

Genetic variation in a tandemly duplicated TPS gene cluster contributes to the diversity of aroma in lychee fruit

Huimin Hu^{1*}, Hongsen Liu^{1*}, Zaohai Zeng¹ , Yaxuan Xiao¹, Yingxiao Mai¹, Yanqing Zhang^{1,2}, Blake C. Meyers³ , Yanwei Hao¹  and Rui Xia¹ 

¹Guangdong Basic Research Center of Excellence for Precise Breeding of Future Crops, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in (South China) at Ministry of Agriculture and Rural Affairs, College of Horticulture, South China Agricultural University, Guangzhou, Guangdong, 510642, China; ²College of Agriculture, Guangxi University, Nanning, Guangxi, 530004, China; ³Department of Plant Sciences, University of California Davis, Davis, CA 95616, USA

Summary

Authors for correspondence:

Yanwei Hao

Email: yanweihao@scau.edu.cn

Rui Xia

Email: rxia@scau.edu.cn

Received: 4 November 2024

Accepted: 4 March 2025

New Phytologist (2025)

doi: 10.1111/nph.70090

Key words: farnesol synthesis, fruit aromatic diversity, genetic variation, lychee, tandem duplication, terpene synthase.

- Fruits undergo a similar ripening process, yet they exhibit a range of differences in color, taste, and shape, both across different species and within the same species. How does this diversity arise? We uncovered a conserved fruit ripening process in lychee fruit in which a NAC transcription factor, LcNAC1, acts as a master regulator. LcNAC1 regulates the expression of two terpene synthase genes, *LcTPSa1* and *LcTPSa2*, which belong to a gene cluster consisting of four *TPS* genes. *LcTPSa1*–*LcTPSa3* are responsible for catalyzing the production of farnesol, which in turn dictates the aromatic diversity in fruit of different lychee varieties.
- Through comparative, transcriptomic, and genomic analyses across various lychee varieties, we found these four *TPS* genes exhibit distinct expression levels due to natural genetic variation. These include copy number variations, presence/absence variations, insertions and deletions, and single nucleotide polymorphisms, many of which affect the binding affinity of LcNAC1.
- A single nucleotide mutation in *LcTPSa1* caused a premature translational termination, resulting in a truncated version of the TPS protein, which surprisingly remains functional.
- All these genomic changes in the LcNAC1-regulated *TPS* genes are likely to contribute to the great aromatic diversity observed in lychee fruit. This diversification of fruit aroma in lychee varieties offers a compelling example of how species- or variety-specific traits evolve – the phenotypic diversity is primarily derived from natural genetic variation accumulated in downstream structural genes within an evolutionarily conserved regulatory circuit.

Introduction

Fruits are essential for the reproduction of flowering plants and serve as a critical source of nutrition in a modern healthy diet (Maupilé *et al.*, 2024). Angiosperm fruits exhibit great diversity in nature, which derives both from different fruit types (including dry and fleshy fruits) and from variation in sizes, shapes, colors, textures, and flavor properties (Xiang *et al.*, 2024). The fruit flavor is mainly formed during the late fruit developmental stage, usually termed as ripening, which is a typical physiological process of fleshy fruits. This process is always accompanied by a series of complex alterations in fruit softening, conversion of starch to sugars, accumulation of pigments, emission of volatile aromatic compounds, and biosynthesis of essential nutrients (Giovannoni *et al.*, 2017; Cao *et al.*, 2022). Although fruits vary greatly in fruit color, aroma, and texture, most undergo similar processes of ripening, suggesting the presence of conserved

regulatory mechanisms. Ethylene is regarded as the key hormone that regulates fruit ripening (Zhu *et al.*, 2024). Transcription factors such as MADS-box (encoded by *MCML*, *Agamous*, *Deficiens*, and *SRF-box* genes), ethylene responsive factors, and the NAC family (NAM, ATAF1/2, and CUC1/2) regulate fruit softening, color alterations, and flavor formation by influencing ethylene production and ethylene response in apples, strawberries, and tomatoes (Li *et al.*, 2019; Deng *et al.*, 2022; Li *et al.*, 2023; Wang *et al.*, 2024). In the nonripening tomato mutant, a truncated NAC resulting from a mutation inhibits the activation of *SLACS2*, reducing ethylene biosynthesis and consequently affecting fruit softening, lycopene accumulation, and other ripening characteristics (Gao *et al.*, 2020).

Although diverse types of fruit undergo similar ripening processes, each fruit produces unique flavor compounds during the late stages of ripening, which are largely associated with the specific activation of flavor-related genes (Colantonio *et al.*, 2022; Cao *et al.*, 2024). The expression of these genes is usually

*These authors contributed equally to this work.

associated with genomic variation (Mishra *et al.*, 2024). For instance, the durian fruit is known for its unique odor and overpowering flavor of which is determined by the upregulation of sulfur-related genes and genomic expansion of genes affecting the accumulation of volatile sulfur compounds (Teh *et al.*, 2017). In cucumber, a 699-bp insertion and 10 single nucleotide polymorphisms (SNPs) in the regulatory region of transcription factor Bt (bitter fruit) affect fruit bitterness (Shang *et al.*, 2014). Furthermore, a SNP in the promoter of terpene synthases 1 (*AaTPS1*) affects the binding affinity of NAC protein, resulting in a higher accumulation of terpenes in certain kiwifruits (Nieuwenhuizen *et al.*, 2015). Therefore, natural variation is a great source of flavor diversity across different fruits.

Lychee (*Litchi chinensis* Sonn.) is an economically important fruit crop cultivated in subtropical and tropical regions; lychee is widely popular with consumers due to its unique aromas and nutritional richness (Hu *et al.*, 2022). The lychee fruit is a true fruit, derived from a superior ovary, and its edible flesh is the aril that develops from the outer integument of the ovule (Fan *et al.*, 2021). Lychee cultivars display significant disparities in the characteristics of their fruits, including shape, color, and flavor (Hu *et al.*, 2022). However, research on the development of lychee fruit and the genetic intricacies underlying their diversity remains relatively limited. The aroma of the lychee aril arises mainly at the fruit ripening stage (as with most other fruits), with significant variation in the types and quantities of terpenoids in lychee fruits of aromatic diversity (Liu *et al.*, 2022). Farnesol is one of the main components of lychee fruit aroma (Hao *et al.*, 2007; Wu *et al.*, 2009; Li *et al.*, 2010). In this study, we profiled a conserved ripening process controlled by a master regulator (LcNAC1) through comprehensive transcriptome and comparative genomic analysis. We further demonstrated that a cluster of *TPS* genes, which control farnesol biosynthesis, is regulated by LcNAC1, and diverse natural genetic variation within these *TPS* loci influences the enzymatic activity of TPS or the regulatory capacity of LcNAC1, contributing to the great aromatic diversity in lychee fruit.

Materials and Methods

Plant materials

Litchi chinensis Sonn. cv ‘Hemaoli’ (‘HML’), utilized for transcriptomic analysis, was grown at the Fruit Tree Research Institute of the Guangdong Academy of Agricultural Sciences, located in Guangdong Province, China. Three lychee trees with consistent growth potential were chosen to serve as three biological replicates. Fruits were harvested at 13 distinct developmental stages: 0 d after pollination (DAP), 3 DAP, 10 DAP, 17 DAP, 24 DAP, 31 DAP, 38 DAP, 45 DAP, 52 DAP, 59 DAP, 66 DAP, 73 DAP, and 80 DAP. This fruit was then dissected into different tissues including ovary, pericarp, seed, and aril (Supporting Information Table S1) before being frozen. Ripe lychee fruit from 37 different varieties was collected from Dongguan Botanical Garden (Guangdong Province). The harvested ripe lychee fruit was randomly divided into three groups, each serving as a biological replicate. Aril samples

were cut into small pieces and immediately frozen in liquid nitrogen for subsequent analysis.

RNA-Seq and data analysis

The aril RNA was extracted using the RNAPrep Pure Plant Plus kit (Polysaccharides & Polyphenolics-rich) (Tiangen, Beijing, China), while the total RNA of the ovary, seed, and pericarp from ‘HML’ was extracted using PureLink™ Plant RNA purification Reagent (Invitrogen, Code No.: 12322012). High-quality RNA (1 µg) from 90 high-quality samples was sent to BioMaker (Beijing, China) for RNA sequencing (RNA-seq) on a NovaSeq 6000 platform. The low-quality bases were removed using the FASTP software (Chen *et al.*, 2018). The STAR software was used to align all sequence data to the reference genome of lychee (Dobin *et al.*, 2012). The expression of lychee genes was quantified using the STRINGTIE software (Pertea *et al.*, 2015). The heatmap was constructed with TBTOOLS (Chen *et al.*, 2023).

Identification of gene co-expression module

The average fragments per kilobase of exon model per million (FPKM) values of genes from samples of the same tissue and developmental stage were calculated to represent gene expression levels. Genes with low expression levels and minimal variation across all samples were filtered out (mean FPKM < 2 and SD < 0.1). The construction of the network and detection of modules were performed using the R-based weighted gene co-expression network analysis (WGCNA) package (v.1.71) (Langfelder & Horvath, 2008). The soft threshold power was determined using the pickSoftThreshold function. The blockwiseModules function was employed to identify gene co-expression modules (parameters: TOMType = ‘unsigned’; minModuleSize = 80; reassignThreshold = 0; pamRespectsDendro = FALSE). The resulting gene co-expression modules were visualized with the CLUSTERGVIS package (v.0.1.1, <https://github.com/junjunlab/ClusterGVis>) to illustrate gene clustering and expression patterns.

Synteny analysis of the *LcNAC1* gene across 21 angiosperm species

Protein-coding genes from 21 angiosperm species (Table S2) were analyzed to identify gene family groups. The longest transcripts for each gene model were retained to reduce redundancy resulting from alternative splicing variation. ORTHOFINDER (v.2.5.4) (Emms & Kelly, 2019) was utilized to identify orthologous gene groups. The single-copy orthogroups for each species were collected and aligned using MUSCLE (v.5.1) (Edgar, 2004), and these alignments were concatenated to create a super alignment matrix. Four fossil calibration times, obtained from the TIMETREE website (<http://www.timetree.org/>) (Kumar *et al.*, 2022), were used to estimate species divergence times: 151.6–170.1 MYA between *Liriodendron tulipifera* and the Mesangiospermae group, 111.4–123.9 MYA for the Pentapetalae group, 109.8–124.4 MYA between *Vitis vinifera* and the Rosids group, and 68.0–85.4 MYA between *Citrus sinensis* and the Sapindaceae group. Using the constructed

phylogenetic tree and orthogroups alignment matrix, the MCMCtree program within the PAML (v.4.9j) (Yang, 2007) package was applied to estimate the divergence time. The MCMCtree running parameters were set as follows: model: HKY85; burn-in: 2000, and sample no.: 20000. Gene synteny analyses were performed using JCVI (v.1.2.7) (Tang *et al.*, 2024). The targeted *LcNAC1* gene and its neighboring genes were selected as initial candidates for exploring conserved evolutionary synteny blocks. Finally, a microsynteny plot was generated using the 'synteny' function with default parameters.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of *LcNAC1*

The protein sequences of the coding genes were extracted from the lychee genome, and gene functional annotation was performed using EGGNOG-MAPPER v.2.1.12 (Cantalapiedra *et al.*, 2021). The alignment parameters used were as follows: --seed_ortholog_evalue 0.001 --score 60 --pidient 40 --query_cover 20 --subject_cover 20. The eggNOG-mapper Helper tool in TBTOOLS (Chen *et al.*, 2023) was subsequently employed to transform the functional annotation outcomes (Table S3). By employing TBtools' KEGG enrichment analysis feature in conjunction with KEGG plants backend background files, and by utilizing the coding genes from the entire genome as the reference, functional enrichment analysis was performed on the 504 NAC target genes within the M2 module.

Identification and characterization of gene family

Two profiles of Hidden Markov Models for specific TPS domains, the Terpene Synthase N-terminal domain (PF01397) and the Terpene Synthase C-terminal domain (PF03936), were retrieved from the Pfam database (v.36.0) (Mistry *et al.*, 2020). The *LcTPS* genes were retrieved from the lychee cv 'FZX' genome database using HMMER (v.3.3.2) (Potter *et al.*, 2018). Protein sequences containing these two domains ($E < 1e-05$) were subsequently confirmed through manual correction using IGV-GSAMAN (v.0.7.11) (Chen *et al.*, 2021). Multiple sequence alignment of the *LcTPS* genes was performed using the MAFFT software (v.7.505) (Kato & Standley, 2013) with default parameters. The TRIMAL software (v.1.4.rev15) (Capella-Gutiérrez *et al.*, 2009) was employed to automatically trim poorly aligned regions. The final alignment was utilized to construct a maximum likelihood phylogenetic tree using the IQTREE software (v.2.2.0.3) (Minh *et al.*, 2020) and Chiplot (Xie *et al.*, 2023).

Evaluation of transcription binding sites

To identify downstream target genes regulated by the *LcNAC1* gene, we obtained the NAC002 binding motif profile (MA2015.1) from the JASPAR database (<http://jaspar.genereg.net/>) (Raulusevičiute *et al.*, 2023). The FIMO program, part of the MEME SUITE software (v.5.5.6) (Bailey *et al.*, 2015), was utilized to scan the promoter sequences (1 kb upstream of the translation start site: ATG) of genes within the same co-expression module.

Analysis of subcellular localization

The recombinant 35S-*LcNAC1*-GFP vectors were constructed using primers listed in Table S2 and were transformed into *Agrobacterium tumefaciens* GV3101::pSoup for transient expression in tobacco (*Nicotiana benthamiana*) leaves. The vector was infiltrated into transgenic tobacco leaves expressing a red fluorescent nuclear marker (Nucleus-RFP). After 48-h infiltration, the leaves were detached for analysis using a confocal laser scanning microscope (LSM 800; Carl Zeiss, Oberkochen, Germany). The acquired images were processed with the LSM Image Browser (Carl Zeiss). The experiment was conducted in triplicate.

Dual-luciferase reporter assays

According to the previously described protocol (Zhang *et al.*, 2024), full-length coding sequences (CDSs) of the transcription factor (Li *et al.*, 2024) were cloned into the pGreen II 62-SK vector as the effector, and the promoters of *LcTPSa1–LcTPSa3* were cloned into the pGreen II 0800-LUC vector as the reporters using the primers listed in Table S2. The above effector and reporter constructs were transformed into *A. tumefaciens* GV3101::pSoup, and the bacteria were injected into tobacco leaves for transient expression assays. After 3d infiltration, the leaf zones of infiltration were harvested for enzyme activity assays of firefly luciferase and renilla luciferase using the DualLuciferase[®] Reporter Assay System (Yeasen, Shanghai, China). At least six independent biological replicates were performed.

Yeast one-hybrid assay

The promoters of the *LcTPSa1–LcTPSa3* were cloned into the pAbAi vector, and the construct was integrated into the genome of the Y1HGOLD yeast strain. The background aureobasidin A resistance (AbA^r) expression of the Y1HGOLD *LcTPSa1–LcTPSa3–pAbAi* strain was tested on a selective synthetic dextrose (SD) medium, uracil. Then, the full length of *LcNAC1* was cloned into the pGADT7 vector for identification. After determining the minimal inhibitory concentration of AbA for the bait strains, the pGADT7-*LcNAC1* prey vectors were transformed into the bait strain and screened on an SD/–Leu/AbA plate. Individual bait–prey interactions were performed to verify the positive recombinant prey vector. All transformations and screenings were examined at least three times. Primers used in this assay are listed in Table S2.

Electrophoretic mobility shift assay

The CDS of *LcNAC1* was cloned into the pGEX-4T-1 vector using the primers listed in Table S2 and then was expressed in *Escherichia coli* (*E. coli*) Rosetta (DE3). Expression and purification of the recombinant protein were performed according to the GST-tag Protein Purification Kit (Beyotime, Shanghai, China) manufacturer's instructions. The electrophoretic mobility shift assay (EMSA) was conducted using the LightShift[®] Chemiluminescent EMSA Kit (Thermo Fisher Scientific, Waltham, MA,

USA) (Zhang *et al.*, 2024). The double-stranded probes with 3' biotin labeling were made by annealing separately synthesized strands. The probes used for EMSAs are listed in Table S2.

Real-time quantitative PCR

Real-time quantitative PCR was conducted with Promega GoTaq[®] qPCR Master Mix (A6001) in a BioRad CFX384 Real-Time PCR Detection System, with each assay being replicated three times both biologically and technically. The *LcGAPDH* and *LcEF1 α* were used as controls for lychee mRNA samples, and the *LcTPSa1–LcTPSa3* primers were designed by using primer3 (<http://frodo.wi.mit.edu/primer3>) and are described in Table S2. The specificity of the primers was tested with melting curves and resequencing of PCR products. The relative expression was calculated using the comparative $2^{-\Delta\Delta C_t}$ method (Pfaffl, 2001).

Heterologous expression in *E. coli* and enzymatic activity assay

The complete open reading frame (ORF) sequences of *LcTPSa1* and *LcTPSa2* genes (excluding the termination codon) were cloned into the pET-32a expression vector using the primers in Table S2. After sequence validation, the pET-32a vector was transformed into *E. coli* Rosetta (DE3). Recombinant protein was purified by using the HisTALON[™] gravity column (TransGen Biotech, Beijing, China) following the manufacturer's instructions. Purified *LcTPSa1* and *LcTPSa2* proteins were examined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and were visualized by Coomassie Brilliant Blue.

Enzyme assay buffer (30 mM HEPES, 25 mM MgCl₂, and 50 mM dithiothreitol) was utilized in the reaction with farnesyl pyrophosphate (FPP) and geranyl pyrophosphate as substrates. The reaction was carried out in a 2-ml sealed glass bottle; 50 μ g of purified protein, 60 μ M of reaction substrate, and the previously mentioned enzyme reaction buffer were added to 200 μ l of the reaction system. The reaction was performed at 30°C for 2 h. The protein expressed by the empty vector and the substrate were heated separately as a negative control. The product catalyzed by the enzyme was collected by solid-phase microextraction (SPME) before gas chromatography mass spectrometry (GC-MS) analysis.

Transient expression of *LcTPSa1*, *LcTPSa2*, and *LcNAC1* in *N. benthamiana*

The full-length CDSs of *LcTPSa1*, *LcTPSa2*, and *LcNAC1* genes were cloned into the modified pEAQ-HT-DEST-1 expression vector using the primers in Table S2, respectively. The modified pEAQ-HT-DEST-1 vector used for transient expression studies in *N. benthamiana* was kindly donated by Dr Zhenhua Liu at Shanghai Jiao Tong University. Expression vectors were introduced into the *A. tumefaciens* strain LBA4404. The bacteria were resuspended in infiltration buffer (10 mM MES, 10 mM MgCl₂, 150 μ M acetosyringone, and pH 5.6) at OD₆₀₀ = 0.65 and incubated for 3 h at

room temperature. After 5 d, infected leaves of *N. benthamiana* were collected by SPME before GC-MS analysis.

Protein structure model prediction and substrate docking simulation

The protein structure models of *LcTPSa1* and *LcTPSa2* were generated using the AlphaFold3 online tool (<https://alphafoldserver.com/>) (Abramson *et al.*, 2024). A Mg²⁺–FPP complex was incorporated into the modeled *LcTPSa2* by aligning it with a distinct TPS structure (PDB entry 5IK0) (Koo *et al.*, 2016) as a template. The AUTODOCK software (v.4.2.6) (Morris *et al.*, 2009) was utilized to prepare and parameterize the receptor protein *LcTPSa2* and the ligand Mg²⁺–FPP complex. Docking grid documents were generated by using the AutoGrid of the sitemap, and AUTODOCK was employed for docking simulation. The optimal pose for interaction analysis was selected, and the protein–ligand interaction diagram was generated with PyMOL (v.2.5.7, <http://www.pymol.org/pymol>).

Free and bound volatiles analysis by GC-MS

Glycosylated volatiles were extracted according to Wu *et al.* (2018). Frozen lychee fruits (5 g) and tobacco leaves (1 g) were ground and homogenized in 30 ml of saturated salt solution. After centrifuging for 20 min at 13 000 g, the supernatant was used as the crude extract. Isolation of glycosidic precursors was conducted by using SPE LC-18 resins (CNW, Duesseldorf, Germany). Free volatile compounds were removed by washing with 25 ml of dichloromethane, and the bound fraction was eluted with 25 ml of methanol. The bound volatile compounds were enzymatically hydrolyzed at 40°C after adding 2 mg of β -d-glucoside glucohydrolase (CAS: 9001-22-3; Sigma-Aldrich) according to previous studies (Bönisch *et al.*, 2014; Yauk *et al.*, 2014). The free aglycones were released for 30 min at 45°C and collected using a SPME fiber coated with 100 μ m of polydimethylsiloxane and divinylbenzene (PDMS-DVB, 3 pk (Red); Supelco Inc., Bellefonte, PA, USA). The released volatiles were identified using GC-MS.

Frozen lychee fruits (5 g) were ground and homogenized in a saturated salt solution. Ethyl caprate (0.864 mg ml⁻¹) was used as an internal standard. Free volatiles were extracted using fibers coated with 100 μ m of PDMS-DVB (Supelco Co.) for 30 min at 45°C and identified using GC-MS as mentioned previously.

An Agilent 7890B gas chromatograph coupled with an Agilent 5977A mass spectrophotometer (Agilent, Palo Alto, CA, USA) equipped with a HP-5MS column (0.30 mm, 30 m, 0.25 μ m; J&W Scientific, Folsom, CA, USA) was applied for the identification of volatile compounds according to methods described above. Helium was used as a carrier gas at a flow rate of 1.0 ml min⁻¹. The temperature program started at 40°C and was increased by 8°C min⁻¹ to 100°C and then to 240°C at 10°C min⁻¹. The column effluent was ionized by electron ionization (EI) at an energy of 70 eV and a source temperature of 230°C. Mass scanning was performed over the range 35–550 aum. Volatile compounds were identified by comparing their EI mass

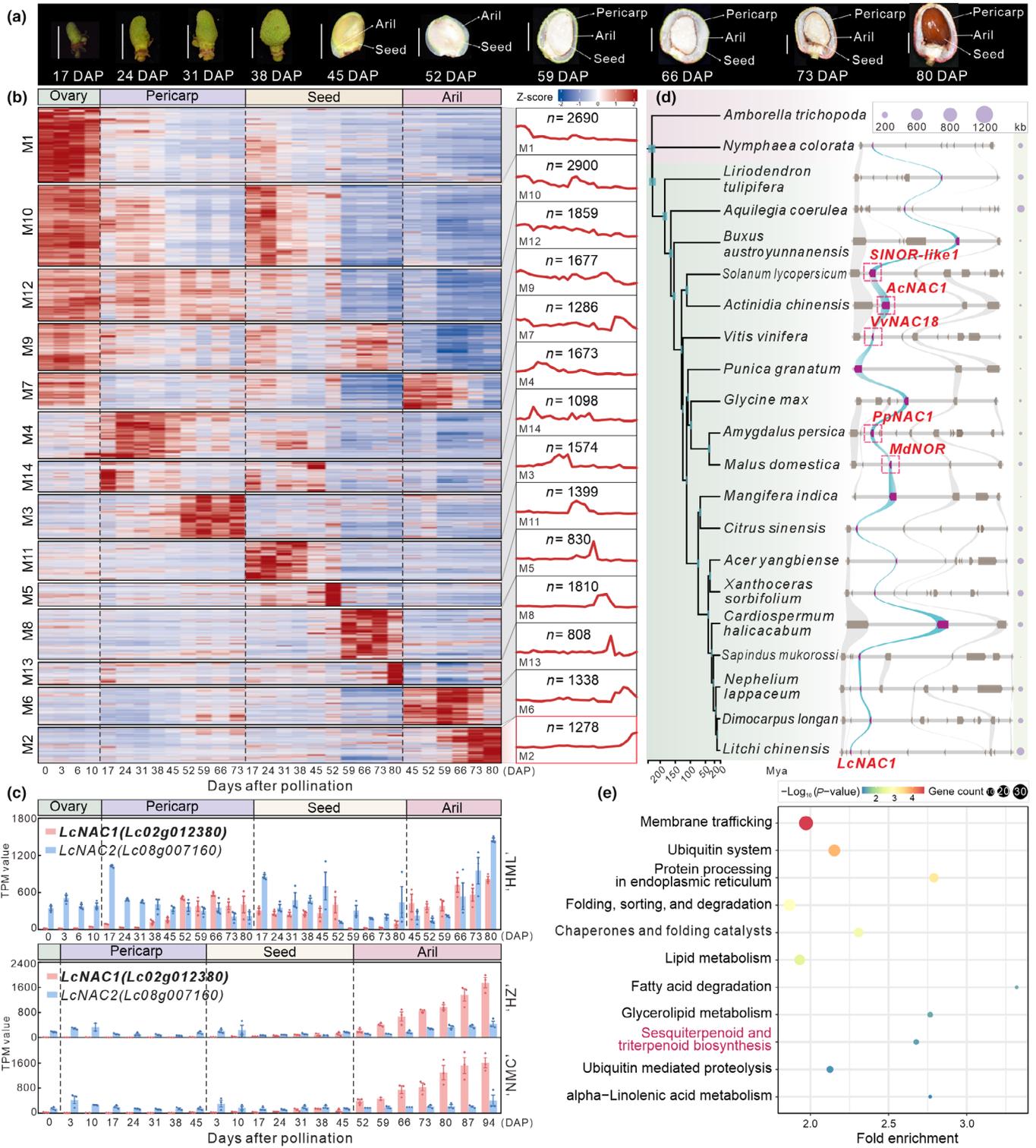


Fig. 1 *LcNAC1* acts as a key regulator during lychee fruit ripening. (a) Photograph of lychee fruit cv 'Hemaoli' (HML) at different developmental stages. DAP, days after pollination. Bars: 1 cm. (b) The relative expression of genes in 14 co-expression modules during lychee fruit development and ripening. The line plots illustrate the common expression pattern of genes in different modules. *n*, the number of genes in modules. (c) Expression profile of *LcNAC1* (Lc02g012380) and *LcNAC2* (Lc08g007160) genes during fruit development and ripening of lychee cv 'HML', 'Huaizi', and 'Nuomici'. TPM, transcripts per million. Data were presented as the mean ± SE. (d) Phylogenetic tree and conserved *LcNAC1* gene blocks in the genomes of 21 plant species. The axis denotes the divergence time of each node (Ma, million years ago), and the green bars indicate the 95% confidence intervals of divergence times in millions of years. The dashed rectangles depict the gene that has been reported in the published literature. The dot size denotes the genome size of each gene block. (e) The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the genes in Module 2, which are predicted to harbor a NAC binding site in their promoters.

spectra with the NIST Mass Spectral Library (NIST14.L) and the retention time of authentic standards.

Statistical analysis

The significance of differences between two samples was assessed by using a two-sided Mann–Whitney test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$) (SPSS 19.0; SPSS Inc., Chicago, IL, USA). Figures were generated with GRAPH PAD PRISM 8.0 (GraphPad Software; San Diego, CA, USA) and Origin 2021 (Origin Lab Corp., Northampton, MA, USA).

Results

LcNAC1 is a conserved master regulator in lychee fruit ripening

Lychee is an exotic tropical and subtropical drupe with a distinct process of fruit development (Zhao *et al.*, 2020; Zhang *et al.*, 2024). The aril (the flesh that is consumed) does not develop until 45 DAP (Fig. 1a; Table S1). In this study, fruits (separated into ovary, pericarp, aril, and seed) at 10 different stages from flowering to fruit ripening of lychee cv ‘HML’ were collected for a comprehensive transcriptome analysis (Fig. 1a). In total, 90 samples were used for RNA sequencing and 22 220 genes were detected with a considerable expression in lychee fruit development (mean FPKM > 2 and SD > 0.1 across all samples). Next, these genes were classified into expression-correlated groups by a WGCNA. A total of 14 co-expression modules were detected (Fig. 1b), with the number of genes in each module ranging from 808 (Module 13) to 2900 (Module 10) (Table S4). Three modules (M2, M6, and M7) exhibited high expression in aril samples. The M2 genes were predominantly expressed in the aril, and their level gradually increased from the peak until harvest (66–80 DAP) (Fig. 1b), suggesting that genes in this module are highly associated with lychee fruit ripening. Notably, there were two NAC genes, namely, *LcNAC1* (*Lc02g012380*) and *LcNAC2* (*Lc08g007160*), among the 35 highly expressed transcription factor genes (total FPKM > 200 across all aril samples) in this module (Table S5). These TF genes were highly expressed also in aril development in two other late-maturing lychee cultivars, ‘Huaizi’ (‘HZ’) and ‘Nuomici’ (‘NMC’) (Figs 1c, S1, S2), and *LcNAC1* exhibited significantly higher expression than *LcNAC2*, consistent with a prominent role of LcNAC1 in lychee fruit ripening (Fig. 1c).

Interestingly, synteny analysis revealed that *LcNAC1* was located in a syntenic block that is highly conserved among many angiosperms (Fig. 1d; Table S6). *LcNAC1* is syntenic to the *NOR-like1* gene in tomato (*SINOR-like1*) and the *MdNOR* gene in apple, and all three genes belong to the same orthologous group in our orthogroup analysis (Fig. S3; Table S7). *SINOR-like1* is a key regulator in tomato fruit ripening, participating in ethylene synthesis, Chl degradation, carotenoid accumulation, and fruit softening processes (Gao *et al.*, 2018; Jia *et al.*, 2024), and *MdNOR* is involved in ethylene-mediated fruit ripening

regulation in apple as well (Giovannoni, 2004). Therefore, lychee fruit likely has a conserved process of fruit ripening controlled by an evolutionarily conserved NAC (*LcNAC1*) transcription factor, similar to many other fruits.

LcNAC1 activates aroma-related *LcTPSa1* and *LcTPSa2* expression in aril

As a proposed master regulator of lychee fruit ripening, what downstream target genes might be regulated by LcNAC1 in lychee? To answer this question, we performed enrichment analysis for the M2 genes bearing potential NAC binding sites (Fig. 1e; Table S8). These potential target genes were predominantly enriched in various molecule metabolic and transport pathways, such as membrane trafficking and amino acid metabolism, suggesting that LcNAC1 is likely involved in regulating a series of fruit ripening-related processes (Fig. 1e). Given that terpenoids are one of the main components of lychee fruit aroma, biosynthesized at the fruit ripening stage, one of the enriched categories, ‘sesquiterpenoid and triterpenoid biosynthesis’, caught our attention for further investigation. The category includes five genes, four of which encode TPS proteins (Table S9). TPS is a crucial terminal enzyme in the biosynthesis of terpenoids and is responsible for the generation of the structural diversity in the superfamily of terpenoid natural products (Jia *et al.*, 2022). Plant TPS genes are typically categorized into seven subfamilies, from *TPS-a* to *TPS-h* (Yan *et al.*, 2023). All four TPS genes belong to the *TPS-a* subfamily (Fig. S4; Table S10). *Lc02g019120* was expressed at a very low level, not only in fruits but also in other tissues such as pericarp at early stages (Figs S5–S7); by contrast, *Lc01g020180*, *Lc01g020190*, and *Lc01g020200* had remarkably high transcript levels in the aril of three lychee cultivars during the fruit ripening process, especially *Lc01g020190* (Figs 2a, S5–S7). Therefore, these three TPS genes, denoted as *LcTPSa1* (*Lc01g020180*), *LcTPSa2* (*Lc01g020190*), and *LcTPSa3* (*Lc01g020200*), were selected for further analyses.

There are three potential LcNAC1 binding sites in the promoter sequences of the *LcTPSa1*–*LcTPSa3* genes, with binding sequence and predicted binding capacity varied among genes (Table S8; Fig. S8). To confirm the regulation of LcNAC1, yeast one-hybrid (Y1H) validation experiments were first conducted and revealed that the LcNAC1 protein failed to interact with the *LcTPSa3* promoter, which harbors only the binding site 3 (Figs 2b, S8). This result suggests that binding site 2, which is predicted to be present only in *LcTPSa1* and *LcTPSa2* (Figs 2c, S8), instead of binding site 3, has NAC protein binding capacity. As anticipated, LcNAC1 was localized to the nucleus of *N. benthamiana* (Fig. 2d), and EMSA proved that LcNAC1 could directly bind to the binding site 2 of the promoters of *LcTPSa1* and *LcTPSa2* (Figs 2e, S9). Dual-luciferase reporter assay demonstrated that the binding of LcNAC1 activated the transcription of both *LcTPSa1* and *LcTPSa2* (Fig. 2f). Collectively, these results suggest that LcNAC1 activates the expression of *LcTPSa1* and *LcTPSa2* during the ripening of lychee fruit.

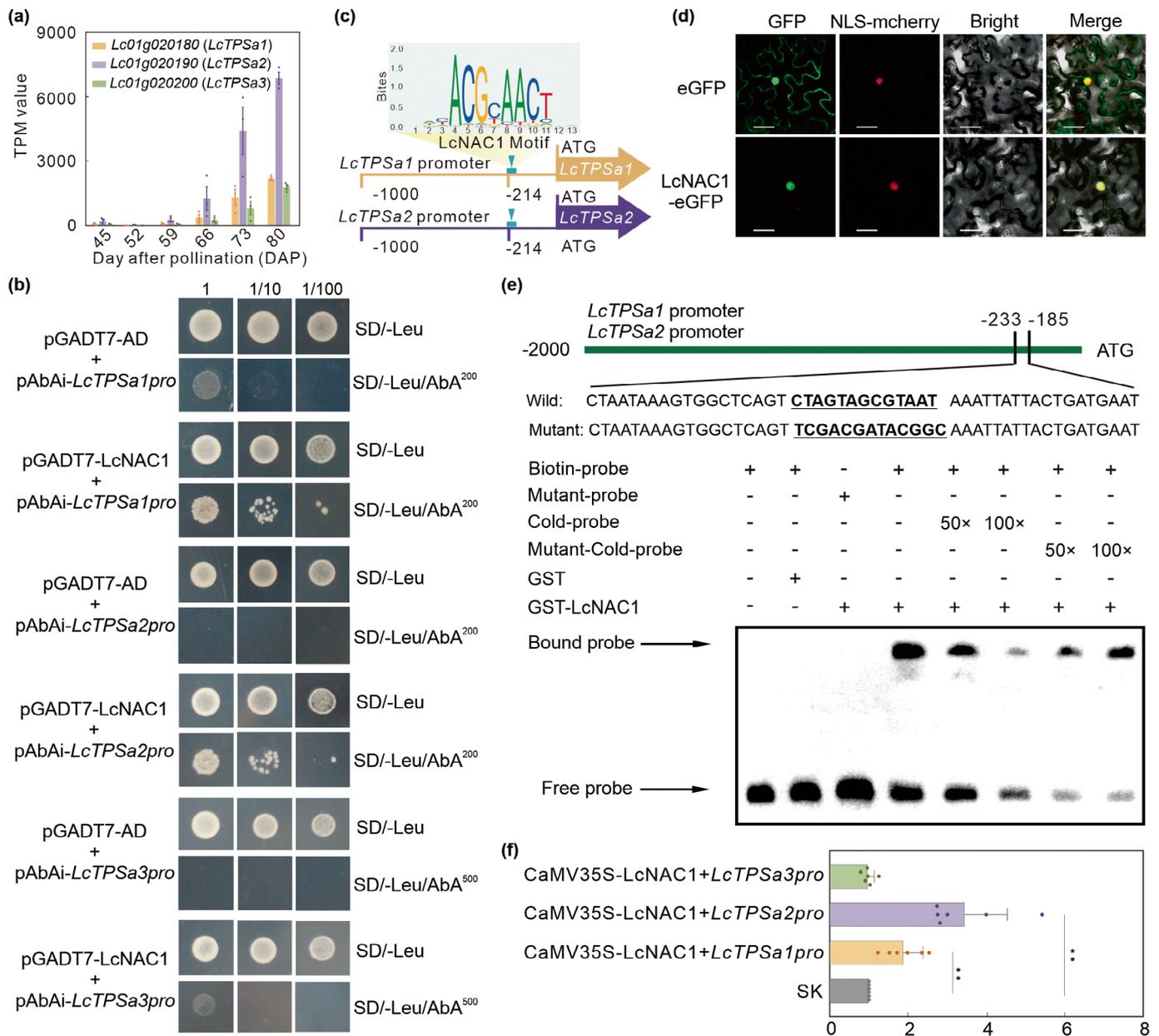


Fig. 2 LcNAC1 activates the transcription of *LcTPSa1* and *LcTPSa2*. (a) Expression pattern of *LcTPSa1*, *LcTPSa2*, and *LcTPSa3* genes during lychee fruit development and ripening. Transcripts per million was used to indicate gene expression levels. (b) Yeast one-hybrid assay revealing the binding of LcNAC1 to the indicated promoter fragments. pGADT7–LcNAC1 served as a prey, and pAbAi–*LcTPSa1pro*, *LcTPSa2pro*, and *LcTPSa3pro* were used as baits. Empty pGADT7 served as a negative control. The transformants were selected on SD/–Leu medium supplemented with varying concentrations of aureobasidin A. Data were presented as the mean \pm SE. [Correction added on 1 May 2025, after first online publication: the authors identified that some of the images in (b) were duplicated. This has now been corrected in this updated figure.] (c) Schematic diagram for the specific LcNAC1 binding sites of *LcTPSa1* and *LcTPSa2*. Motif logo derived from JASPAR TF binding profile (MA2015.1) is provided above the green background. (d) Subcellular localization of LcNAC1 in *Nicotiana benthamiana* leaves. GFP, GFP channel; NLS-mcherry, transgenic tobacco plants with red fluorescence in the nucleus; Bright-field, light microscopy image; Merge, merged image of the GFP and RFP channels. Bars, 50 μ m. (e) Electrophoretic mobility shift assay shows that LcNAC1 directly binds to the motifs in the *LcTPSa1* and *LcTPSa2* promoters. Recombinant LcNAC1 (2 μ g) was incubated with biotin-labeled probes or an unlabeled DNA probe with intact (competitor) or mutated (mutant probe) binding motifs. (f) LcNAC1 activated the expression of *LcTPSa1* and *LcTPSa2* *in vivo*, as shown by transient dual-luciferase reporter assays. Means and SE were calculated from six replicates (**, *P* value < 0.01, two-sided Mann–Whitney test).

A cluster of tandemly duplicated *TPS* genes is responsible for the farnesol biosynthesis in lychee fruit

Further examination revealed another *TPS* gene, *LcTPSa4* (*Lc01g020210*) present alongside the *LcTPSa1*–*LcTPSa3*,

forming a four-gene, tandemly duplicated gene cluster in the ‘HML’ genome (Fig. 3a). Interestingly, we found that the gene copy number of this *TPS* cluster varied in the genomes of different lychee varieties, with the *LcTPSa1* gene absent in some genomes (Figs 3a, S10, S11). *LcTPSa1* was present in both

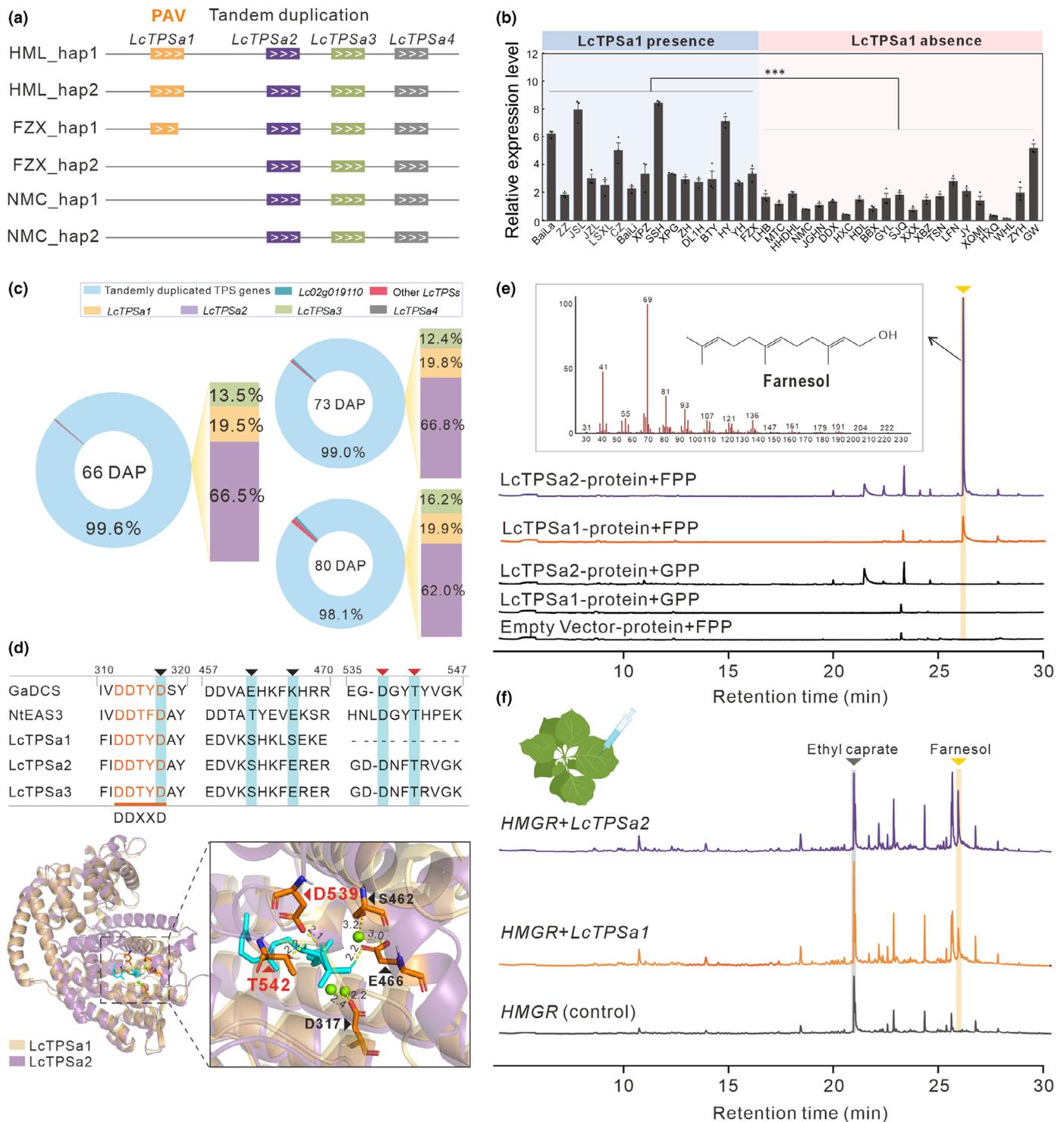


Fig. 3 Tandemly duplicated LcTPSs catalyze farnesol biosynthesis both *in vitro* and *in vivo*. (a) Diagram illustrating the distribution of tandemly duplicated *LcTPSa1–LcTPSa4* genes among the haplotype genomes of three representative varieties. HML, ‘Hemaoli’, FZX, ‘Feizixiao’, and NMC, ‘Nuomici’. Presence/absence variation (PAV) of *LcTPSa1* was shown. (b) Expression level analysis of total *LcTPSa1–LcTPSa3* using reverse transcriptional quantitative polymerase chain reaction across 37 lychee varieties. Error bars indicate SEs from three replicates, and the letters above the bars indicate a significant difference (***, P value < 0.001, two-sided Mann–Whitney test). (c) Percentage of *LcTPSa1–LcTPSa4* expression to the total expression of all *TPS* genes during lychee fruit ripening (66 d after pollination (DAP), 73 DAP, and 80 DAP). (d) Protein sequence analysis, homology modeling, and molecular docking of *LcTPSa1* and *LcTPSa2* with substrate magnesium–farnesyl pyrophosphate (Mg^{2+} -FPP) (FPP in cyan and magnesium in green). A magnified view of the docking model emphasizes the five amino acids in *LcTPSa2* that interact with Mg^{2+} -FPP and their associated hydrogen bonds. (e) Enzymatic activity assay of *LcTPSa1* (480 aa) and *LcTPSa2* (563 aa) protein with FPP as substrate *in vitro*. (f) Transient overexpression of *LcTPSa1* and *LcTPSa2* increases farnesol content in *Nicotiana benthamiana* leaves. The *N. benthamiana* leaves transformed with an empty vector (pEAQ) served as the control group (WT). *m/z*, mass-to-charge ratio.

haplotypes of ‘HML’ but was completely absent in a homozygous state in ‘NMC’ and hemizygous in ‘Feizixiao’ (‘FZX’), in which it is specifically present in FZX_hap1 but absent in FZX_hap2 (Fig. 3a). Indeed, this presence/absence variation (PAV) of the *LcTPSa1* gene impacted the total expression level of this *TPS* cluster, with the total expression level of *LcTPSa1–LcTPSa3* significantly higher in lychee varieties having *LcTPSa1* than in those without it (Fig. 3b). For these four clustered *TPS* genes, we noted that the *LcTPSa4* copy was not expressed at all (Fig. S8), likely due to the change in promoter sequences, and their expression of the other three (*LcTPSa1–LcTPSa3*) accounted for the majority of the expression of all lychee *TPS* genes (54 in total) during ‘HML’ fruit ripening (Figs 3c, S5), indicating their predominant role in terpene biosynthesis in ripe lychee fruit. Additionally, these three *TPS* genes exhibited varying levels of expression in ‘HML’ fruits, with the *LcTPSa2* showing the highest expression, accounting for > 60% of total *TPS* abundance, followed by *LcTPSa1* and *LcTPSa3*, contributing ~19% and ~13%, respectively, to the total expression (Fig. 3c). A similar expression pattern for these clustered *TPS* genes was observed in both ‘HZ’ and ‘NMC’ as well (Figs S5–S7, S12), with the exception of *LcTPSa1*, which is absent in these two cultivars. These expression differences suggest that these *TPS* genes are likely under distinct levels of transcriptional regulation. The absence of the LcNAC1 binding site 2 in the *LcTPSa3* promoter may explain the lower expression level of *LcTPSa3* compared with *LcTPSa1* and *LcTPSa2* (Fig. S8).

A functionally important difference was also identified between LcTPSa1 and LcTPSa2; that is, *LcTPSa1* bears a 1-bp deletion in Exon 7 compared with *LcTPSa2*, leading to a premature termination codon and a deletion of 82 amino acids at the C terminus of LcTPSa1 (Fig. S13). Molecular docking with a full-length LcTPSa2 protein revealed that five amino acids are critical for the correct docking of FPP, a substrate of the TPS enzyme, into the enzymatic site. By contrast, 539 Asp (D) and 542 Thr (T) in LcTPSa2 are missing in LcTPSa1 due to the premature translational termination, which may impact the binding affinity of TPS to the substrate (Figs 3d, S14). To explore whether this deletion affects or blocks its biological function, we conducted an *in vitro* enzymatic activity assay. We found that both the recombinant LcTPSa1 and LcTPSa2 enzymes could catalyze the conversion of FPP to farnesol (Figs 3e, S15; Table S11). Given the unavailability of a transgenic system in lychee, we further validated the enzymatic function by *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves. We found that leaves transiently expressing *LcTPSa1* and *LcTPSa2* produced a sesquiterpene profile very similar to that detected in the *in vitro* assay (Fig. 3f; Table S12). These results indicate that LcTPSa1, in spite of the deletion at the C terminus, still has the enzymatic activity capable of catalyzing farnesol biosynthesis. However, tobacco leaves expressing *LcTPSa1* yielded a much lower amount of farnesol than those expressing *LcTPSa2*, suggesting that the sequence deletion does affect LcTPSa1 activity (Fig. 3f). Previously, farnesol was reported to be a major contributor to lychee fruit aroma and flavor quality (Hao *et al.*, 2007; Li *et al.*, 2010). To confirm these earlier reports, ripe lychee fruits

from six different varieties were harvested for free and bound volatile terpene profiling, and farnesol was indeed found to be the main bound terpenoid compound in lychee fruit (Fig. S16; Table S13). The farnesol content in lychee cv ‘HZ’ fruits increased during fruit development and ripening, showing a positive correlation with the trend of *LcTPSa1–LcTPSa3* expression levels (Figs S6, S17). Taken together, the *TPS* genes in the cluster (*LcTPSa1*, *LcTPSa2*, and *LcTPSa3*) are likely the most important *TPS* genes responsible for the biosynthesis of farnesol in lychee fruit, and their genetic variation contributes to the aromatic diversity among lychee varieties.

Variation in the LcNAC1 binding site of *LcTPSa2* further drives the divergence of farnesol biosynthesis

The higher expression and greater enzymatic activity of LcTPSa2 suggest that it is the key factor essential for the formation of lychee aroma. The abundance of variation in the lychee genome and the remarkable aromatic diversity of lychee fruit across different cultivars (Hu *et al.*, 2022; Liu *et al.*, 2022) prompted us to scrutinize the genomic variation in the *LcTPSa2* locus among different lychee varieties. Intriguingly, we discovered three types of polymorphism, referred to here as G-type, A-type, and insertions and deletions (InDel), situated within the LcNAC1 binding site in the *LcTPSa2* promoter. For instance, a 7-bp deletion is present in the haplotype genome 1 (hap1) of ‘FZX’, A- and G-type in the hap1 and hap2 of ‘Baitangying’ (‘BTY’), respectively, and A-type in both haplotypes in ‘NMC’ (Fig. 4a). We further assessed the haplotypic expression level of *LcTPSa2* in fruits across five lychee varieties and found that there was a significant difference in allelic expression in the ‘FZX’ and ‘BTY’ accessions, in which the LcNAC1 binding sites are heterozygous (Figs S18, S19), whereas no significant expression difference was detected in other varieties with a homozygous LcNAC1 binding site. This suggests that the sequence polymorphism in the LcNAC1 binding site in the *LcTPSa2* promoter does affect the binding affinity of LcNAC1 and subsequently the expression of *LcTPSa2*. Specifically, the deletion in the hap1 of ‘FZX’ may prevent LcNAC1 binding, leading to the lower expression level, while the G-type allele in the hap2 of ‘BTY’ exhibited a relatively higher expression level than the allele of hap1 (Fig. 4b). To confirm the influence of the A/G variation on the binding affinity of LcNAC1, we conducted an EMSA experiment and found that LcNAC1 bound to both types of probes (G- and A-type) containing the NAC binding site *in vitro*, but the A-type probe showed a weaker band than the G-type probe (Fig. 4c). The different expression levels of *LcTPSa2* across various lychee cultivars can be traced back to a 7-bp deletion impeding the LcNAC1 binding to the *LcTPSa2* promoter, alongside an A/G mutation altering the efficiency of the LcNAC1 interaction with this regulatory site. Consequently, LcNAC1 could be a crucial regulator in the farnesol synthesis process in lychee fruit. To test this, we conducted a co-infiltration experiment involving *LcTPSa2* with its native promoter (*proLcTPSa2*) and *LcNAC1* in *N. benthamiana* leaves. Our results revealed that a higher level of farnesol was observed in the leaves cotransformed with LcNAC1 and *proLcTPSa2::LcTPSa2*

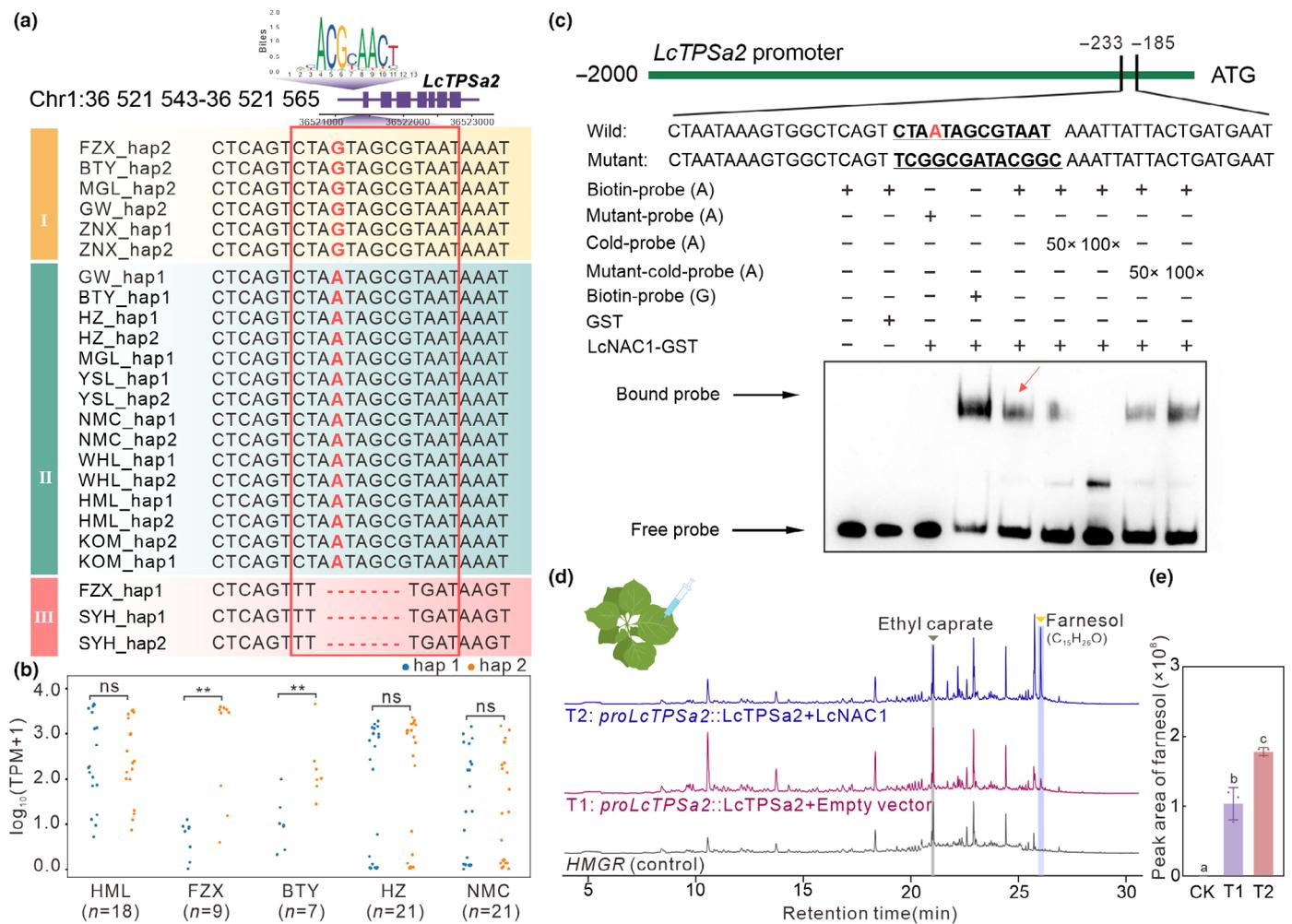


Fig. 4 Genetic variation in the promoter of *LcTPSa2* affects the binding ability of LcNAC1. (a) Sequence alignment of the LcNAC1 binding site across 12 lychee germplasms. The FZX_hap2 was used as the reference genome for the location of LcNAC1. (b) Comparison of *LcTPSa2* expression levels between haplotypes (hap1 and hap2) across five lychee cultivars, analyzed using a two-sided Mann–Whitney test (**, P value < 0.01 ; ns, P value > 0.05). (c) Electrophoretic mobility shift assay shows that a single nucleotide polymorphism in the lychee *TPSa2* promoter (*LcTPSa2*^{G type} and *LcTPSa2*^{A type}) effected the binding affinity of the LcNAC1 protein. (d) Ectopic expression of *LcNAC1* in *N. benthamiana* leaves. GC–MS analysis of farnesol from *N. benthamiana* leaves overexpressing *LcNAC1*. The *N. benthamiana* leaves transformed with pEAQ served as the control group (WT). Control, tobacco leaves were infiltrated with empty vector (pEAQ) *Agrobacterium* suspension; T1, tobacco leaves were infiltrated with a mixture of *proLcTPSa2::LcTPSa2* and empty vector *Agrobacterium* suspension at a 1 : 1 ratio; T2, tobacco leaves were infiltrated with a mixture of *proLcTPSa2::LcTPSa2* and LcNAC1 *Agrobacterium* suspension at a 1 : 1 ratio. (e) Peak area of products generated in *N. benthamiana* leaves overexpressing different combinations. Results represent the mean values \pm SE of three biological replicates. The lowercase letters (a, b, c) indicate significant difference ($P < 0.05$, one-way ANOVA with Fisher's least significant difference test).

(Fig. 4d,e; Table S12). In summary, the sequence variation in the LcNAC1 binding site of the *LcTPSa2* promoter drives the divergence of farnesol biosynthesis in lychee fruit by modulating LcNAC1 binding affinity.

Discussion

The diversity of lychee fruit is evident in its color, flavor, and shape across diverse cultivars. While exhibiting these notable differences, lychee varieties share a uniform fruit ripening process, likely conserved in other plants. Our findings reveal that the process of ripening of lychee fruit resembles a river system (Fig. 5) in which the main course of the river symbolizes the core ripening process

controlled by the master regulator LcNAC1, which is conserved across species. The branching of the main river into tributaries represents the diversification process of *TPS* genes, through genetic variation including copy number variations, PAVs, InDels, and SNPs. This variation contributes to the aromatic diversity symbolized by even finer branches downstream of the tributaries (Fig. 5). This diversification process is likely divergent across species because of the species-dependent genetic variation.

Fruit ripening is a developmental process controlled by a few key transcription factors (Shi *et al.*, 2021). Climacteric fruits continue to ripen postharvest, characterized by increased respiration and ethylene production (Satekge & Magwaza, 2022). In climacteric fruits, three types of transcriptional circuits have been

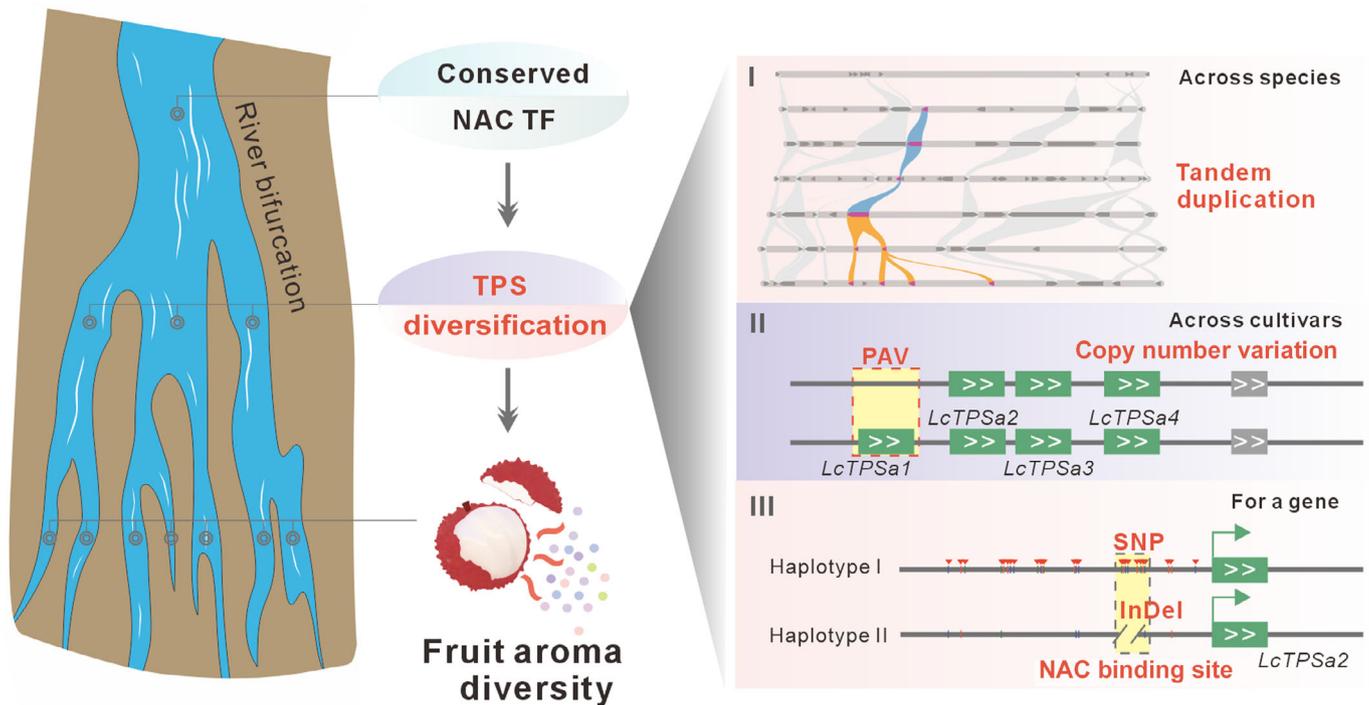


Fig. 5 A proposed model for the great aromatic diversity arising from the various natural genetic variations accumulated in the cluster of *TPS* genes. The ripening process of lychee fruit resembles a river system. The core ripening process controlled by the master regulator, *LcNAC1*, is symbolized by the main course of the river, implying a regulatory process widely conserved across various fruits. The diversification process of *TPS* genes is likely species-specific and represented by the branching of the main river into tributaries. A plethora of genetic variation, including copy number variation, presence/absence variation, insertions and deletions, and single nucleotide polymorphisms, contributes to the *TPS* diversification, ultimately leading to the remarkable aromatic diversity symbolized by these even finer branches downstream of the tributaries.

identified in the regulation of fruit ripening: the tomato MADS-type circuit involving RIN (a NAC protein), the peach NAC-type circuit, and the banana dual-loop circuit that incorporates both MADS and NAC (Lü *et al.*, 2018). Thus, the MADS and NAC transcription factors are the two major players governing the fruit ripening process. Our study revealed that *LcNAC1*, the ortholog of tomato *NOR-like1*, has both expression and evolutionary similarity to *NAC* genes known for regulating fruit ripening in crops such as tomato, grape, peach, and apple (Giovannoni, 2004; Gao *et al.*, 2018; Niu *et al.*, 2022; Cao *et al.*, 2023; Cao *et al.*, 2024). Genes potentially regulated by *LcNAC1* were notably enriched in various pathways associated with fruit ripening, particularly the one involving terpene synthase genes that contribute to the delightful aroma of lychee fruit. These findings underline what we hypothesize to be the crucial role of the *LcNAC1* transcription factor in the regulation of fruit ripening in lychee.

Our study also found that substantial natural genetic variation in *TPS* genes, the downstream target genes of *LcNAC1*, may affect their expression levels and contribute to the aromatic diversity in lychee fruit. These *TPS* genes, responsible for farnesol biosynthesis, form a tandemly duplicated cluster in the genome of lychee. Collinear synteny analysis across seven Sapindaceae species revealed that, with the exception of maples, which lack these *TPS* genes, all other species possess between one and four copies (Fig. S20; Table S6). Lychee is the only species having all the four *TPS* genes,

indicating that this tandem duplication event occurred specifically within the lychee genome. Thus, this expansion of *TPS* genes, which are highly active during fruit ripening, likely plays a significant role in defining the unique flavor of lychee fruit.

In addition, other genetic variants such as InDels, PAVs, and SNPs were also found for these clustered *TPS* genes. These types of variants are common in plant genomes; they can modify gene dosage and expression levels, leading to phenotypic changes that affect plant height, flowering time, and seed dormancy (Che *et al.*, 2022; Chen *et al.*, 2022; Xu *et al.*, 2022). For instance, a 21-bp InDel in the promoter of sugar transporter protein 1 contributes to soluble solid content in tomatoes (Wang *et al.*, 2023). Apart from these, we also discovered a truncated version of a *TPS* gene, *LcTPSa1*, resulting from a single nucleotide mutation that causes a premature translational termination, yet the truncated protein remains as functional as the full-length version. This opens the possibility of engineering typical *TPS* proteins into shorter versions, which could be valuable in synthetic biology in which gene size is currently a critical factor for the assembly of large gene sets.

Lychee is one of the most attractive tropical or subtropical fruits on the international market, prized for its exceptional nutritional profile, exotic flavor, and appealing fruit color. Compared with other fruits in the Sapindaceae family, lychee fruit exhibits greater diversity in color, flavor, and metabolites. These distinct characteristics likely result from its unique domestication,

cultivation, and breeding processes, during which genetic variation contributing to its unique fruit traits has accumulated and been selected. The genetic variation characterized in this study not only enhances our understanding of the formation of lychee fruit aroma but also provides valuable resources for DNA-marker-based breeding programs or biotechnological engineering aimed at developing lychee cultivars with even more desirable fruit traits. Furthermore, this diversification of fruit aroma in lychee germplasm offers a compelling example of how species- or variety-specific traits evolve; that is, the phenotypic diversity is primarily derived from natural genetic variation in downstream structural genes regulated by a widely conserved regulatory circuit.

Acknowledgements

This work was supported by the Key Area Research and Development Program of Guangdong Province (2022B0202070003). This work was also supported by the National Science Foundation of China (nos. 32072547 and 32372665). The authors are grateful to Dr Zhenhua Liu's laboratory (Shanghai Jiao Tong University) for sharing with us the modified pEAQ-HT-DEST-1 vector, and the authors also thank Mrs Yin Zheng for her assistance with GC-MS analysis.

Competing Interests

None declared.

Author Contributions

HH, YH and RX designed the research. HH, HL, ZZ, YX, YM and YZ performed the research. HH and HL analyzed the data. HH, HL, BCM, YH and RX wrote the paper. HH and HL contributed equally to this work.

ORCID

Yanwei Hao  <https://orcid.org/0000-0002-6465-4885>
Blake C. Meyers  <https://orcid.org/0000-0003-3436-6097>
Rui Xia  <https://orcid.org/0000-0003-2409-1181>
Zaohai Zeng  <https://orcid.org/0009-0006-0789-0175>

Data availability

Sequencing raw reads in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under the accession BioProject no. PRJNA1163705.

References

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J *et al.* 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**: 493–500.
- Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic Acids Research* **43**: W39–W49.
- Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M, Bitz O, Schwab W. 2014. Activity-based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from grape. *Plant Physiology* **166**: 23–39.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. EGGNOG-MAPPER v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* **38**: 5825–5829.
- Cao K, Wang B, Fang W, Zhu G, Chen C, Wang X, Li Y, Wu J, Tang T, Fei Z *et al.* 2022. Combined nature and human selections reshaped peach fruit metabolome. *Genome Biology* **23**: 146.
- Cao X, Li X, Su Y, Zhang C, Wei C, Chen K, Grierson D, Zhang B. 2023. Transcription factor PpNAC1 and DNA demethylase PpDML1 synergistically regulate peach fruit ripening. *Plant Physiology* **194**: 2049–2068.
- Cao X, Su Y, Zhao T, Zhang Y, Cheng B, Xie K, Yu M, Allan A, Klee H, Chen K *et al.* 2024. Multi-omics analysis unravels chemical roadmap and genetic basis for peach fruit aroma improvement. *Cell Reports* **43**: 114623.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. TRIMAL: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
- Che G, Pan Y, Liu X, Li M, Zhao J, Yan S, He Y, Wang Z, Cheng Z, Song W *et al.* 2022. Natural variation in CRABS CLAW contributes to fruit length divergence in cucumber. *Plant Cell* **35**: 738–755.
- Chen C, Li J, Feng J, Liu B, Feng L, Yu X, Li G, Zhai J, Meyers BC, Xia R. 2021. sRNAanno – a database repository of uniformly annotated small RNAs in plants. *Horticulture Research* **8**: 45.
- Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y *et al.* 2023. TTools-II: a “one for all, all for one” bioinformatics platform for biological big-data mining. *Molecular Plant* **16**: 1733–1742.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. FASTP: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.
- Chen Z, Ke W, He F, Chai L, Cheng X, Xu H, Wang X, Du D, Zhao Y, Chen X *et al.* 2022. A single nucleotide deletion in the third exon of FT-D1 increases the spikelet number and delays heading date in wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal* **20**: 920–933.
- Colantonio V, Ferrão LFV, Tieman DM, Bliznyuk N, Sims C, Klee HJ, Munoz P, Resende MFR. 2022. Metabolomic selection for enhanced fruit flavor. *Proceedings of the National Academy of Sciences, USA* **119**: e2115865119.
- Deng H, Chen Y, Liu Z, Liu Z, Shu P, Wang R, Hao Y, Su D, Pirrello J, Liu Y *et al.* 2022. SIERF.F12 modulates the transition to ripening in tomato fruit by recruiting the co-repressor TOPLESS and histone deacetylases to repress key ripening genes. *Plant Cell* **34**: 1250–1272.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Emms DM, Kelly S. 2019. ORTHOFINDER: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.
- Fan S, Wang D, Xie H, Wang H, Qin Y, Hu G, Zhao J. 2021. Sugar transport, metabolism and signaling in fruit development of *Litchi chinensis* Sonn: a review. *International Journal of Molecular Sciences* **22**: 11231.
- Gao Y, Wei W, Fan Z, Zhao X, Zhang Y, Jing Y, Zhu B, Zhu H, Shan W, Chen J *et al.* 2020. Re-evaluation of the nor mutation and the role of the NAC-NOR transcription factor in tomato fruit ripening. *Journal of Experimental Botany* **71**: 3560–3574.
- Gao Y, Wei W, Zhao X, Tan X, Fan Z, Zhang Y, Jing Y, Meng L, Zhu B, Zhu H *et al.* 2018. A NAC transcription factor, NOR-like1, is a new positive regulator of tomato fruit ripening. *Horticulture Research* **5**: 75.
- Giovannoni J, Nguyen C, Ampofo B, Zhong S, Fei Z. 2017. The epigenome and transcriptional dynamics of fruit ripening. *Annual Review of Plant Biology* **68**: 61–84.
- Giovannoni JJ. 2004. Genetic regulation of fruit development and ripening. *Plant Cell* **16**: S170–S180.
- Hao J, Xu Y, Li C, Dou H, Gu H, Hu W. 2007. Analysis of aromatic compositions in different breeds of litchi by SPME/GC-MS methods. *Food Science* **28**: 12.

- Hu G, Feng J, Xiang X, Wang J, Salojärvi J, Liu C, Wu Z, Zhang J, Liang X, Jiang Z *et al.* 2022. Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. *Nature Genetics* 54: 73–83.
- Jia H, Xu Y, Deng Y, Xie Y, Gao Z, Lang Z, Niu Q. 2024. Key transcription factors regulate fruit ripening and metabolite accumulation in tomato. *Plant Physiology* 195: 2256–2273.
- Jia Q, Brown R, Köllner TG, Fu J, Chen X, Wong GK-S, Gershenzon J, Peters RJ, Chen F. 2022. Origin and early evolution of the plant terpene synthase family. *Proceedings of the National Academy of Sciences, USA* 119: e2100361119.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software V.7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Koo HJ, Vickery CR, Xu Y, Louie GV, O'Maille PE, Bowman M, Nartey CM, Burkart MD, Noel JP. 2016. Biosynthetic potential of sesquiterpene synthases: product profiles of Egyptian Henbane prenaspirodiene synthase and related mutants. *The Journal of Antibiotics* 69: 524–533.
- Kumar S, Suleski M, Craig JM, Kasprowitz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TIME TREE 5: an expanded resource for species divergence times. *Molecular Biology and Evolution* 39: msac174.
- Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559.
- Li C, Hao J, Zhong H, Xu Y. 2010. Free and glycosidically bound volatile flavor compounds in fruit of *Litchi chinensis* Huaizhi. *Food Science* 31: 24.
- Li J, Chen C, Zeng Z, Wu F, Feng J, Liu B, Mai Y, Chu X, Wei W, Li X *et al.* 2024. SapBase: a central portal for functional and comparative genomics of Sapindaceae species. *Journal of Integrative Plant Biology* 66: 1561–1570.
- Li S, Chen K, Grierson D. 2019. A critical evaluation of the role of ethylene and MADS transcription factors in the network controlling fleshy fruit ripening. *New Phytologist* 221: 1724–1741.
- Li X, Martín-Pizarro C, Zhou L, Hou B, Wang Y, Shen Y, Li B, Posé D, Qin G. 2023. Deciphering the regulatory network of the NAC transcription factor FvRIF, a key regulator of strawberry (*Fragaria vesca*) fruit ripening. *Plant Cell* 35: 4020–4045.
- Liu Z, Zhao M, Li J. 2022. Aroma volatiles in litchi fruit: a mini-review. *Horticulturae* 8: 1166.
- Lü P, Yu S, Zhu N, Chen Y-R, Zhou B, Pan Y, Tzeng D, Fabi JP, Argyris J, Garcia-Mas J *et al.* 2018. Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. *Nature Plants* 4: 784–791.
- Maupilé L, Chaib J, Boualem A, Bendahmane A. 2024. Parthenocarpy, a pollination-independent fruit set mechanism to ensure yield stability. *Trends in Plant Science* 29: 1254–1265.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.
- Mishra S, Srivastava AK, Khan AW, Tran L-SP, Nguyen HT. 2024. The era of panomics-driven gene discovery in plants. *Trends in Plant Science* 29: 995–1005.
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ *et al.* 2020. Pfam: the protein families database in 2021. *Nucleic Acids Research* 49: D412–D419.
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009. AUTODOCK4 and AUTODOCKTOOLS4: automated docking with selective receptor flexibility. *Journal of Computational Chemistry* 30: 2785–2791.
- Nieuwenhuizen NJ, Chen X, Wang MY, Match AJ, Perez RL, Allan AC, Green SA, Atkinson RG. 2015. Natural variation in monoterpene synthesis in kiwifruit: transcriptional regulation of terpene synthases by NAC and ETHYLENE-INSENSITIVE3-like transcription factors. *Plant Physiology* 167: 1243–1258.
- Niu Z, Zhao Y, Chen Z, Xuan L, Niu S, Chu F, Yang L. 2022. Screening, cloning and expression analysis of NAC transcription factors related to grape fruit ripening. *Journal of Fruit Science* 39: 1137–1146.
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. STRINGTIE enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33: 290–295.
- Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29: e45.
- Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. *Nucleic Acids Research* 46: W200–W204.
- Raulusevičius I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, Lemma RB, Lucas J, Chêneby J, Baranasic D *et al.* 2023. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* 52: D174–D182.
- Satekge TK, Magwaza LS. 2022. Postharvest application of 1-methylcyclopropene (1-MCP) on climacteric fruits: factors affecting efficacy. *International Journal of Fruit Science* 22: 595–607.
- Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W *et al.* 2014. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346: 1084–1088.
- Shi Y, Vrebalo J, Zheng H, Xu Y, Yin X, Liu W, Liu Z, Sorensen I, Su G, Ma Q *et al.* 2021. A tomato LATERAL ORGAN BOUNDARIES transcription factor, *SLOB1*, predominantly regulates cell wall and softening components of ripening. *Proceedings of the National Academy of Sciences, USA* 118: e2102486118.
- Tang H, Krishnakumar V, Zeng X, Xu Z, Taranto A, Lomas JS, Zhang Y, Huang Y, Wang Y, Yim WC *et al.* 2024. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta* 3: e211.
- Teh BT, Lim K, Yong CH, Ng CCY, Rao SR, Rajasegaran V, Lim WK, Ong CK, Chan K, Cheng VKY *et al.* 2017. The draft genome of tropical fruit durian (*Durio zibethinus*). *Nature Genetics* 49: 1633–1641.
- Wang J, Sun Q, Ma C, Wei M, Wang C, Zhao Y, Wang W, Hu D. 2024. MdWRKY31-MdNAC7 regulatory network: orchestrating fruit softening by modulating cell wall-modifying enzyme MdXTH2 in response to ethylene signalling. *Plant Biotechnology Journal* 22: 3244–3261.
- Wang Y, Shi C, Ge P, Li F, Zhu L, Wang Y, Tao J, Zhang X, Dong H, Gai W *et al.* 2023. A 21-bp InDel in the promoter of STP1 selected during tomato improvement accounts for soluble solid content in fruits. *Horticulture Research* 10: uhad009.
- Wu B, Cao X, Liu H, Zhu C, Klee H, Zhang B, Chen K. 2018. UDP-glucosyltransferase PpUGT85A2 controls volatile glycosylation in peach. *Journal of Experimental Botany* 70: 925–936.
- Wu Y, Pan Q, Qu W, Duan C. 2009. Comparison of volatile profiles of nine litchi (*Litchi chinensis* Sonn.) cultivars from Southern China. *Journal of Agricultural and Food Chemistry* 57: 9676–9681.
- Xiang Y, Zhang T, Zhao Y, Dong H, Chen H, Hu Y, Huang C, Xiang J, Ma H. 2024. Angiosperm-wide analysis of fruit and ovary evolution aided by a new nuclear phylogeny supports association of the same ovary type with both dry and fleshy fruits. *Journal of Integrative Plant Biology* 66: 228–251.
- Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. 2023. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Research* 51: W587–W592.
- Xu F, Tang J, Wang S, Cheng X, Wang H, Ou S, Gao S, Li B, Qian Y, Gao C *et al.* 2022. Antagonistic control of seed dormancy in rice by two bHLH transcription factors. *Nature Genetics* 54: 1972–1982.
- Yan X-M, Zhou S-S, Liu H, Zhao S-W, Tian X-C, Shi T-L, Bao Y-T, Li Z-C, Jia K-H, Nie S *et al.* 2023. Unraveling the evolutionary dynamics of the TPS gene family in land plants. *Frontiers in Plant Science* 14: 1273648.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.
- Yauk Y, Ged C, Wang MY, Match AJ, Tessarotto L, Cooney JM, Chervin C, Atkinson RG. 2014. Manipulation of flavour and aroma compound sequestration and release using a glycosyltransferase with specificity for terpene alcohols. *The Plant Journal* 80: 317–330.
- Zhang Y, Zeng Z, Hu H, Zhao M, Chen C, Ma X, Li G, Li J, Liu Y, Hao Y *et al.* 2024. MicroRNA482/2118 is lineage-specifically involved in gibberellin signalling via the regulation of GID1 expression by targeting noncoding PHAS genes and subsequently instigated phasiRNAs. *Plant Biotechnology Journal* 22: 819–832.
- Zhao L, Wang K, Wang K, Zhu J, Hu Z. 2020. Nutrient components, health benefits, and safety of litchi (*Litchi chinensis* Sonn.): a review. *Comprehensive Reviews in Food Science and Food Safety* 19: 2139–2163.
- Zhu Q, Tan Q, Gao Q, Zheng S, Chen W, Galaud J, Li X, Zhu X. 2024. Calmodulin-like protein CML15 interacts with PP2C46/65 to regulate papaya fruit ripening via integrating calcium, ABA and ethylene signals. *Plant Biotechnology Journal* 22: 1703–1723.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Relative expression of genes in 15 co-expression modules during lychee cv ‘Huaizi’ fruit development and ripening.

Fig. S2 Relative expression of genes in 14 co-expression modules during lychee cv ‘Nuomici’ fruit development and ripening.

Fig. S3 Phylogenetic tree of the LcNAC1 family orthologous genes across 21 species.

Fig. S4 Phylogenetic relationships of lychee TPS genes.

Fig. S5 Expression pattern of *TPS* genes during lychee cv ‘Hemaoli’ fruit development.

Fig. S6 Expression pattern of *TPS* genes during lychee cv ‘Huaizi’ fruit development.

Fig. S7 Expression pattern of *TPS* genes during lychee cv ‘Nuomici’ fruit development.

Fig. S8 Alignment of promoter sequences and schematic diagram for the LcNAC1 binding sites of *LcTPSa1–LcTPSa4* genes.

Fig. S9 Sodium dodecyl sulfate–polyacrylamide gel electrophoresis analysis of GST-LcNAC1 protein.

Fig. S10 Alignment coverage map of PacBio HiFi sequencing reads from 12 lychee cultivars at the *LcTPSa1–LcTPSa4* locus, utilizing ‘FZX_hap2’ as the reference genome.

Fig. S11 Gene synteny analysis among 12 lychee genomes.

Fig. S12 Relative expression percentage of *LcTPSa1–LcTPSa4* genes out of the total TPS gene expression during ripening in ‘Huaizi’ and ‘Nuomici’ lychee cultivars.

Fig. S13 Nucleotide sequence alignment of the tandemly duplicated *LcTPSa1–LcTPSa4* genes.

Fig. S14 Comparison of the deduced amino acid sequences of *LcTPSa1–LcTPSa4* proteins with GaDCS from *Gossypium arborum* (28) and NtEAS3 from *Nicotiana tabacum* (23).

Fig. S15 Sodium dodecyl sulfate–polyacrylamide gel electrophoresis analysis of LcTPSa1-His and LcTPSa2-His proteins.

Fig. S16 Metabolic profiling of ripe lychee fruits.

Fig. S17 Farnesol content during lychee cv ‘Huaizi’ fruits development.

Fig. S18 IGV screenshot illustrating the alignment of ‘Feizixiao’ (FZX) PacBio HiFi data and the fruit transcriptome data, using ‘FZX_hap2’ as the reference genome.

Fig. S19 IGV screenshot illustrating the alignment of ‘Baitangying’ (BTY) PacBio HiFi data and the fruit transcriptome data, using ‘BTY_hap2’ as the reference genome.

Fig. S20 Conserved LcTPSs gene blocks in the genomes of seven Sapindaceae species.

Table S1 RNA-seq sample information in this study.

Table S2 Primers and genome information used in this study.

Table S3 Gene information of lychee (*Litchi chinensis* Sonn.) genome.

Table S4 Co-expression modules and Fragments Per Kilobase of exon model per Million (FPKM) values of genes in lychee fruit as shown in Figs 1(b), S4, and S5.

Table S5 Fragments Per Kilobase of exon model per Million (FPKM) value of the 35 highly expressed transcription factor genes in Module 2 across cv ‘Hemaoli’ fruit aril development.

Table S6 LcNAC1 and LcTPSa synteny genes as shown in Figs 1(d), S20.

Table S7 Results of orthogroup genes in 21 plant species.

Table S8 Predicted results of NAC target genes.

Table S9 KEGG pathway enrichment results for the genes containing NAC binding sites in co-expression Module 2 as shown in Fig. 1(e).

Table S10 Classification results of TPS genes.

Table S11 Enzymatic assay products of LcTPSa1 and LcTPSa2 upon farnesyl pyrophosphate as substrate.

Table S12 Transient expression products of LcTPSa1, LcTPSa2, and LcNAC1 in tobacco leaves.

Table S13 Metabolic profiling of six lychee ripe fruits.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.