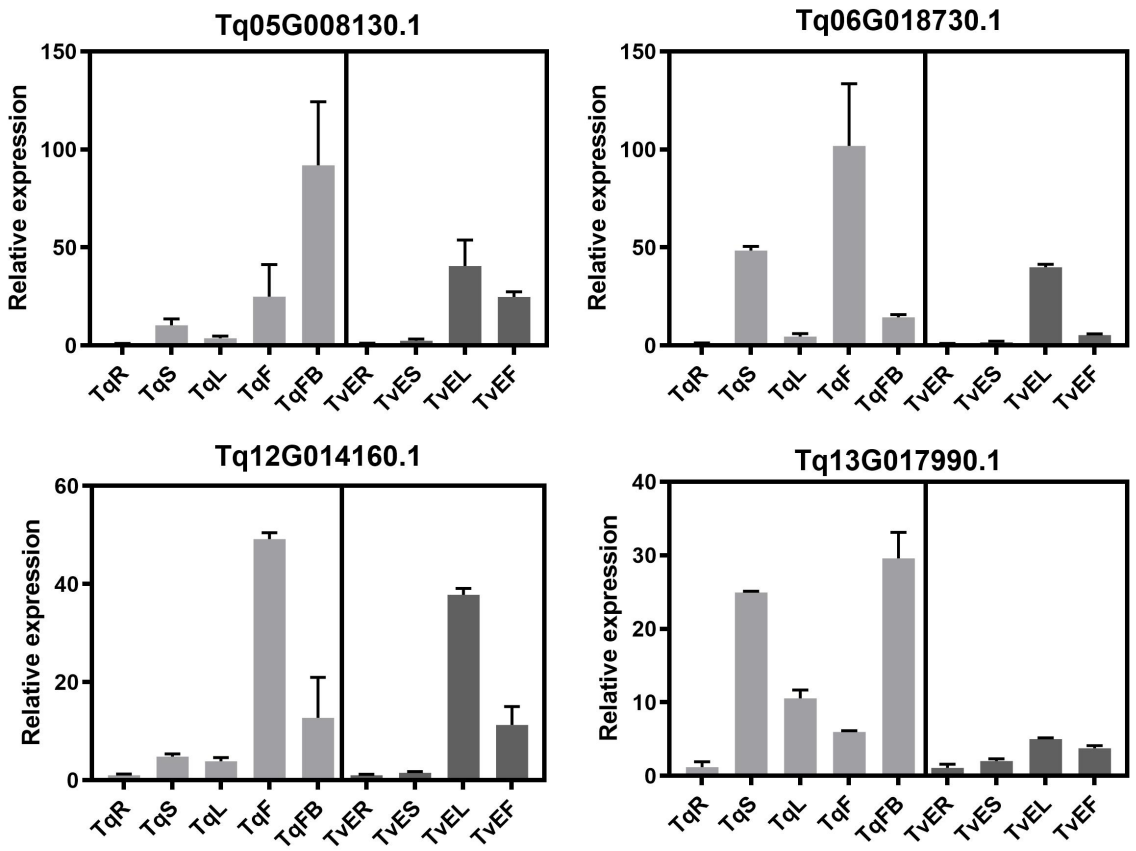
Supplemental Figure 15. Chromatogram of γ-terpinene, thymol, and carcacrol chemical standards. The X-axis represents the retention time of the peak outflow, and the Y-axis represents the response value of the chromatographic peak.



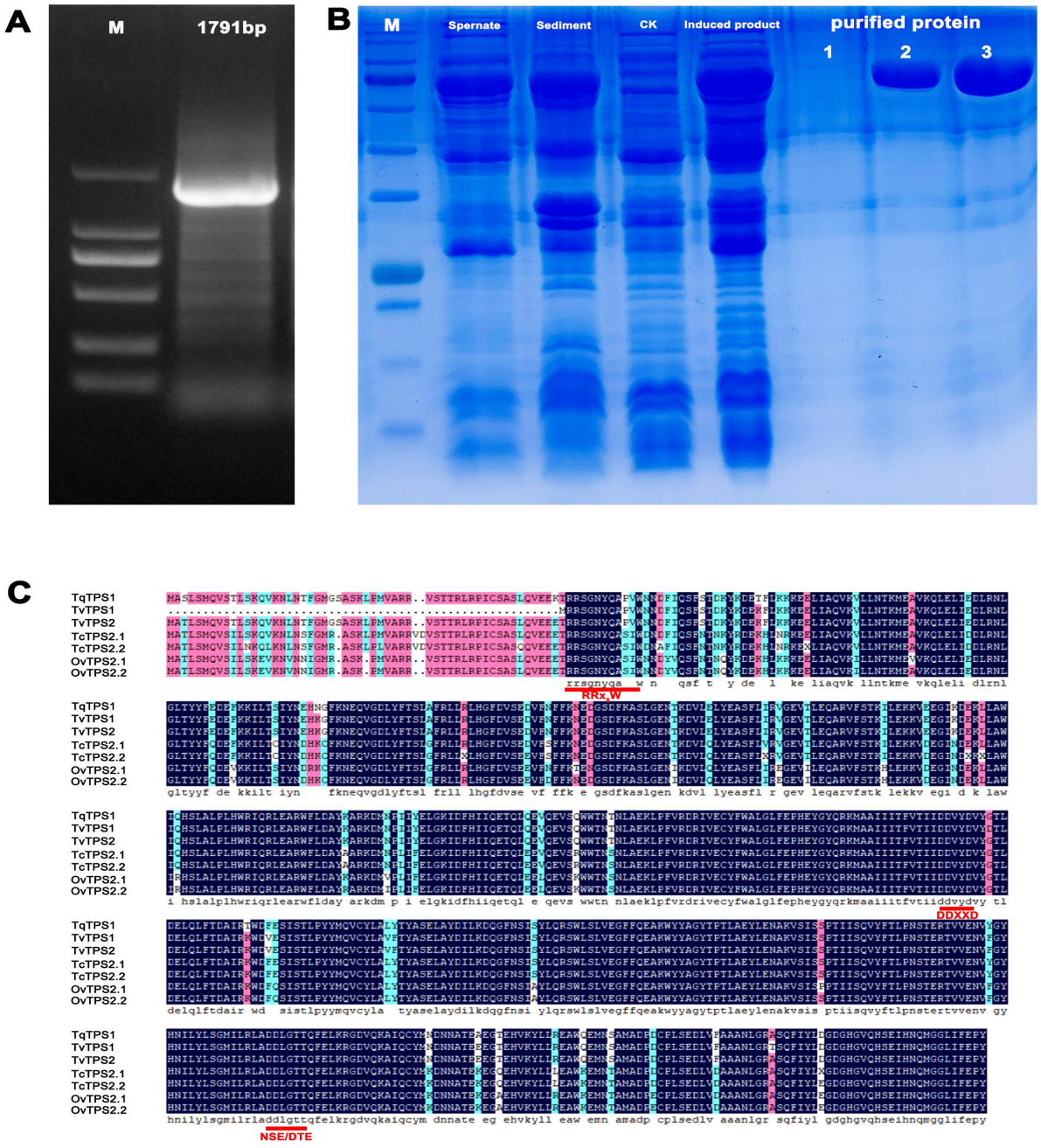
Supplemental Figure 16. Expression heatmap of SDR- and CYP-encoding genes involved in monoterpenoids biosynthesis. (A) Heatmap of SDR-encoding genes. **(B)** Heatmap of CYP-encoding genes. TqR, *T. quinquecostatus* root; TqS, *T. quinquecostatus* stem; TqL, *T. quinquecostatus* leaf; TqF, *T. quinquecostatus* flower; TqFB, *T. quinquecostatus* flower bud stage; TqFH, *T. quinquecostatus* flower half-bloom stage; TqFF, *T. quinquecostatus* flower full-bloom stage; TvER, *T. vulgaris* 'Elsbeth' root; TvES, *T. vulgaris* 'Elsbeth' stem; TvEL, *T. vulgaris* 'Elsbeth' leaf; TvEF, *T. vulgaris* 'Elsbeth' flower. The bar at the right top marked the normalized FPKM values for both in *T. quinquecostatus* and *T. vulgaris* 'Elsbeth' (red denoted high expression and blue indicated low expression, and expression data were Z-score standardized to -3 to 3 per gene.). All of expression genes showing minimum two-fold changes at $p < 0.05$ (Student's two tailed t-test).



Supplemental Figure 17. Identification of co-expression network modules in thymes. (A) Network heat map plot. Branches in the hierarchical clustering dendrograms correspond to modules. Color-coded module membership is indicated by the colored bars below and to the right of the dendrograms. In the heat map, high co-expression interconnectedness is indicated by progressively more saturated yellow and red colors. Modules correspond to blocks of highly interconnected genes. Genes with high intramodular connectivity are located at the tip of the module branches because they display the highest interconnectedness with the rest of the genes in the module. **(B)** Module eigengene adjacency heat map. The heat map shows the relatedness of the 17 co-expression modules identified in WGCNA, with red indicating highly related and blue indicating not related.



Supplemental Figure 18. Quantitative real-time PCR analysis of the *Tq05G008130.1* (*HD-ZIP*), *Tq06G018730.1* (*CYP*), *Tq12G014160.1* (*MYB*), and *Tq13G017990.1* (*SDR*) transcript levels in different tissues of *T. quinquecostatus* and *T. vulgaris* 'Elsbeth', respectively. Amplification of the 18S rRNA gene was used as an internal control. Bars represent the mean and SD of values obtained from three biological replicates.



Supplemental Figure 19. Cloning and *in vitro* gene expression in vitro of CDS in *TqTPS1* and amino acid sequence alignement of other TPS proteins. (A) PCR product of CDS in *TqTPS1*. **(B)** Overexpression of *TqTPS1* analyzed by SDS-PAGE**. (C)** Amino acid sequence alignment of TPS proteins from *T. quinquecostatus*, *T. caespititius*, *T. vulgaris*, and *Origanum vulgare*. The conserved motifs (RRX_8W, DDXXD and NSE/DTE) are highlighted with red lines.