Molecular evolution of terpene synthase underlying the diversification of isoprene emission in Fagaceae

Yuka Ikezaki[1], Shuichi N. Kudo[1], Taichi Nakata[1], Sora Koita[2], Ryosuke Munakata[2], Kazufumi Yazaki[2], Takeshi Torimaru[3], Nobuhiro Tomaru[4], Sachiko Isobe[5,6], Hideki Hirakawa[5,7], Junko Kusumi[8]*, Akiko Satake[1]*

1. Department of Biology, Faculty of Science, Kyushu University, Fukuoka 819-0395, Japan.
2. Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji, 611-0011 Japan.
3. Graduate School of Bioresources, Mie University, Tsu, 514-8507 Japan.
4. Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8603 Japan.
5. Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan.
6. Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, 113-8657, Japan.
7. Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Fukuoka 819-0395, Japan.
8. Department of Environmental Changes, Faculty of Social and Cultural Studies, Kyushu University, Fukuoka 819-0395, Japan.

*Corresponding author:
Akiko Satake
Email: satake.akiko.269@m.kyushu-u.ac.jp
Kusumi Junko
Email: jkusumi@scs.kyushu-u.ac.jp

Author Contributions: Y.I. and A.S. designed research; Y.I., S.N.K., N.T., S.K., R.M., K.Y., S.I., H.H., J.K. and A.S. performed research; Y.I., S.N.K., N.T., S.K., R.M., K.Y., S.I., H.H., J.K. and A.S. analyzed data; T.T. and N.T. provided genome data; Y.I. and A.S. wrote the paper with inputs from all the authors.

Competing Interest Statement: The authors declare no conflict of interest.

Classification: Biological Sciences, Evolution

Keywords: Biogenic volatile compounds, isoprene synthase, ancestral sequence reconstruction, positive selection, convergent evolution

Abstract

Plants emit a wide range of volatile organic compounds, among which isoprene is the most abundant and atmospherically influential. Although oak species are major contributors to isoprene emission, there is considerable variation in isoprene

emission capacity within the Fagaceae family. To unravel the evolutionary origins of isoprene emission, we investigated the molecular evolution of terpene synthase (TPS) genes across eight species from four genera within the Fagaceae. We identified a Fagaceae-specific TPS-b subclade in which potential isoprene synthase (IspS) activity evolved independently in two gene lineages within subgenus *Quercus*. Seasonal gene expression patterns showed that one lineage exhibited high expression particularly in summer, whereas the other displayed consistently low expression across all seasons. Ancestral sequence reconstruction revealed that the acquisition of a diagnostic amino acid residue for IspS function (a substitution from Thr338 to Phe338) arose convergently in these lineages and was subject to positive selection, suggesting adaptive evolution. Ancestral-enzyme assays targeting the gene lineage with high gene expression revealed that the early protein primarily produced monoterpenes—particularly *trans*-β-ocimene—from geranyl diphosphate (GPP), whereas their descendants shifted substrate preference to dimethylallyl diphosphate (DMAPP), evolving into dedicated isoprene synthases. Our results indicate that IspS activity was not ancestral in Fagaceae, but evolved approximately 56 million years ago within the subgenus *Quercus*, and has been retained ever since. These findings emphasize the roles of enzyme structural innovation and regulatory shifts in the diversification of volatile terpenoid biosynthesis.

Significance Statement

Isoprene, the most abundantly emitted biogenic volatile organic compound, influences atmospheric chemistry and protects plants from thermal stress. Yet, its evolutionary origins remain unclear. By analyzing TPS genes across eight Fagaceae species, we demonstrated that high isoprene synthase activity was not ancestral in Fagaceae, but evolved approximately 56 million years ago within the subgenus *Quercus*. Ancestral sequence reconstruction revealed that IspS activity arose through a convergent amino acid substitution under positive selection. Enzyme assays further showed that this functional shift was accompanied by substrate preference changes. Our study provides the genomic evidence of convergent evolution of isoprene biosynthesis in trees and highlights the molecular basis underpinning ecological adaptation through specialized metabolism.

Main Text

Introduction

Plants release a wide spectrum of biogenic volatile organic compounds (BVOCs), which include molecules with diverse structural and physicochemical characteristics (1-3). These compounds, emitted from various plant tissues such as leaves, flowers, fruits, and roots, enter the atmosphere in significant quantities (4). BVOCs serve as a dynamic interface between terrestrial ecosystems and the atmosphere, playing an important role in climate and atmospheric chemistry (5). The annual global emission of BVOCs is estimated to be around 1000 Tg C, with terpenoids—including isoprene, monoterpenes, and sesquiterpenes (6, 7). Among

these, isoprene alone accounts for roughly half of the total BVOC, underscoring its ecological and atmospheric significance (6).

In the plant kingdom, isoprene emission is distributed heterogeneously across plant taxa (8-10). It is particularly common in fast-growing plants, especially hardwood trees such as oaks (*Quercus* spp.), *Populus* spp. (11), and *Eucalyptus* spp. (10). Isoprene emission has also been reported in mosses (12), ferns (13), and legumes (14), indicating its occurrence across diverse plant groups. Isoprene is synthesized by isoprene synthase (IspS), a class I terpenoid synthase (TPS) enzyme that converts dimethylallyl diphosphate (DMAPP) into isoprene at a key terminal branching point of the methylerythritol phosphate (MEP) pathway (15, 16). The TPS family likely originated in ancestral land plants following their divergence from green algae (17). In angiosperms, IspS belongs to the TPS-b subfamily (15, 18). In contrast, recent research has shown that IspS in mosses belongs to the TPS-c subfamily (19), suggesting that IspS has multiple evolutionary origins within the plant kingdom. In angiosperms, IspS appears to have evolved multiple times, likely from closely related terpene synthases. Among these, *trans*-β-ocimene synthases are considered a key precursor, reflecting the functional diversification of terpene synthases in plants (15, 18).

An intriguing and well-studied example of the uneven distribution of isoprene emission within a plant clade can be found in the Fagaceae family (8, 9). In this family, isoprene emission is particularly prominent in the genus *Quercus* (oaks), whereas other genera, such as *Fagus*, *Castanopsis*, and *Lithocarpus* exhibit little to no isoprene emission (9, 20, 21). Intriguingly, within the genus *Quercus*, isoprene emission capacity exhibits significant variation, with a mix of strong isoprene emitters, such as *Q. rubra* and *Q. robur*, and non-emitters, such as *Q. ilex* and *Q. suber* (22). Based on analyses associating isoprene emission capacity with molecular phylogeny within the Fagaceae, Loreto (20) and Monson et al. (9) concluded that isoprene emission was likely an ancestral trait in the genus *Quercus* and was subsequently lost and not restored in certain species within the Quercus sections *Cerris* and *Ilex*. These findings lay a critical foundation for reevaluating the evolutionary history of isoprene emission in the Fagaceae. The growing availability of high-quality genomic and transcriptomic data, even for non-model Fagaceae species (23-25), presents a valuable opportunity to reconstruct the evolutionary trajectory of isoprene emission by examining genome evolution and the expression patterns of *IspS*.

In the present study, we investigated the evolutionary trajectory of isoprene emission in eight species representing four genera within the Fagaceae family. Using comparative genomics, phylogenetics, and ancestral sequence reconstruction, we uncovered a convergent evolutionary origin of *IspS*, shedding light on a novel evolutionary pathway to isoprene emission in angiosperms.

Results

Identification of TPS genes in Fagaceae. To identify TPS genes in Fagaceae, we first obtained genome sequences of eight Fagaceae trees from four genera using public databases (Table S1). These included five species from the genus *Quercus* (*Q. robur*, *Q. lobata*, *Q. rubra*, *Q. acutissima*, and *Q. glauca*), and one

136 species each from the genera *Lithocarpus* (*L. edulis*), *Castanopsis* (*C. hystrix*), and
137 *Fagus* (*F. crenata*) (Table S1). Isoprene emission factors vary significantly among
138 these eight species (Table S2), highlighting their suitability for studying the
139 evolution of isoprene emission. A comparative genomic analysis of these species
140 revealed a highly conserved syntenic structure across genera, including *Quercus*,
141 *Lithocarpus*, and *Castanopsis* (Fig. 1*A*). However, the genomic structure of the
142 genus *Fagus* differed significantly from that of the other genera (Fig. 1*A*),
143 suggesting extensive genome rearrangements following its divergence from the
144 rest of the Fagaceae.
145 To identify TPS genes in our target Fagaceae species, we conducted
146 HMMER searches using the annotated protein sequences from the genomes of
147 eight Fagaceae species against the metal-binding domain (PF03936) and the N-
148 terminal domain (PF01397). Each TPS candidate gene was further examined and
149 filtered based on exon-intron structure, including the number and locations of
150 exons (Fig. S1), by comparing with previously reported TPSs (26), as well as total
151 sequence length (Fig. S2). The number of identified TPS genes varied across
152 species, ranging from a minimum of 24 in *C. hystrix* to a maximum of 67 in *Q.*
153 *glauca* (Fig. S3).
154 The rooted TPS phylogenetic tree constructed from our TPS dataset along
155 with sequences from *Solanum lycopersicum* (tomato) and *Arabidopsis thaliana*,
156 revealed a clear distinction among the TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g
157 subfamilies (Fig. 1*B*). TPS-a members in Fagaceae formed a distinct clade,
158 separate from those of tomato and *A. thaliana*, but showing a closer evolutionary
159 relationship to tomato TPS-a genes (Fig. 1*B*). The TPS-c, TPS-g, and TPS-e/f
160 subfamilies were present in all Fagaceae species, except for the TPS-e/f
161 subfamily, which was absent in *Q. acutissima* (Fig. 1*B*).
162 The TPS-b subfamily represents the largest gene group within Fagaceae
163 species (Fig. 1*B*). The three TPS-b genes in tomato form a distinct clade from other
164 tomato and Arabidopsis TPS-b genes (red asterisk in Fig. 1*B*). These three tomato
165 genes are known to synthesize acyclic monoterpenes, with one of them, TPS47,
166 encoding a cytosolic isoprene synthase (27).
167 Many genes in the TPS-a and TPS-b subfamilies are arranged in tandem
168 gene arrays on chromosomes 8 and 9, respectively, suggesting their origin from
169 tandem gene duplication events (Fig. S4*A*). Overall, the genomic locations of these
170 genes were largely conserved across *Quercus*, *Lithocarpus*, and *Castanopsis*.
171 However, in *Fagus*, significant genome rearrangements were observed, indicating
172 a distinct evolutionary trajectory (Fig. S4*B*).
173
174 **Isoprene score of TPS-b genes in Fagaceae.** To investigate the evolutionary
175 history of *IspS* in Fagaceae, we focused on the TPS-b subfamily that includes
176 known *IspS* from other plant species. We designated the clade containing the
177 tomato *IspS* as TPS-b-target1 (TPS-b-t1) and the clade containing the *Q. serrata*
178 *IspS* as TPS-b-target1 (TPS-b-t3) (Fig. 2*A*). Between the TPS-b-t1 and TPS-b-t3,
179 we identified a distinct clade, which we designate as TPS-b-target2 (TPS-b-t2)
180 (Fig. 2*A*).Notably, both TPS-b-t2 and TPS-b-t3 consist exclusively of genes from

Fagaceae, suggesting that these clades are specific to this family and potentially involved in the evolution of *IspS* in Fagaceae.

Sequences of genes included in these three clades were aligned and used to calculate an isoprene score (Fig. 2*B*). The isoprene score serves as a useful screening tool to identify enzymes with exclusive IspS activity based on sequence information and the structural analysis of IspS (15). This score is calculated using the four conserved amino acid residues (known as the diagnostic tetrad) that are characteristic of IspS. Two of the residues (F338 and F485 in Fig. 2*B*) were suggested to restrict the size of the substrate-binding pocket and play a key role in determining enzyme specificity (15). While F338 and F485 of the diagnostic tetrad residues are considered highly specific to IspS, the other positions are considered redundant (18). In IspS in some angiosperm families, valine (V446) is found instead of serine (18). Therefore, we treated both serine and valine as conserved residues at this position and denoted it as S/V446. The isoprene score ranges from 0 to 4, with a score of 4 representing full conservation of the diagnostic tetrad and thus a high likelihood of exclusive IspS activity.

The multiple sequence alignment revealed that the isoprene score for sequences in the TPS-b-t1 clade was consistently one, whereas the IspS sequence in tomato had a score of 3 (Fig. 2*B*). In the TPS-b-t2 and TPS-b-t3 clades, sequences from the species in the subgenus *Quercus* (*Q. robur*, *Q. lobata*, *Q. rubra*, and *Q. serrata*) were found to have an isoprene score of 4. Specifically, *Q. robur* contained three, *Q. lobata* had two, and *Q. rubra* had one sequence with an isoprene score of 4. In contrast, the maximum isoprene score for sequences from *Lithocarpus*, *Castanopsi*s, and *Fagus* was only three. Many sequences lacked F338, an amino acid considered crucial for isoprene synthase activity (15). These patterns in isoprene score align with isoprene emission capacities, where *Q. robur*, *Q. lobata*, *Q. rubra*, and *Q. serrata* are known to exhibit high isoprene emission, while other species show lower emission levels (Table S2)

Within the twin clades TPS-b-t2 and TPS-b-t3, *Fagus* sequences were found only in TPS-b-t2, whereas sequences from other species were present in both clades. Most genes within TPS-b-t2 and TPS-b-t3 were located on chromosome 8, with the exception of *F. crenata* (Fig. S4). Interestingly, two genes in *Q. robur* and one in *Q. lobata*, all with an isoprene score of 4, were located on chromosome 3 (Fig. S4), suggesting chromosomal translocation events.

Seasonal expression patterns of TPS genes. To investigate the seasonal expression patterns of TPS genes, we conducted RNA-seq on leaf and bud tissues collected monthly over a two-year period from *Q. acutissima* growing under natural conditions in Fukuoka, Japan. Additionally, we incorporated previously published field transcriptome data from *Q. glauca* and *L. edulis* (25), which were sampled at the same study site as *Q. acutissima*. On average, 36% and 41% of TPS genes were expressed in leaves (Fig. 3*A*) and buds (Fig. S6*A*), respectively. Within the TPS-b clade, 42% of genes were expressed in leaves, including 6 out of 14 in *Q. acutissima*, 10 out of 28 in *Q. glauca*, and 2 out of 24 in *L. edulis*. Seasonal expression profiles in leaves revealed that most expressed genes exhibited peak expression in summer, although a small fraction of genes showed peaked

expression during autumn–winter (e.g. Qa12 and Qg1 in the TPS-a clade in Fig. 3*A*). Notably, expression patterns differed markedly among the three target TPS subclades i.e., TPS-b-t1, t2, and t3. In leaves, only genes in the TPS-b-t3 clade showed clear and consistent summer expression peaks (Fig. 3*D*), whereas expression levels in the TPS-b-t1 and TPS-b-t2 clades were low in all species (Fig. 3*B, C*). These results suggest that TPS-b-t3 genes are primarily responsible for terpenoid biosynthesis. Furthermore, the expression pattern of *Q, serrata IspS* (28) within the TPS-b-t3 clade also showed strong expression (Fig. 3*B*-*D*), providing additional support for the functional role of this clade in isoprene biosynthesis. In contrast, in bud tissues, expression levels of genes within the TPS-b-t3 clade were consistently low, and genes in the other two clades, TPS-b-t1 and t2 were relatively high (Fig. S6*B*-*D*), implying a tissue-specific regulation of TPS gene expression.

Convergent evolution and positive selection in *IspS* candidates of Fagaceae. To determine whether the genes with an isoprene score of 4 were gained or lost during evolution, we employed the ancestral sequence reconstruction (ASR) method. This approach relies on statistical principles to infer the sequences of putative ancestral genes based on the amino acid or nucleotide sequences of extant species. For this analysis, we used nucleotide sequences from 17 Fagales species belonging to the TPS-b clade (Table S3) and reconstructed putative ancestral sequences using codeml implemented in PAML (https://github.com/abacus-gene/paml).

The estimated ancestral sequences of the last common node of the TPS-b-t1, TPS-b-t2, and TPS-b-t3 clades (node N2; Fig. 4*A*) had an isoprene score of 2, with the conserved residuals S/V446 and F485 (Fig. 4*B*). Subsequently, N505 was gained in the common ancestor of the TPS-b-t2 and TPS-b-t3 clades (node N4; Fig. 4*A*, *B*), prior to their duplication. Following this duplication, F338 was acquired independently in each of the TPS-b-t2 and TPS-b-t3 clades (nodes N6 and 8; Fig. 4*A B*), specifically in the lineage leading to the three species within subgenus *Quercus* (*Q. robur*, *Q lobata*, and *Q. rubra*).

Notably, the gain of F338 occurred independently in both the TPS-b-t2 and the TPS-b-t3 clades, despite requiring two nucleotide substitutions for the transition from ACT (T) to TTT (F). We also employed GRaSP (https://github.com/charles-abreu/GRaSP) to infer the ancestral sequences and reached the same conclusion that the gain of F338 occurred independently in both the TPS-b-t2 and TPS-b-t3 clades (Fig. S7). Although the inferred ancestral sequences for the distant ancestor at nodes N1–N3 differed slightly between PAML and GRaSP, the evolutionary pattern across the remaining nodes was consistent (Fig. S7). These results suggest that the acquisition of isoprene-producing function occurred through the stepwise accumulation of mutations, potentially driven by convergent molecular evolution in the TPS-b-t2 and TPS-b-t3 clades.

We next tested for evidence of positive selection in the clades that acquired F338 (predefined as foreground branches) by applying the branch-site model in codeml (29, 30). This approach identifies amino acid sites where the nonsynonymous to synonymous substitution rate ratio (ω) exceeds one, indicative

of positive selection. The analysis revealed significant evidence of positive selection in the foreground branches (likelihood ratio test, $P = 1.027\times10^{-6}$) with four residuals identified as candidates under positive selection (Bayes Empirical Bayes posterior probability > 0.95) (Table S4; Fig. S8). Notably, the residue corresponding to F338 was included among these candidates, suggesting that the convergent gain of this residue in the TPS-b-t2 and TPS-b-t3 clades was likely driven by positive selection.

***in vitro* enzyme assays demonstrated that the *IspS* evolved from monoterpene synthase.** To evaluate the evolutionary trajectory of enzymatic function, ancestral TPS sequences were artificially synthesized (Dataset S1) and subjected to enzymatic characterization using an *Escherichia coli* expression system. Because genes within the TPS-b-t3 clade exhibited high expression levels, we focused our analyses on this clade. Enzyme assays using purified recombinant proteins were conducted with either DMAPP (for IspS activity) or GPP (for mono-TPS activity) as a substrate. The ancestral enzyme at node N7 exhibited mono-TPS activity, specifically producing *trans*- and *cis*-β-ocimene, while the enzyme at node N8 exhibited IspS activity only (Fig. 4*C*, *D*). Ancestral proteins reconstructed using both PAML and GRaSP showed identical enzymatic activities at nodes N7 and N8 (Fig. 4*C*, *D*; Fig. S7*C*, *D*). These results indicate that *IspS* with an isoprene score of 4 evolved from β-ocimene synthases, supporting the functional shift during evolution.

Discussion

Using comparative genomics, ancestral-sequence reconstruction, and *in vitro* assays of ancestral enzymes, we found that *IspS* candidates with an isoprene score of 4 arose independently twice in the subgenus *Quercus*—once in the TPS-b-t3 clade and once in the TPS-b-t2 clade—within the TPS-b subfamily in Fagaceae. Because the subgenus *Quercus* diverged from its common ancestor roughly 56 million years ago (31), we infer that the onset of substantial isoprene emission in Fagaceae to the Eocene. Moreover, in all species we examined (*Q. robur*, *Q lobata*, and *Q. rubra*), once the IspS activity evolved, it has been retained. This contrasts with rosids, where a single early gain of isoprene synthase was followed by multiple independent losses (32). Fagaceae therefore provide a complementary pattern—multiple, convergent gains coupled with long-term retention—highlighting diverse evolutionary routes to isoprene emission in angiosperms.

The early TPS-b-t3 proteins preferred GPP and mainly produced β-ocimene, whereas their descendants evolved to favour DMAPP and became dedicated isoprene synthases in the subgenus *Quercus*. The shift would correlate with pocket-narrowing substitutions—most notably the substitution from Thr to Phe at the 338 residue—that exclude the larger GPP substrate. These results confirmed the previous view in which the ocimene synthases is ancestral TPS for which IspS evolved by active site size reduction following acquisition of F338 (33). The combination of branch-level and site-level likelihood tests detected positive selection at this residue, indicating adaptive convergence toward isoprene

emission. We also identified three additional residues under positive selection, suggesting that other sites may contribute to isoprene emission, although the effects of these sites on substrate specificity are currently unknown. Isoprene provides a threefold heat shield—membrane stabilization, ROS scavenging, and Ca^{2+}-mediated heat-shock signaling (34)—mechanisms that likely protect photosynthesis and sustain growth in *Quercus* species with thin, deciduous leaves under rising temperatures.

Terpenoids represent the largest and most diverse class of natural products in terrestrial plants (35). Comparative and functional genomics can chart the continuous evolutionary innovation of terpenoid-biosynthetic pathways with unprecedented resolution. Such insight will deepen our understanding of how chemical diversity has shaped plant adaptation and diversification on the earth.

Materials and Methods

Genome synteny analysis

To compare the genome structures of Fagaceae species across genera, published genome sequences of eight Fagaceae species from four genera were obtained from the public databases and used for subsequent analyses (Table S1). We first performed an all-vs-all BLASTp search using protein sequences with a cut-off e-value of 1E-5 to identify homologous relationships across the genomes. To investigate genomic synteny, we detected syntenic blocks within or between genomes using MCScanX (https://github.com/wyp1125/MCScanX), based on the results of the above BLASTp searches. The synteny blocks were visualized using SynVisio (https://synvisio.github.io). To clarify syntenic relationships across chromosomes in each species, we renumbered chromosomes according to the chromosomal order in *Q. robur*. Additionally, chromosomes oriented opposite to those of *Q. robur* were reversed. These adjusted chromosomes were used for subsequent analyses.

Phylogenetic analysis of TPS genes

Alignment and phylogenetic tree construction were conducted using the build function of ETE3 v3.1.3 (https://etetoolkit.org). Sequence alignment was performed with MAFFT v6.861b (https://mafft.cbrc.jp), alignment trimming with trimAl v1.4.rev (https://vicfero.github.io/trimal), and phylogenetic tree construction with RAxML v8.2.11 (https://cme.h-its.org/exelixis/web/software/raxml) using the GAMMA JTT model and 500 bootstrap replications. The resulting phylogenetic tree was visualized using the online tool iTOL (https://itol.embl.de). The analyzed TPS genes were classified into five subclade groups: TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g, based on reference genes from Arabidopsis and tomato. The chromosomal locations of these classified genes were visualized using TBtools-II v2.2.10 (https://github.com/CJ-Chen/TBtools-II).

Identification of TPS family members

We retrieved candidate genes of the TPS family based on two conserved motifs: the metal-binding domain (PF03936) and the N-terminal domain (PF01397). Protein sequences from eight Fagaceae species were used for a domain search

365 against the Pfam-A database ver 35.0 (36) using HMMER ver 3.3.2
366 (http://hmmer.org) with a cut-off e-value of 1E-5. For genes with more than two
367 isoforms, the longest isoform was selected.
368 To exclude candidate genes with short sequence lengths, we utilized the
369 conserved exon-intron arrangement patterns found in TPS genes (26). Across a
370 broad range of plant species, these patterns are well-documented and categorized
371 into three classes: Class I (12–14 introns), Class II (9 introns), and Class III (6
372 introns) (26). Notably, Class II TPS genes are specific to gymnosperms, meaning
373 that only Class I and Class III TPS genes are present in the angiosperm species
374 analyzed in this study. To identify putative TPS genes, we generated histograms
375 depicting the intron count for each gene within each species. The majority of genes
376 exhibited exon-intron patterns consistent with either Class I or Class III (Fig. S1).
377 Based on this distribution, we selected genes containing more than five exons.
378 Furthermore, some genes remained too short for detailed analysis. To
379 address this, genes encoding proteins shorter than 400 amino acids were excluded
380 (Fig. S2). To classify the subgroups of TPS genes, we incorporated protein
381 sequences of well-characterized TPS genes from *S. lycopersicum* and *A. thaliana*
382 as references for phylogenetic analysis. These reference sequences were
383 obtained from Table S7 in Zhou and Pichersky (27).
384 Additionally, *IspS* and *IspS-like* from *Q. serrata* were obtained from previous
385 study (28) and included in the analysis. To root the phylogenetic tree, we included
386 the *ent-copalyl diphosphate synthase* sequence from *Picea sitchensis* as an
387 outgroup.
388
389 **Phylogenetic analysis of TPS genes**
390 Alignment and phylogenetic tree construction were conducted using the build
391 function of ETE3 v3.1.3 (https://etetoolkit.org). Sequence alignment was
392 performed with MAFFT v6.861b (https://mafft.cbrc.jp), alignment trimming with
393 trimAl v1.4.rev (https://vicfero.github.io/trimal), and phylogenetic tree construction
394 with RAxML v8.2.11 (https://cme.h-its.org/exelixis/web/software/raxml) using the
395 GAMMA JTT model and 500 bootstrap replications. The resulting phylogenetic tree
396 was visualized using the online tool iTOL (https://itol.embl.de). The analyzed TPS
397 genes were classified into five subclade groups: TPS-a, TPS-b, TPS-c, TPS-e/f,
398 and TPS-g, based on reference genes from Arabidopsis and tomato. The
399 chromosomal locations of these classified genes were visualized using TBtools-II
400 v2.2.10 (https://github.com/CJ-Chen/TBtools-II).
401
402 **Isoprene synthase gene analysis**
403 To investigate molecular evolution of *IspS* in Fagaceae, we utilized published *IspS*
404 sequences from *S. lycopersicum* (27) and *Q. serrata* (28), both belonging to the
405 TPS-b subfamily. Since an adjacent subclade was located near the subclades
406 containing these sequences, we focused on the three subclades that include 44
407 Fagaceae genes.
408 These genes of three subclades were aligned by MAFFT v7.520
409 (https://mafft.cbrc.jp) with the reference *IspS* sequence from *Populus alba*

410 (AB198180) (37), *S. lycopersicum* (27) and *Q. serrata* (28) to calculate an isoprene
411 score (15).
412
413 **Sample collection for RNA sequencing**
414 To evaluate expression of TPS genes in natural conditions, we analyzed
415 transcriptome data obtained from leaf and bud tissues in three Fagaceae species,
416 *Q. acutissima*, *Q. glauca*, and *L. edulis*. We generated data for *Q. acutissima*,
417 whereas those for *Q. glauca* and *L. edulis* were obtained from the previous
418 study (25). For all data, samples were collected every four weeks spanning from
419 2020–2021 in *Q. acutissima*, 2022–2023 in *Q. glauca* and 2019–2021 in *L. edulis*
420 in Fukuoka prefectures in Japan. For each species, we targeted three individuals
421 and each sample was collected from three branches per individual, except for *L.*
422 *edulis*. Given the consistent expression patterns observed among individuals in a
423 previous study (24), we focused on only one individual for *L. edulis*. Study sites for
424 *Q. acutissima* and *Q. glauca* is the biodiversity reserve on the Ito campus of
425 Kyushu University (33º35′ N, 130º12′ E, 20 to 57 m a.s.l). *L. edulis* were studied at
426 the Imajyuku Field Activity Center (33º33′ N, 130º16′ E, 84 to 111 m a.s.l). Leaf
427 samples were collected from spring to summer in the deciduous species *Q.*
428 *acutissima*, whereas in the evergreen species *Q. glauca* and *L. edulis*, sampling
429 was conducted throughout the year.
430 RNA-seq samples were collected and preserved in 2 ml microtubes
431 containing 1.5 ml of RNA-stabilizing reagent (RNAlater; Thermo Fisher Scientific,
432 Waltham, MA, USA), then stored at −80°C until RNA extraction, following the
433 protocol described by Kudo et al. (25).
434
435 **RNA extraction and RNA-seq analysis**
436 Total RNA was extracted from leaf samples following the protocol described
437 previously (38). RNA integrity was assessed using the method outlined previously
438 (25). For each sample, 5–6 μg of total RNA was sent to Hangzhou Veritas Genetics
439 Medical Institute Co., Ltd., where cDNA libraries were prepared using the NEBNext
440 Ultra II RNA Library Prep Kit for Illumina. Transcriptome sequencing (150 bp
441 paired-end reads) was performed on an Illumina NovaSeq 6000 platform (Illumina,
442 San Diego, CA, USA). Gene expression quantification followed a three-step
443 pipeline: (1) quality filtering was performed using fastp v0.20.0 with default settings
444 (https://github.com/OpenGene/fastp); (2) RNA-seq reads were aligned to the
445 reference genome using TopHat v2.1.1
446 (https://ccb.jhu.edu/software/tophat/index.shtml); and (3) transcript abundance
447 was quantified using StringTie v1.3.5 (https://ccb.jhu.edu/software/stringtie).
448
449 **Seasonal expression patterns of TPS genes in Fagaceae**
450 To investigate the seasonal expression patterns of TPS genes, we extracted TPS
451 gene transcripts from RNA-seq data of *Q. acutissima*, *Q. glauca*, and *L. edulis*. For
452 each gene, we calculated the mean expression level in transcripts per million
453 (TPM) across three individuals, except for *L. edulis*, for which data were available
454 from only one individual (25). TPM values were transformed using $\log_2$(TPM +
455 0.01). Genes with a mean TPM > 1 across all samples were considered to be

expressed. A heatmap was generated in R ver. 4.3.1 using the 'pheatmap' package, with gene order arranged according to TPS clades (Fig. 3*A*). In addition to these three species, we also obtained expression data for *IspS* and *IspS*-like in *Q. serrata* (28), which were used to plot seasonal changes in expression levels. To generate the heatmap, we calculated the average gene expression levels for each month by aggregating data from different individuals over a two-year period for each species.

Ancestral amino acid sequence inference

To infer the ancestral amino acid sequences of putative *IspS* in *Quercus* species, we analyzed 95 representative nucleotide sequences, which have no ambiguity, of TPS-b subfamily members from Fagales. These sequences include genes identified through the genomic survey of eight Fagaceae species described above, genes from *Q. serrata*, and additional genes retrieved via BLAST searches against the database. The BLAST search identified genes from *Q. suber*, *L. litseifolius*, *Castanea mollissima*, *Juglans regia*, *J. maicrocarpa x J. regida*, *Carya illinoinensis*, *Corylus avellana*, and *Alnus glutinosa*, resulting in a dataset comprising genes from 17 Fagales species. (Table S3). Additionally, seven TPS-g subfamily genes were used as an outgroup.

Sequence alignment was performed using MACSE v2 (https://www.agap-ge2pop.org/macse), which can account for the underlying codon structure of protein-coding sequences and generate both nucleotide (NT) and amino acid (AA) alignments. Phylogenetic reconstruction and ancestral amino acid sequence inference were conducted separately for NT and AA alignments. ML phylogenetic trees were constructed with RAxML v8.2.11 (https://cme.h-its.org/exelixis/web/software/raxml) with 1000 bootstrap replications. The GAMMA GTR model (GTR+I+G) and the GAMMA JTT model (JTT+I+G) were used for the NT and AA alignment, respectively. The resulting ML trees exhibited compatible topology, particularly the splits of the major clusters. Two strongly supported clusters of *IspS* with the maximum score were identified in both trees. Ancestral amino acid sequences were inferred at each internal node using codeml in PAML (https://github.com/abacus-gene/paml) with the F3x4 codon model for the NT alignment. We also employed GRaSP (https://github.com/charles-abreu/GRaSP) to infer the ancestral sequences from the AA alignment and estimate the location of gaps in the ancestral sequences.

Detection of positive selection

To investigate whether amino acid sites in putative *IspS* (foreground) have undergone positive selection compared with the other TPS-b subfamily members (background), we applied the branch-site test for positive selection (29). The NT alignment and MLtree used for the ancestral sequence inference were employed in this analysis. Alignment gaps were removed prior to the analyses. Previous studies have shown that gene conversion and recombination events could lead to false positives (39, 40). To account for this, we applied the GARD method implemented in HYPHY (https://hyphy.org) to detect potential gene conversion or recombination events. No significant evidence of gene conversion or

502 recombination was found in the genes, so subsequent analyses were conducted
503 using the full-length alignments.
504 All branches within two clades of the genes with the isoprene score of 4
505 were set as foreground branches and the branch-site tests were conducted using
506 codeml (30). The null model was set to model = 2, NSsites = 2, fix_omega = 1 and
507 omega = 1, and the alternative model (Model A) was with model = 2, NSsites = 2
508 and fix_omega = 0 (to estimate ω_2). The Bayes empirical Bayes (BEB) method
509 was used to identify sites potentially under positive selection on the foreground
510 branches.
511
512 **Materials and reagents for *in vivo* enzymatic characterization**
513 DMAPP and GPP were used as substrates, with DMAPP purchased from
514 TargetMol (Boston, USA) and GPP from Sigma-Aldrich (St. Louis, USA). As
515 standard compounds, isoprene was obtained from Tokyo Chemical Industry Co.,
516 Ltd. (Tokyo, Japan), and β-ocimene was purchased from Cayman Chemical (Ann
517 Arbor, USA).
518
519 **Construction of *E. coli* expression vectors**
520 To express the ancestral sequences estimated at nodes N7 and N8 by PAML
521 (*N7_paml* and *N7_paml*) and GRaSP (*N7_grasp* and *N8_grasp*) in *E. coli*, their
522 putative transit peptides at N-terminus were removed and the truncated CDSs
523 were further codon-optimized for *E. coli*. The optimized sequences were
524 synthesized by GeneArt (Thermo Fisher Scientific). These sequences were
525 amplified by PCR using KOD One® (Toyobo, Osaka, Japan) with the primer pairs
526 shown in Table S5. The amplification program is composed of an initial
527 denaturation at 94°C for 1 min followed by 40 cycles of denaturation at 98°C for
528 10 sec, annealing at 51°C for 5 sec, and extension at 68°C for 10 sec, with a final
529 extension at 68°C for 5 min. The amplicons were individually inserted into
530 pET22b(+) vector (Novagen, Madison, US) by In-Fusion reaction (Takara Bio,
531 Kusatsu, Japan) using NdeI and BamHI sites, which enables the production of
532 proteins with the histidine tag (6×His). The sequence of *N8_paml* was constructed
533 by introducing a single mutation to *N8_grasp* by PCR using KOD One® (Toyobo)
534 and the primer pair *N8_paml*_Fw and *N8_paml*_Rv. This amplification was carried
535 out with an initial denaturation at 94°C for 1 min, followed by 40 cycles of
536 denaturation at 98°C for 10 sec, annealing at 51°C for 5 sec, and extension at
537 68°C for 40 sec, with a final extension at 68°C for 5 min. The amplicon was
538 introduced into the pET22b(+) vector in the same process as described above.
539
540 **Enzymatic characterization**
541 Constructs *ΔTP-N7_paml-6×His*, *ΔTP-N7_grasp-6×His*, *ΔTP-N8_paml-6×His*,
542 and *ΔTP-N8_grasp-6×His* were individually introduced into *E. coli* BL21 (DE3) for
543 recombinant protein production. Protein preparation was performed as previously
544 described (Koita et al.) (28). Protein concentrations were measured using a Qubit®
545 2.0 Fluorometer and the Qubit Protein Assay Kit (Thermo Fisher Scientific).
546 A standard reaction mixture (200 μl) contained 0.5 mM DMAPP or GPP, 20
547 mM $MgCl_2$, and 160 μl of purified enzyme. For assays using DMAPP, 27–32 μg of

enzyme were used, whereas 1.4–1.7 μg were used for GPP-based assays. Reactions were incubated at 37 °C for 30 minutes in 2 ml glass vials (Thermo Fisher Scientific). GC-MS analysis of the enzymatic products was carried out as described in Koita et al. (28).

Data availability

The RNAseq data are available from the PRJDB35868.

Acknowledgments

We thank Kayoko Ohta for her technical support with RNA extraction. We also thank Koharu Yamaguchi, Ryota Aoyagi, and Ryosuke Imai for their help with sample collection for RNAseq, Sangwan Kim for his support with registration of RNA sequence data to DDBJ, and Kenji Fukushima for his insightful comments on the bioinformatic analyses. This study was supported by grant-in-Aid for Transformative Research Areas “Plant-Climate Feedback” (no. JP23H04965 to A.S., K.Y., R.M., S.I., H.H; JP23H04966 to A.S., JP23H04967 to K.Y., R.M).

References

1. J. Kesselmeier, M. Staudt, Biogenic volatile organic compounds (VOC): An overview on emission, physiology and ecology. *J. Atmos. Chem.* **33**, 23-88 (1999).

2. J. Laothawornkitkul, J. E. Taylor, N. D. Paul, C. N. Hewitt, Biogenic volatile organic compounds in the Earth system. *New Phytologist.* **183**, 27-51 (2009).

3. A. Satake, *et al.*, Plant Molecular Phenology and Climate Feedbacks Mediated by BVOCs. *Annu Rev Plant Biol.* **75**, 605-627 (2024).

4. M. Antonelli, *et al.*, Forest volatile organic compounds and their effects on human health: A state-of-the-art review. *Int. J. Environ. Res. Public Health.* **17**, 6506 (2020).

5. N. Lee Ng, *et al.*, Nitrate radicals and biogenic volatile organic compounds: Oxidation, mechanisms, and organic aerosol. *Atmos Chem Phys.* **17**, 2103-2162 (2017).

6. A. Guenther, A global model of natural volatile organic compound emissions. *J. Geophys. Res.:Atmos.* **100**, 8873-8892 (1995).

7. A. B. Guenther, *et al.*, The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geosci. Model Dev.* **5**, 1471–1492 (2012).

8. P. C. Harley, R. K. Monson, M. T. Lerdau, Ecological and evolutionary aspects of isoprene emission from plants. *Oecologia.* **118**, 109-123 (1999).

9. R. K. Monson, R. T. Jones, T. N. Rosenstiel, J. P. Schnitzler, Why only some plants emit isoprene. *Plant Cell Environ.* **36**, 503–516 (2013).

10. K. G. S. Dani, I. M. Jamie, I. C. Prentice, B. J. Atwell, Evolution of isoprene emission capacity in plants. *Trends Plant Sci.* **19**, 439-446 (2014).

11. B. Miller, C. Oschinski, W. Zimmer, First isolation of an isoprene synthase gene from poplar and successful expression of the gene in Escherichia coli. *Planta.* **213**, 483-487 (2001).

12. D. T. Hanson, S. Swanson, L. E. Graham, T. D. Sharkey, Evolutionary significance of isoprene emission from mosses. *Am. J. Bot.* **86**, 634–639 (1999).

13. D. T. Tingey, R. C. Evans, E. H. Bates, M. L. Gumpertz, Isoprene emissions and photosynthesis in three ferns – The influence of light and temperature. *Physiol. Plant.* **69**, 609-616 (1987).

14. P. Harley, *et al.*, Variation in potential for isoprene emissions among Neotropical forest sites. *Glob Chang Biol.* **10**, 630-650 (2004).

15. T. D. Sharkey, D. W. Gray, H. K. Pell, S. R. Breneman, L. Topper, Isoprene synthase genes form a monophyletic clade of acyclic terpene synthases in the Tps-b terpene synthase family. *Evolution.* **67**, 1026–1040 (2013).

16. G. M. Silver, R. Fall, Characterization of aspen isoprene synthase, an enzyme responsible for leaf isoprene emission to the atmosphere. *J. Biol. Chem.* **270**, 13010-13016 (1995).

17. Q. Jia, *et al.*, Origin and early evolution of the plant terpene synthase family. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2100361119 (2022).

18. M. Li, *et al.*, In Planta Recapitulation of Isoprene Synthase Evolution from Ocimene Synthases. *Mol. Biol. Evol.* **34**, 2583–2599 (2017).

19. T. Kawakami, S. Miyazaki, H. Kawaide, Molecular characterization of a moss isoprene synthase provides insight into its evolution. *FEBS Lett.* **597**, 2133–2142 (2023).

20. F. Loreto, Distribution of isoprenoid emitters in the Quercus genus around the world: Chemo-taxonomical implications and evolutionary considerations based on the ecological function of the trait. *Perspect. Plant Ecol. Evol. Syst.* **5**, 185-192 (2002).

21. A. Tani, N. Masui, T. W. Chang, M. Okumura, Y. Kokubu, Basal emission rates of isoprene and monoterpenes from major tree species in Japan: interspecies and intraspecies variabilities. *Prog. Earth Planet. Sci.* **11**, 42 (2024).

22. F. Loreto, S. Fineschi, Reconciling functions and evolution of isoprene emission in higher plants. *New Phytol.* **206**, 578-582 (2015).

624 23. A. Kremer, *et al.*, Genomics of Fagaceae. *Tree Genet. Genomes.* **8**, 583-610
625 (2012).

626 24. A. Satake, K. Ohta, N. Takeda-Kamiya, K. Toyooka, J. Kusumi, Seasonal
627 gene expression signatures of delayed fertilization in Fagaceae. *Mol. Ecol.*
628 **32**, 4801-4813 (2023).

629 25. S. N. Kudo, *et al.*, The evolution of gene expression in seasonal
630 environments. bioRxiv [Preprint] (2025).
631 http://biorxiv.org/lookup/doi/10.1101/2025.03.25.645377 (accessed 20 April
632 2025).

633 26. S. C. Trapp, R. B. Croteau, Genomic Organization of Plant Terpene
634 Synthases and Molecular Evolutionary Implications. *Genetics.* **158**, 811–832.
635 (2001).

636 27. F. Zhou, E. Pichersky, The complete functional characterisation of the
637 terpene synthase family in tomato. *New Phytol.* **226**, 1341–1360 (2020).

638 28. S. Koita, *et al.*, Molecular basis behind the isoprene emission diversity in
639 Fagaceae. bioRxiv [Preprint] (2025).
640 http://biorxiv.org/lookup/doi/10.1101/2025.05.17.654689 (accessed 20 May
641 2025).

642 29. J. Zhang, R. Nielsen, Z. Yang, Evaluation of an improved branch-site
643 likelihood method for detecting positive selection at the molecular level. *Mol.*
644 *Biol. Evol.* **22**, 2472–2479 (2005).

645 30. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol.*
646 *Evol.* **24**, 1586–1591 (2007).

647 31. B. F. Zhou, *et al.*, Phylogenomic analyses highlight innovation and
648 introgression in the continental radiations of Fagaceae across the Northern
649 Hemisphere. *Nat. Commun.* **13**, 1320 (2022).

650 32. J. Shang, *et al.*, Evolution of the biosynthetic pathways of terpene scent
651 compounds in roses. *Curr. Biol.* **34**, 3550-3563.e8 (2024).

652 33. D. W. Gray, S. R. Breneman, L. A. Topper, T. D. Sharkey, Biochemical
653 characterization and homology modeling of methylbutenol synthase and
654 implications for understanding hemiterpene synthase evolution in plants. *J.*
655 *Biol. Chem.* **286**, 20582-20590 (2011).

656 34. Z. Zuo, S. M. Weraduwage, T. Huang, T. D. Sharkey, How volatile
657 isoprenoids improve plant thermotolerance. *Trends Plant Sci.*, in press. doi:
658 10.1016/J.TPLANTS.2025.05.004.

659 35. F. Chen, D. Tholl, J. Bohlmann, E. Pichersky, The family of terpene
660 synthases in plants: A mid-size family of genes for specialized metabolism

661 that is highly diversified throughout the kingdom. *Plant J.* **66**, 212–229
662 (2011).

663 36. J. Mistry, *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids*
664 *Res.* **49**, D412–D419 (2021).

665 37. K. Sasaki, K. Ohara, K. Yazaki, Gene expression and characterization of
666 isoprene synthase from Populus alba. *FEBS Lett.* **579**, 2514-2518 (2005).

667 38. Y. Miyazaki, *et al.*, Nitrogen as a key regulator of flowering in Fagus crenata:
668 Understanding the physiological mechanism of masting by gene expression
669 analysis. *Ecol. Lett.* **17**, 1299-1309 (2014).

670 39. M. Anisimova, R. Nielsen, Z. Yang, Effect of recombination on the accuracy
671 of the likelihood method for detecting positive selection at amino acid sites.
672 *Genetics.* **164**, 1229-1236 (2003).

673 40. C. Casola, M. W. Hahn, Gene conversion among paralogs results in
674 moderate false detection of positive selection using likelihood methods. *J.*
675 *Mol. Evol.* **68**, 679–687 (2009).

676
677

Figures and Tables

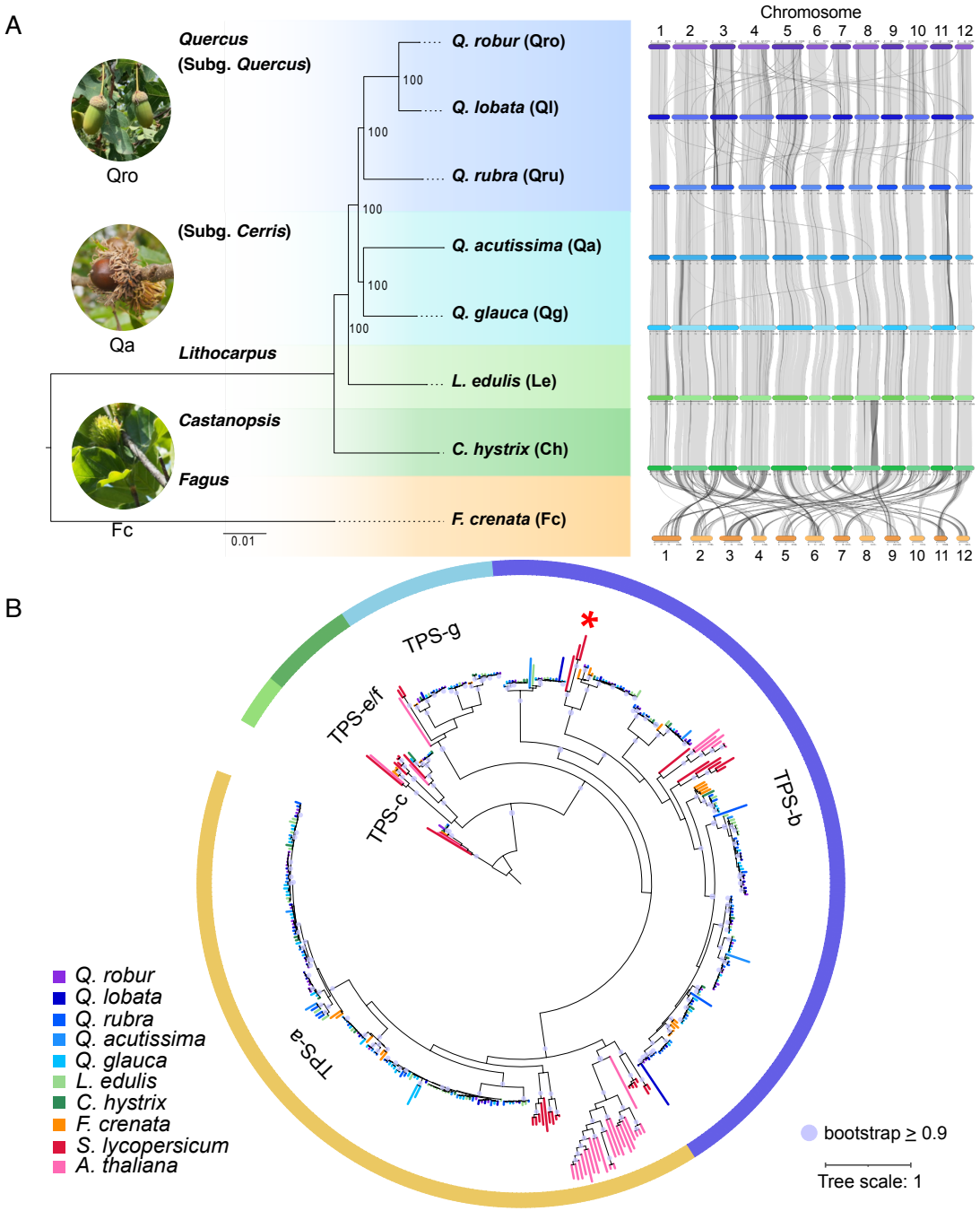


Figure 1. Syntenic relationships of Fagaceae genomes and the phylogenetic analysis of TPS genes. (*A*) Syntenic relationships among eight Fagaceae genomes. The phylogenetic tree was constructed based on whole-genome sequences from eight Fagaceae species, with bootstrap values shown at each branch. Chromosomes of each species are arranged along the corresponding branches of the phylogenetic tree. Syntenic relationships are illustrated with black and gray lines, where black indicates reverse alignments and gray represents standard (collinear) alignments. Chromosome numbers are labeled accordingly.

689 (*B*) Molecular phylogenetic tree of TPS genes in the eight Fagaceae species. TPS
690 genes from *S. lycopersicum* and *A. thaliana* were used as references to define five
691 TPS subclades (TPS-a, -b, -c, -e/f, and -g). The three TPS-b genes in tomato
692 including isoprene synthase were shown by red asterisk. The *ent*-copalyl
693 diphosphate (*ent*-CPP) synthase gene from *Picea sitchensis* was used as the
694 outgroup. The branches whose bootstrap values were 0.9 or more were shown by
695 purple circles.
696
697

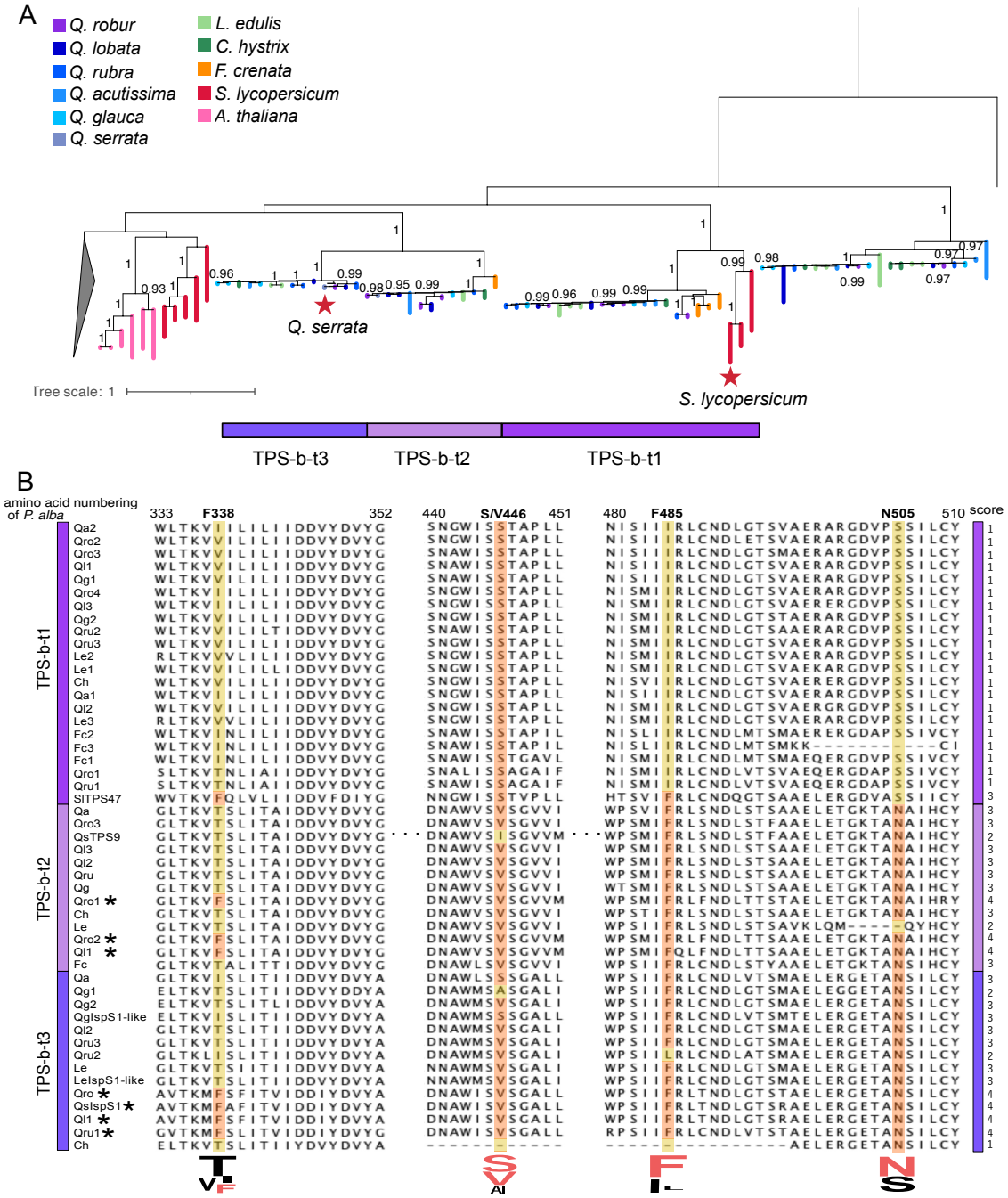


Figure 2. Phylogenetic tree and multiple sequence alignment of TPS genes in the TPS-b Subclade. (*A*) Molecular phylogenetic tree of TPS genes classified within the TPS-b subclade. The phylogeny was inferred with RAxML v8.2.11 under the GAMMA JTT substitution model, using 500 bootstrap replicates to assess node support. A TPS-g gene from *S. lycopersicum* (TPSgSlTPS37) was used as an outgroup. Two genes previously characterized with high isoprene synthase activity—score 3 in *S. lycopersicum* and score 4 in *Q. serrata*—are marked with red stars. (*B*) Multiple sequence alignment of TPS genes from the three TPS-b subclades: TPS-b-t1, TPS-b-t2, and TPS-b-t3. Four conserved amino acid positions used to compute isoprene scores are highlighted by red and yellow

709 shading, red means conserved amino acid at this position and yellow means
710 another. Genes with the highest isoprene score (score 4) are indicated by asterisks.
711
712

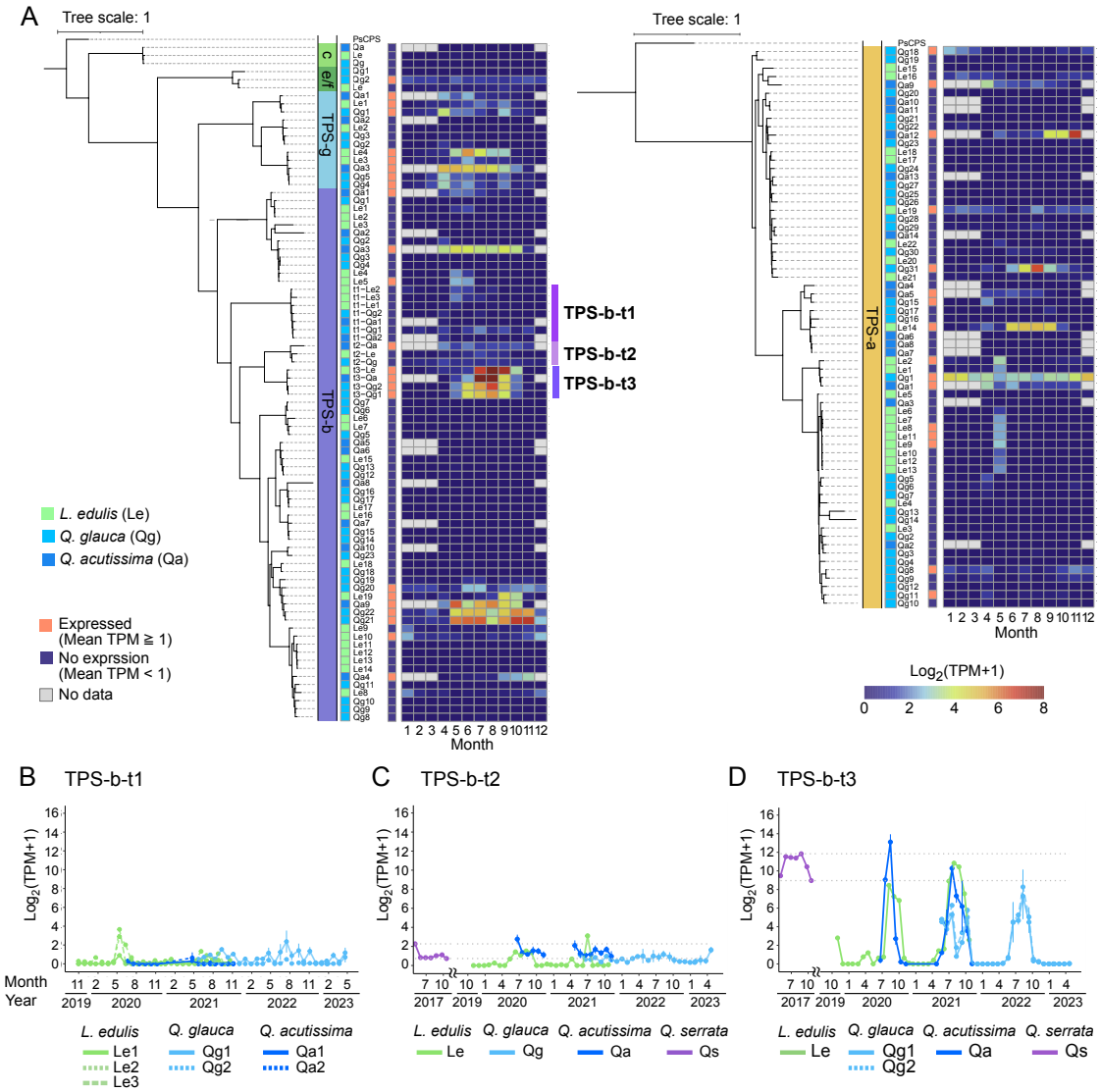
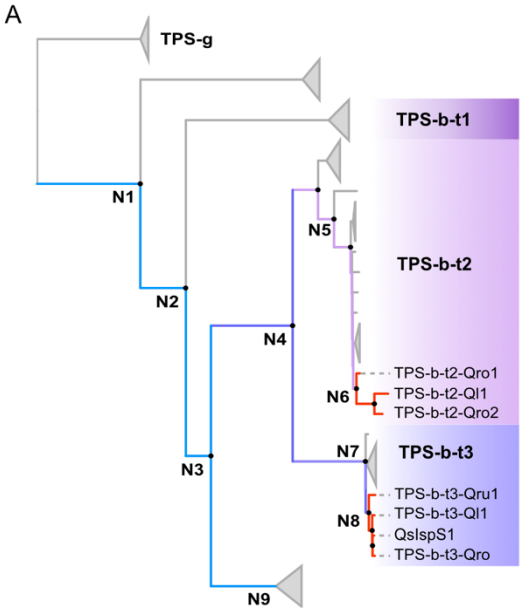


Figure 3. Seasonal expression patterns of TPS genes in leaves. (*A*) Heatmap showing the seasonal expression patterns of TPS genes in three Fagaceae species: *Q. acutissima* (blue), *Q. glauca* (light blue), and *L. edulis* (green). TPS-c genes are not included in the heatmap, as no TPS-c genes were consistently expressed in all three species. Gray squares indicate missing data due to leaf shedding. Genes were considered expressed if the mean of TPM values across all observation periods was ≥1. (*B-D*) Seasonal expression patterns of TPS genes within the subclades TPS-b-t1, TPS-b-t2, and TPS-b-t3. In addition to the three species mentioned above, data from *Q. serrata* were incorporated from a previous study (Koita et al. 2025).



B

Node	F338	S/V446	F485	N505	Isoprene score
N1	I	V	F	T	2
N2	I	V	F	T	2
N3	I	V	F	K	2
N4	T	V	F	N	3
N5	T	V	F	N	3
N6	F	V	F	N	4
N7	T	V	F	N	3
N8	F	V	F	N	4
N9	N	V	F	K	2

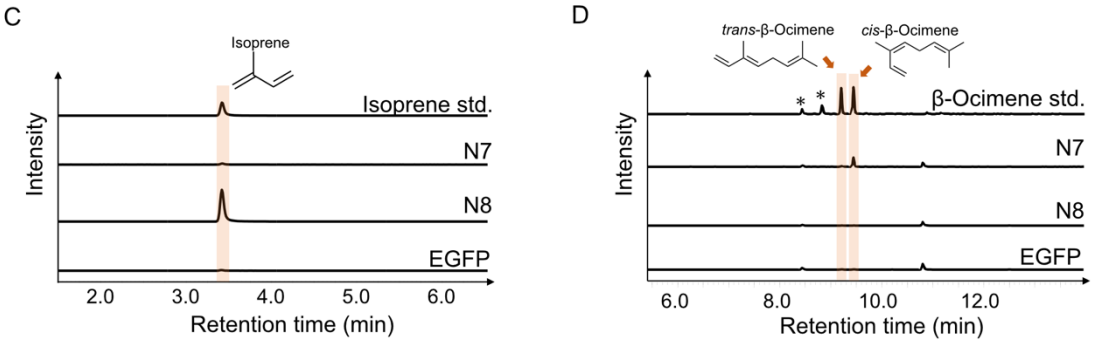


Figure 4. Results of ancestral sequence reconstruction of Fagaceae TPS genes using PAML. (*A*) Maximum-likelihood phylogeny of the TPS gene set used for ancestral reconstructions. The phylogeny was inferred with RAxML v8.2.11 under the GAMMA GTR substitution model, using 1,000 bootstrap replicates to assess node support. Internal nodes N1–N8 mark the bifurcation points analysed. Branch colours denote the inferred isoprene score—blue (2), purple (3) and red (4). The three focal clades (TPS-b-t1, TPS-b-t2 and TPS-b-t3) are highlighted in distinct colours for ease of reference. (*B*) Inferred amino-acid states at the diagnostic tetrad and the corresponding isoprene score for each ancestral node (N1–N8). The posterior probability (PP) of the most likely codon/amino-acid state was as follows: N1 – S/V446: GCA (Ala, PP = 0.263) → GTA (Val, PP = 0.600); F485: CTT (Leu, PP = 0.523) → TTT (Phe, PP = 0.405). N4 – T505: AAA (Lys, PP = 0.774) → AAT (Asn, PP = 0.956). N6 – F338: ACT (Thr, PP = 0.999) → TTT (Phe, PP = 0.970). N8 – F338: ACT (Thr, PP = 0.996) → TTT (Phe, PP = 0.990). (*C*) and (*D*) Enzyme assays of the ancestral sequence products using DMAPP and GPP as substrates. The reaction products were detected at *m/z* 67 for isoprene (*C*) and *m/z* 93 for monoterpenes (*D*) as the substrate. The *cis* or *trans* configurations of β-ocimene were identified based on spectral annotation. The chromatograms, except for the

745 β-ocimene standard, are plotted on the same scale. Asterisks represent peaks of
746 impurities.
747
748
749 **Additional files**
750
751 **Supplementary file 1**
752 Supplementary figures (Fig. S1 to Fig. S8) and supplementary tables (Table S1 to
753 Table S5) were included in this file.
754
755 **Supplementary Dataset**
756 **Dataset S1**
757 Multiple sequence alignment of ancestral sequences of Fagaceae TPS genes
758 using PAML and GRaSP. The four conserved amino acid residues used for
759 calculating isoprene score were shown by black frame.
760