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 authors:**

Akiko Satake

Email: satake.akiko.269@m.kyushu-u.ac.jp

Junko Kusumi

Email: jkusumi@scs.kyushu-u.ac.jp

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 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*. Ancestral sequence reconstruction revealed that the acquisition of a diagnostic amino acid residue for IspS function 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 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.

Introduction

Plants release a wide spectrum of biogenic volatile organic compounds (BVOCs), which include molecules with diverse structural and physicochemical characteristics (Kesselmeier and Staudt, 1999; Laothawornkitkul et al., 2009; Satake et al., 2024). These compounds, emitted from various plant tissues such as leaves, flowers, fruits, and roots, enter the atmosphere in significant quantities (Antonelli et al., 2020). BVOCs serve as a dynamic interface between terrestrial ecosystems and the atmosphere, playing an important role in climate and atmospheric chemistry (Lee Ng et al., 2017). The annual global emission of BVOCs is estimated to be around 1000 Tg C, with terpenoids—including isoprene, monoterpenes, and sesquiterpenes (Guenther, 1995; Guenther et al., 2012).

Among these, isoprene alone accounts for roughly half of the total BVOC, underscoring its ecological and atmospheric significance (Guenther, 1995).

In the plant kingdom, isoprene emission is distributed heterogeneously across plant taxa (Dani et al., 2014; Harley et al., n.d.; Monson et al., 2013). It is particularly common in fast-growing plants, especially hardwood trees such as oaks (*Quercus* spp.), *Populus* spp. (Miller et al., 2001), and *Eucalyptus* spp. (Dani et al., 2014). Isoprene emission has also been reported in mosses (Hanson et al., 1999), ferns (Tingey et al., 1987), and legumes (Harley et al., 2004), 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 (Sharkey et al., 2013; Silver and Fall, 1995). The TPS family likely originated in ancestral land plants following their divergence from green algae (Jia et al., 2022). In angiosperms, IspS belongs to the TPS-b subfamily (Li et al., 2017; Sharkey et al., 2013). In contrast, recent research has shown that IspS in mosses belongs to the TPS-c subfamily (Kawakami et al., 2023), 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 (Li et al., 2017; Sharkey et al., 2013).

An intriguing and well-studied example of the uneven distribution of isoprene emission within a plant clade can be found in the Fagaceae family (Harley et al., n.d.; Monson et al., 2013). 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 (Loreto, 2002; Monson et al., 2013; Tani et al., 2024). 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* (Loreto and Fineschi, 2015). Based on analyses associating isoprene emission capacity with molecular phylogeny within the Fagaceae, Loreto, 2022 and Monson *et al.*, 2013 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 (Kremer et al., 2012; Kudo et al., 2025; Satake et al., 2023), 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 species each from the genera *Lithocarpus* (*L. edulis*), *Castanopsis* (*C. hystrix*), and *Fagus* (*F. crenata*) (Table S1). Isoprene emission factors vary significantly among these eight species (Table S2), highlighting their suitability for studying the evolution of isoprene emission. A comparative genomic analysis of these species revealed a highly conserved syntenic structure across genera, including *Quercus*, *Lithocarpus*, and *Castanopsis* (Fig. 1*A*). However, the genomic structure of the genus *Fagus* differed significantly from that of the other genera (Fig. 1*A*), suggesting extensive genome rearrangements following its divergence from the rest of the Fagaceae.

To identify TPS genes in our target Fagaceae species, we conducted HMMER searches using the annotated protein sequences from the genomes of eight Fagaceae species against the metal-binding domain (PF03936) and the N-terminal domain (PF01397). Each TPS candidate gene was further examined and filtered based on exon-intron structure, including the number and locations of exons (Fig. S1), by comparing with previously

120 reported TPSs (Trapp & Corteau), as well as total sequence length (Fig. S2). The number
121 of identified TPS genes varied across species, ranging from a minimum of 24 in *C. hystrix*
122 to a maximum of 67 in *Q. glauca* (Fig. S3).

123 The rooted TPS phylogenetic tree constructed from our TPS dataset along with
124 sequences from *Solanum lycopersicum* (tomato) and *Arabidopsis thaliana*, revealed a clear
125 distinction among the TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g subfamilies (Fig. 1*B*).
126 TPS-a members in Fagaceae formed a distinct clade, separate from those of tomato and *A.*
127 *thaliana*, but showing a closer evolutionary relationship to tomato TPS-a genes (Fig. 1*B*).
128 The TPS-c, TPS-g, and TPS-e/f subfamilies were present in all Fagaceae species, except
129 for the TPS-e/f subfamily, which was absent in *Q. acutissima* (Fig. 1*B*).

130 The TPS-b subfamily represents the largest gene group within Fagaceae species
131 (Fig. 1*B*). The three TPS-b genes in tomato form a distinct clade from other tomato and
132 Arabidopsis TPS-b genes (red asterisk in Fig. 1*B*). These three tomato genes are known to
133 synthesize acyclic monoterpenes, with one of them, TPS47, encoding a cytosolic isoprene
134 synthase (Zhou and Pichersky, 2020).

135 Many genes in the TPS-a and TPS-b subfamilies are arranged in tandem gene arrays
136 on chromosomes 8 and 9, respectively, suggesting their origin from tandem gene
137 duplication events (Fig. S4*A*). Overall, the genomic locations of these genes were largely
138 conserved across *Quercus*, *Lithocarpus*, and *Castanopsis*. However, in *Fagus*, significant
139 genome rearrangements were observed, indicating a distinct evolutionary trajectory (Fig.
140 S4*B*).

141

142 **Isoprene score of TPS-b genes in Fagaceae.** To investigate the evolutionary history of
143 *IspS* in Fagaceae, we focused on the TPS-b subfamily that includes known *IspS* from other
144 plant species. We designated the clade containing the tomato *IspS* as TPS-b-target1 (TPS-
145 b-t1) and the clade containing the *Q. serrata IspS* as TPS-b-target1 (TPS-b-t3) (Fig. 2*A*).
146 Between the TPS-b-t1 and TPS-b-t3, we identified a distinct clade, which we designate as
147 TPS-b-target2 (TPS-b-t2) (Fig. 2*A*). Notably, both TPS-b-t2 and TPS-b-t3 consist

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

150 Sequences of genes included in these three clades were aligned and used to
151 calculate an isoprene score (Fig. 2*B*). The isoprene score serves as a useful screening tool
152 to identify enzymes with exclusive IspS activity based on sequence information and the
153 structural analysis of IspS (Sharkey et al., 2013). This score is calculated using the four
154 conserved amino acid residues (known as the diagnostic tetrad) that are characteristic of
155 IspS. Two of the residues (F338 and F485 in Fig. 2*B*) were suggested to restrict the size of
156 the substrate-binding pocket and play a key role in determining enzyme specificity
157 (Sharkey et al., 2013). While F338 and F485 of the diagnostic tetrad residues are
158 considered highly specific to IspS, the other positions are considered redundant (Li et al.,
159 2017). In IspS in some angiosperm families, valine (V446) is found instead of serine (Li et
160 al., 2017). Therefore, we treated both serine and valine as conserved residues at this
161 position and denoted it as S/V446. The isoprene score ranges from 0 to 4, with a score of
162 4 representing full conservation of the diagnostic tetrad and thus a high likelihood of
163 exclusive IspS activity.

164 The multiple sequence alignment revealed that the isoprene score for sequences in
165 the TPS-b-t1 clade was consistently one, whereas the IspS sequence in tomato had a score
166 of 3 (Fig. 2*B*). In the TPS-b-t2 and TPS-b-t3 clades, sequences from the species in the
167 subgenus *Quercus* (*Q. robur*, *Q. lobata*, *Q. rubra*, and *Q. serrata*) were found to have an
168 isoprene score of 4. Specifically, *Q. robur* contained three, *Q. lobata* had two, and *Q. rubra*
169 had one sequence with an isoprene score of 4. In contrast, the maximum isoprene score for
170 sequences from *Lithocarpus*, *Castanopsi*s, and *Fagus* was only three. Many sequences
171 lacked F338, an amino acid considered crucial for isoprene synthase activity (Sharkey et
172 al., 2013). These patterns in isoprene score align with isoprene emission capacities, where
173 *Q. robur*, *Q. lobata*, *Q. rubra*, and *Q. serrata* are known to exhibit high isoprene emission,
174 while other species show lower emission levels (Table S2)

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

179 isoprene score of 4, were located on chromosome 3 (Fig. S4), suggesting chromosomal
180 translocation events.
181
182 **Seasonal expression patterns of TPS genes.** To investigate the seasonal expression
183 patterns of TPS genes, we conducted RNA-seq on leaf and bud tissues collected monthly
184 over a two-year period from *Q. acutissima* growing under natural conditions in Fukuoka,
185 Japan. Additionally, we incorporated previously published field transcriptome data from
186 *Q. glauca* and *L. edulis* (Kudo et al., 2025), which were sampled at the same study site as
187 *Q. acutissima*. On average, 36% and 41% of TPS genes were expressed in leaves (Fig. 3*A*)
188 and buds (Fig. S6*A*), respectively. Within the TPS-b clade, 42% of genes were expressed
189 in leaves, including 6 out of 14 in *Q. acutissima*, 10 out of 28 in *Q. glauca*, and 2 out of 24
190 in *L. edulis*. Seasonal expression profiles in leaves revealed that most expressed genes
191 exhibited peak expression in summer, although a small fraction of genes showed peaked
192 expression during autumn–winter (e.g., Qa12 and Qg1 in the TPS-a clade in Fig. 3*A*).
193 Notably, expression patterns differed markedly among the three target TPS subclades i.e.,
194 TPS-b-t1, t2, and t3. In leaves, only genes in the TPS-b-t3 clade showed clear and
195 consistent summer expression peaks (Fig. 3*D*), whereas expression levels in the TPS-b-t1
196 and TPS-b-t2 clades were low in all species (Fig. 3*B, C*). These results suggest that TPS-
197 b-t3 genes are primarily responsible for terpenoid biosynthesis. Furthermore, the
198 expression pattern of *Q, serrata IspS* (Koita et al., 2025) within the TPS-b-t3 clade also
199 showed strong expression (Fig. 3*B-D*), providing additional support for the functional role
200 of this clade in isoprene biosynthesis. In contrast, in bud tissues, expression levels of genes
201 within the TPS-b-t3 clade were consistently low, and genes in the other two clades, TPS-
202 b-t1 and t2 were relatively high (Fig. S6*B-D*), implying a tissue-specific regulation of TPS
203 gene expression.
204
205 **Convergent evolution and positive selection in *IspS* candidates of Fagaceae.** To
206 determine whether the genes with an isoprene score of 4 were gained or lost during
207 evolution, we employed the ancestral sequence reconstruction (ASR) method. This
208 approach relies on statistical principles to infer the sequences of putative ancestral genes
209 based on the amino acid or nucleotide sequences of extant species. For this analysis, we

210 used nucleotide sequences from 17 Fagales species belonging to the TPS-b clade (Table
211 S3) and reconstructed putative ancestral sequences using codeml implemented in PAML
212 (https://github.com/abacus-gene/paml).
213 The estimated ancestral sequences of the last common node of the TPS-b-t1, TPS-
214 b-t2, and TPS-b-t3 clades (node N2; Fig. 4*A*) had an isoprene score of 2, with the conserved
215 residuals S/V446 and F485 (Fig. 4*B*). Subsequently, N505 was gained in the common
216 ancestor of the TPS-b-t2 and TPS-b-t3 clades (node N4; Fig. 4*A*, *B*), prior to their
217 duplication. Following this duplication, F338 was acquired independently in each of the
218 TPS-b-t2 and TPS-b-t3 clades (nodes N6 and 8; Fig. 4*A B*), specifically in the lineage
219 leading to the three species within subgenus *Quercus* (*Q. robur*, *Q lobata*, and *Q. rubra*).
220
221 Notably, the gain of F338 occurred independently in both the TPS-b-t2 and the
222 TPS-b-t3 clades, despite requiring two nucleotide substitutions for the transition from ACT
223 (T) to TTT (F). We also employed GRaSP (https://github.com/charles-abreu/GRaSP) to
224 infer the ancestral sequences and reached the same conclusion that the gain of F338
225 occurred independently in both the TPS-b-t2 and TPS-b-t3 clades (Fig. S7). Although the
226 inferred ancestral sequences for the distant ancestor at nodes N1–N3 differed slightly
227 between PAML and GRaSP, the evolutionary pattern across the remaining nodes was
228 consistent (Fig. S7). These results suggest that the acquisition of isoprene-producing
229 function occurred through the stepwise accumulation of mutations, potentially driven by
230 convergent molecular evolution in the TPS-b-t2 and TPS-b-t3 clades.
231 We next tested for evidence of positive selection in the clades that acquired F338
232 (predefined as foreground branches) by applying the branch-site model in codeml (Yang,
233 2007; Zhang et al., 2005). This approach identifies amino acid sites where the
234 nonsynonymous to synonymous substitution rate ratio (ω) exceeds one, indicative of
235 positive selection. The analysis revealed significant evidence of positive selection in the
236 foreground branches (likelihood ratio test, $P = 1.027\times10^{-6}$) with four residuals identified
237 as candidates under positive selection (Bayes Empirical Bayes posterior probability > 0.95
238) (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 (Zhou et al., 2022), 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 (Shang et al., 2024). 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 (Gray et al., 2011). 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 (Zuo et al., 2025)—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 (Chen et al., 2011). 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

300 chromosomes according to the chromosomal order in *Q. robur*. Additionally,
301 chromosomes oriented opposite to those of *Q. robur* were reversed. These adjusted
302 chromosomes were used for subsequent analyses.
303
304 **Phylogenetic analysis of TPS genes**
305 Alignment and phylogenetic tree construction were conducted using the build function of
306 ETE3 v3.1.3 (https://etetoolkit.org). Sequence alignment was performed with MAFFT
307 v6.861b (https://mafft.cbrc.jp), alignment trimming with trimAl v1.4.rev
308 (https://vicfero.github.io/trimal), and phylogenetic tree construction with RAxML v8.2.11
309 (https://cme.h-its.org/exelixis/web/software/raxml) using the GAMMA JTT model and
310 500 bootstrap replications. The resulting phylogenetic tree was visualized using the online
311 tool iTOL (https://itol.embl.de). The analyzed TPS genes were classified into five subclade
312 groups: TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g, based on reference genes from
313 Arabidopsis and tomato. The chromosomal locations of these classified genes were
314 visualized using TBtools-II v2.2.10 (https://github.com/CJ-Chen/TBtools-II).
315
316 **Identification of TPS family members**
317 We retrieved candidate genes of the TPS family based on two conserved motifs: the metal-
318 binding domain (PF03936) and the N-terminal domain (PF01397). Protein sequences from
319 eight Fagaceae species were used for a domain search against the Pfam-A database ver
320 35.0 (Mistry et al., 2021) using HMMER ver 3.3.2 (http://hmmer.org) with a cut-off e-
321 value of 1E-5. For genes with more than two isoforms, the longest isoform was selected.
322 To exclude candidate genes with short sequence lengths, we utilized the conserved
323 exon-intron arrangement patterns found in TPS genes (Trapp and Croteau, 2001). Across
324 a broad range of plant species, these patterns are well-documented and categorized into
325 three classes: Class I (12–14 introns), Class II (9 introns), and Class III (6 introns) (Trapp
326 & Croteau, 2001). Notably, Class II TPS genes are specific to gymnosperms, meaning that
327 only Class I and Class III TPS genes are present in the angiosperm species analyzed in this
328 study. To identify putative TPS genes, we generated histograms depicting the intron count
329 for each gene within each species. The majority of genes exhibited exon-intron patterns

330 consistent with either Class I or Class III (Fig. S1). Based on this distribution, we selected
331 genes containing more than five exons.
332 Furthermore, some genes remained too short for detailed analysis. To address this,
333 genes encoding proteins shorter than 400 amino acids were excluded (Fig. S2). To classify
334 the subgroups of TPS genes, we incorporated protein sequences of well-characterized TPS
335 genes from *S. lycopersicum* and *A. thaliana* as references for phylogenetic analysis. These
336 reference sequences were obtained from Table S7 in Zhou and Pichersky (Zhou and
337 Pichersky, 2020).
338 Additionally, *IspS* and *IspS-like* from *Q. serrata* were obtained from previous study
339 (Koita et al., 2025)and included in the analysis. To root the phylogenetic tree, we included
340 the *ent-copalyl diphosphate synthase* sequence from *Picea sitchensis* as an outgroup.
341
342 **Phylogenetic analysis of TPS genes**
343 Alignment and phylogenetic tree construction were conducted using the build function of
344 ETE3 v3.1.3 (https://etetoolkit.org). Sequence alignment was performed with MAFFT
345 v6.861b (https://mafft.cbrc.jp), alignment trimming with trimAl v1.4.rev
346 (https://vicfero.github.io/trimal), and phylogenetic tree construction with RAxML v8.2.11
347 (https://cme.h-its.org/exelixis/web/software/raxml) using the GAMMA JTT model and
348 500 bootstrap replications. The resulting phylogenetic tree was visualized using the online
349 tool iTOL (https://itol.embl.de). The analyzed TPS genes were classified into five subclade
350 groups: TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g, based on reference genes from
351 Arabidopsis and tomato. The chromosomal locations of these classified genes were
352 visualized using TBtools-II v2.2.10 (https://github.com/CJ-Chen/TBtools-II).
353
354 **Isoprene synthase gene analysis**
355 To investigate molecular evolution of *IspS* in Fagaceae, we utilized published *IspS*
356 sequences from *S. lycopersicum* (Zhou and Pichersky, 2020) and *Q. serrata* (Koita et al.,
357 2025), both belonging to the TPS-b subfamily. Since an adjacent subclade was located near
358 the subclades containing these sequences, we focused on the three subclades that include
359 44 Fagaceae genes.

These genes of three subclades were aligned by MAFFT v7.520 (https://mafft.cbrc.jp) with the reference *IspS* sequence from *Populus alba* (AB198180) (Sasaki et al., 2005), *S. lycopersicum* (Zhou and Pichersky, 2020)and *Q. serrata* (Koita et al., 2025) to calculate an isoprene score (Sharkey et al., 2013).

Sample collection for RNA sequencing

To evaluate expression of TPS genes in natural conditions, we analyzed transcriptome data obtained from leaf and bud tissues in three Fagaceae species, *Q. acutissima*, *Q. glauca*, and *L. edulis*. We generated data for *Q. acutissima*, whereas those for *Q. glauca* and *L. edulis* were obtained from the previous study (Kudo et al., 2025). For all data, samples were collected every four weeks spanning from 2020–2021 in *Q. acutissima*, 2022–2023 in *Q. glauca* and 2019–2021 in *L. edulis* in Fukuoka prefectures in Japan. For each species, we targeted three individuals and each sample was collected from three branches per individual, except for *L. edulis*. Given the consistent expression patterns observed among individuals in a previous study (Satake et al., 2023), we focused on only one individual for *L. edulis*. Study sites for *Q. acutissima* and *Q. glauca* is the biodiversity reserve on the Ito campus of Kyushu University (33º35′ N, 130º12′ E, 20 to 57 m a.s.l). *L. edulis* were studied at the Imajyuku Field Activity Center (33º33′ N, 130º16′ E, 84 to 111 m a.s.l). Leaf samples were collected from spring to summer in the deciduous species *Q. acutissima*, whereas in the evergreen species *Q. glauca* and *L. edulis*, sampling was conducted throughout the year.

RNA-seq samples were collected and preserved in 2 ml microtubes containing 1.5 ml of RNA-stabilizing reagent (RNAlater; Thermo Fisher Scientific, Waltham, MA, USA), then stored at −80°C until RNA extraction, following the protocol described by Kudo *et al.*, 2025.

RNA extraction and RNA-seq analysis

Total RNA was extracted from leaf samples following the protocol described previously (Miyazaki et al., 2014). RNA integrity was assessed using the method outlined previously (Kudo et al., 2025). For each sample, 5–6 µg of total RNA was sent to Hangzhou Veritas Genetics Medical Institute Co., Ltd., where cDNA libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina. Transcriptome sequencing (150 bp

391 paired-end reads) was performed on an Illumina NovaSeq 6000 platform (Illumina, San
392 Diego, CA, USA). Gene expression quantification followed a three-step pipeline: (1)
393 quality filtering was performed using fastp v0.20.0 with default settings
394 (https://github.com/OpenGene/fastp); (2) RNA-seq reads were aligned to the reference
395 genome using TopHat v2.1.1 (https://ccb.jhu.edu/software/tophat/index.shtml); and (3)
396 transcript abundance was quantified using StringTie v1.3.5
397 (https://ccb.jhu.edu/software/stringtie).
398
399 **Seasonal expression patterns of TPS genes in Fagaceae**
400 To investigate the seasonal expression patterns of TPS genes, we extracted TPS gene
401 transcripts from RNA-seq data of *Q. acutissima*, *Q. glauca*, and *L. edulis*. For each gene,
402 we calculated the mean expression level in transcripts per million (TPM) across three
403 individuals, except for *L. edulis*, for which data were available from only one individual
404 (Kudo et al., 2025). TPM values were transformed using $\log_2(\text{TPM} + 1)$. Genes with a
405 mean TPM > 1 across all samples were considered to be expressed. A heatmap was
406 generated in R ver. 4.3.1 using the 'pheatmap' package, with gene order arranged according
407 to TPS clades (Fig. 3*A*). In addition to these three species, we also obtained expression data
408 for *IspS* and *IspS*-like in *Q. serrata* (Koita et al., 2025), which were used to plot seasonal
409 changes in expression levels. To generate the heatmap, we calculated the average gene
410 expression levels for each month by aggregating data from different individuals over a two-
411 year period for each species.
412
413 **Ancestral amino acid sequence inference**
414 To infer the ancestral amino acid sequences of putative *IspS* in *Quercus* species, we
415 analyzed 95 representative nucleotide sequences, which have no ambiguity, of TPS-b
416 subfamily members from Fagales. These sequences include genes identified through the
417 genomic survey of eight Fagaceae species described above, genes from *Q. serrata*, and
418 additional genes retrieved via BLAST searches against the database. The BLAST search
419 identified genes from *Q. suber*, *L. litseifolius*, *Castanea mollissima*, *Juglans regia*, *J.*
420 *maicrocarpa x J. regida*, *Carya illinoinensis*, *Corylus avellana*, and *Alnus glutinosa*,

421 resulting in a dataset comprising genes from 17 Fagales species. (Table S3). Additionally,
422 seven TPS-g subfamily genes were used as an outgroup.
423 Sequence alignment was performed using MACSE v2 (https://www.agap-
424 ge2pop.org/macse), which can account for the underlying codon structure of protein-
425 coding sequences and generate both nucleotide (NT) and amino acid (AA) alignments.
426 Phylogenetic reconstruction and ancestral amino acid sequence inference were conducted
427 separately for NT and AA alignments. ML phylogenetic trees were constructed with
428 RAxML v8.2.11 (https://cme.h-its.org/exelixis/web/software/raxml) with 1000 bootstrap
429 replications. The GAMMA GTR model (GTR+I+G) and the GAMMA JTT model
430 (JTT+I+G) were used for the NT and AA alignment, respectively. The resulting ML trees
431 exhibited compatible topology, particularly the splits of the major clusters. Two strongly
432 supported clusters of *IspS* with the maximum score were identified in both trees. Ancestral
433 amino acid sequences were inferred at each internal node using codeml in PAML
434 (https://github.com/abacus-gene/paml) with the F3x4 codon model for the NT alignment.
435 We also employed GRaSP (https://github.com/charles-abreu/GRaSP) to infer the ancestral
436 sequences from the AA alignment and estimate the location of gaps in the ancestral
437 sequences.
438
439 **Detection of positive selection**
440 To investigate whether amino acid sites in putative *IspS* (foreground) have undergone
441 positive selection compared with the other TPS-b subfamily members (background), we
442 applied the branch-site test for positive selection (Zhang et al., 2005). The NT alignment
443 and MLtree used for the ancestral sequence inference were employed in this analysis.
444 Alignment gaps were removed prior to the analyses. Previous studies have shown that gene
445 conversion and recombination events could lead to false positives (Anisimova et al., 2003;
446 Casola and Hahn, 2009). To account for this, we applied the GARD method implemented
447 in HYPHY (https://hyphy.org) to detect potential gene conversion or recombination events.
448 No significant evidence of gene conversion or recombination was found in the genes, so
449 subsequent analyses were conducted using the full-length alignments.
450 All branches within two clades of the genes with the isoprene score of 4 were set
451 as foreground branches and the branch-site tests were conducted using codeml (Yang,

452 2007). The null model was set to model = 2, NSsites = 2, fix_omega = 1 and omega = 1,
453 and the alternative model (Model A) was with model = 2, NSsites = 2 and fix_omega = 0
454 (to estimate ω_2). The Bayes empirical Bayes (BEB) method was used to identify sites
455 potentially under positive selection on the foreground branches.
456
457 **Materials and reagents for *in vivo* enzymatic characterization**
458 DMAPP and GPP were used as substrates, with DMAPP purchased from TargetMol
459 (Boston, USA) and GPP from Sigma-Aldrich (St. Louis, USA). As standard compounds,
460 isoprene was obtained from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan), and β-
461 ocimene was purchased from Cayman Chemical (Ann Arbor, USA).
462
463 **Construction of *E. coli* expression vectors**
464 To express the ancestral sequences estimated at nodes N7 and N8 by PAML (*N7_paml* and
465 *N7_paml*) and GRaSP (*N7_grasp* and *N8_grasp*) in *E. coli*, their putative transit peptides
466 at N-terminus were removed and the truncated CDSs were further codon-optimized for *E.*
467 *coli*. The optimized sequences were synthesized by GeneArt (Thermo Fisher Scientific).
468 These sequences were amplified by PCR using KOD One® (Toyobo, Osaka, Japan) with
469 the primer pairs shown in Table S5. The amplification program is composed of an initial
470 denaturation at 94°C for 1 min followed by 40 cycles of denaturation at 98°C for 10 sec,
471 annealing at 51°C for 5 sec, and extension at 68°C for 10 sec, with a final extension at 68°C
472 for 5 min. The amplicons were individually inserted into pET22b(+) vector (Novagen,
473 Madison, US) by In-Fusion reaction (Takara Bio, Kusatsu, Japan) using NdeI and BamHI
474 sites, which enables the production of proteins with the histidine tag (6×His). The sequence
475 of *N8_paml* was constructed by introducing a single mutation to *N8_grasp* by PCR using
476 KOD One® (Toyobo) and the primer pair *N8_paml*_Fw and *N8_paml*_Rv. This
477 amplification was carried out with an initial denaturation at 94°C for 1 min, followed by
478 40 cycles of denaturation at 98°C for 10 sec, annealing at 51°C for 5 sec, and extension at
479 68°C for 40 sec, with a final extension at 68°C for 5 min. The amplicon was introduced
480 into the pET22b(+) vector in the same process as described above.
481
482 **Enzymatic characterization**

483 Constructs *ΔTP-N7_paml-6×His*, *ΔTP-N7_grasp-6×His*, *ΔTP-N8_paml-6×His*, and *ΔTP-*
484 *N8_grasp-6×His* were individually introduced into *E. coli* BL21 (DE3) for recombinant
485 protein production. Protein preparation was performed as previously described (Koita *et*
486 *al.*, 2025). Protein concentrations were measured using a Qubit® 2.0 Fluorometer and the
487 Qubit Protein Assay Kit (Thermo Fisher Scientific).
488 A standard reaction mixture (200 µl) contained 0.5 mM DMAPP or GPP, 20 mM
489 $MgCl_2$, and 160 µl of purified enzyme. For assays using DMAPP, 27–32 µg of enzyme
490 were used, whereas 1.4–1.7 µg were used for GPP-based assays. Reactions were incubated
491 at 37 °C for 30 minutes in 2 ml glass vials (Thermo Fisher Scientific). GC-MS analysis of
492 the enzymatic products was carried out as described in Koita *et al.*, 2025.
493

494 ## Acknowledgments

495 We thank Kayoko Ohta for her technical support with RNA extraction. We also thank
496 Koharu Yamaguchi, Ryota Aoyagi, and Ryosuke Imai for their help with sample collection
497 for RNAseq, Sangwan Kim for his support with registration of RNA sequence data to
498 DDBJ, and Kenji Fukushima for his insightful comments on the bioinformatic analyses.
499

500 ## Additional information

501 **Data availability**

502 The RNAseq data that support the findings of this study are openly available in DDBJ,
503 Bioproject ID: PRJDB35868.
504

505 **Funding**

506 This study was supported by grant-in-Aid for Transformative Research Areas "Plant-
507 Climate Feedback" (no. JP23H04965 to A.S., K.Y., R.M., S.I., H.H; JP23H04966 to A.S.,
508 JP23H04967 to K.Y., R.M).
509

510 **Competing Interests**

511 The authors declare no conflict of interests.
512

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.

Additional files

Supplementary Information. Supplementary figures (Fig. S1 to Fig. S8) and supplementary tables (Table S1 to S5) were included int this file.

Supplementary 1.

Supplementary Dataset. Supplementary dataset (Dataset S1) was included in this file. It is multiple sequence alignment of ancestral sequences of Fagaceae TPS genes using PAML and GRaSP.

Dataset S1.

References

Anisimova M, Nielsen R, Yang Z (2003) **Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites.** *Genetics* **164**:1229-1236. doi:10.1093/genetics/164.3.1229.

Antonelli M, Donelli D, Barbieri G, Valussi M, Maggini V, Firenzuoli F (2020) **Forest volatile organic compounds and their effects on human health: A state-of-the-art review.** *Int J Environ Res Public Health* **17**:6506. doi:10.3390/ijerph17186506.

Casola C, Hahn MW (2009) **Gene conversion among paralogs results in moderate false detection of positive selection using likelihood methods.** *J Mol Evol* **68**: 679–687. doi:10.1007/s00239-009-9241-6.

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) **The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom.** *Plant Journal* **66**:212–229. doi:10.1111/j.1365-313X.2011.04520.x.

Dani KGS, Jamie IM, Prentice IC, Atwell BJ (2014) **Evolution of isoprene emission capacity in plants.** *Trends Plant Sci* **19**: 439-446. doi:10.1016/j.tplants.2014.01.009.

Gray DW, Breneman SR, Topper LA, Sharkey TD (2011) **Biochemical characterization and homology modeling of methylbutenol synthase and implications for understanding hemiterpene synthase evolution in plants.** *Journal of Biological Chemistry* **286:**20582-20590. doi:10.1074/jbc.M111.237438.

Guenther A (1995) **A global model of natural volatile organic compound emissions.** *J Geophys Res* **100:**8873-8892. doi:10.1029/94JD02950.

Guenther AB, Jiang X, Heald CL, Sakulyanontvittaya T, Duhl T, Emmons LK, Wang X (2012) **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. doi:10.5194/gmd-5-1471-2012.

Hanson DT, Swanson S, Graham LE, Sharkey TD (1999) **Evolutionary significance of isoprene emission from mosses.** *Am J Bot* **86**:634–639. doi:10.2307/2656571.

Harley P, Vasconcellos P, Vierling L, Pinheiro CCDS, Greenberg J, Guenther A, Klinger L, De Almeida SS, Neill D, Baker T, Phillips O, Malhi Y (2004) **Variation in potential for isoprene emissions among Neotropical forest sites.** *Glob Chang Biol* **10:**630-650. doi:10.1111/j.1529-8817.2003.00760.x.

Harley PC, Monson RK, Lerdau MT (1999) **Ecological and evolutionary aspects of isoprene emission from plants.** *Oecologia* **118**:109-123. doi: 10.1007/s004420050709.

Jia Q, Brown R, K TG, Fu J, Chen X, Ka-Shu Wong G, Gershenzon J, Peters RJ, Chen F (2022) **Origin and early evolution of the plant terpene synthase family.** *Proc. Natl. Acad. Sci. U. S. A* 119: e2100361119. doi:10.1073/pnas.2100361119.

Kawakami T, Miyazaki S, Kawaide H (2023) **Molecular characterization of a moss isoprene synthase provides insight into its evolution.** *FEBS Lett* **597**:2133–2142. doi:10.1002/1873-3468.14691.

Kesselmeier J, Staudt M (1999) **Biogenic volatile organic compounds (VOC): An overview on emission, physiology and ecology.** *J Atmos Chem* **33**:23-88. doi:10.1023/A:1006127516791.

574 Koita S, Munakata R, Fukushima K, Nagano A, Ikezaki Y, Satake A, Saito T, Miura K,
575 Sugiyama A, Yazaki K (2025) **Molecular basis behind the isoprene emission**
576 **diversity in Fagaceae.** *bioRxiv* doi:10.1101/2025.05.17.654689.
577 Kremer A, Abbott AG, Carlson JE, Manos PS, Plomion C, Sisco P, Staton ME, Ueno S,
578 Vendramin GG (2012) **Genomics of Fagaceae.** *Tree Genet Genomes* **8**:583-610.
579 doi:10.1007/s11295-012-0498-3.
580 Kudo SN, Ikezaki Y, Kusumi J, Hirakawa H, Isobe S, Satake A (2025) **The evolution of**
581 **gene expression in seasonal environments.** *bioRxiv*
582 doi:10.1101/2025.03.25.645377.
583 Laothawornkitkul J, Taylor JE, Paul ND, Hewitt CN (2009) **Biogenic volatile organic**
584 **compounds in the Earth system: Tansley review.** *New Phytologist* **183**:27-51.
585 doi:10.1111/j.1469-8137.2009.02859.x.
586 Lee Ng N, Brown SS, Archibald AT, Atlas E, Cohen RC, Crowley JN, Day DA,
587 Donahue NM, Fry JL, Fuchs H, Griffin RJ, Guzman MI, Herrmann H, Hodzic A,
588 Iinuma Y, Kiendler-Scharr A, Lee BH, Luecken DJ, Mao J, McLaren R, Mutzel A,
589 Osthoff HD, Ouyang B, Picquet-Varrault B, Platt U, Pye HOT, Rudich Y,
590 Schwantes RH, Shiraiwa M, Stutz J, Thornton JA, Tilgner A, Williams BJ, Zaveri
591 RA (2017) **Nitrate radicals and biogenic volatile organic compounds: Oxidation,**
592 **mechanisms, and organic aerosol.** *Atmos Chem Phys* **17**:2103-2162.
593 doi:10.5194/acp-17-2103-2017.
594 Li M, Xu J, Algarra Alarcon A, Carlin S, Barbaro E, Cappellin L, Velikova V, Vrhovsek
595 U, Loreto F, Varotto C (2017) **In Planta Recapitulation of Isoprene Synthase**
596 **Evolution from Ocimene Synthases.** *Mol Biol Evol* **34**:2583–2599.
597 doi:10.1093/molbev/msx178.
598 Loreto F (2002) **Distribution of isoprenoid emitters in the Quercus genus around the**
599 **world: Chemo-taxonomical implications and evolutionary considerations based**
600 **on the ecological function of the trait.** *Perspect Plant Ecol Evol Syst* **5**:185-192.
601 doi:10.1078/1433-8319-00033.
602 Loreto F, Fineschi S (2015) Reconciling functions and evolution of isoprene emission in
603 higher plants. *New Phytologist* **206**:578-582. doi:10.1111/nph.13242.

604 Miller B, Oschinski C, Zimmer W (2001) **First isolation of an isoprene synthase gene**
605 **from poplar and successful expression of the gene in Escherichia coli.** *Planta*
606 **213**:483-487. doi:10.1007/s004250100557.
607 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL,
608 Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A (2021) **Pfam:**
609 **The protein families database in 2021.** *Nucleic Acids Res* **49:** D412–D419.
610 doi:10.1093/nar/gkaa913.
611 Miyazaki Y, Maruyama Y, Chiba Y, Kobayashi MJ, Joseph B, Shimizu KK, Mochida K,
612 Hiura T, Kon H, Satake A (2014) **Nitrogen as a key regulator of flowering in**
613 **Fagus crenata: Understanding the physiological mechanism of masting by gene**
614 **expression analysis.** *Ecol Lett* **17:**1299-1309. doi:10.1111/ele.12338.
615 Monson RK, Jones RT, Rosenstiel TN, Schnitzler JP (2013) **Why only some plants emit**
616 **isoprene.** *Plant Cell Environ* **36**:503–516. doi:10.1111/pce.12015.
617 Sasaki K, Ohara K, Yazaki K (2005) **Gene expression and characterization of isoprene**
618 **synthase from Populus alba.** *FEBS Lett* **579:**2514-2518.
619 doi:10.1016/j.febslet.2005.03.066.
620 Satake A, Hagiwara T, Nagano AJ, Yamaguchi N, Sekimoto K, Shiojiri K, Sudo K
621 (2024) **Plant Molecular Phenology and Climate Feedbacks Mediated by**
622 **BVOCs.** *Annu Rev Plant Biol* **75:**605-627. doi:10.1146/annurev-arplant-060223-
623 032108.
624 Satake A, Ohta K, Takeda-Kamiya N, Toyooka K, Kusumi J (2023) **Seasonal gene**
625 **expression signatures of delayed fertilization in Fagaceae.** *Mol Ecol* **32**:4801-
626 4813. doi:10.1111/mec.17079.
627 Shang J, Feng D, Liu H, Niu L, Li R, Li Y, Chen M, Li A, Liu Z, He Y, Gao X, Jian H,
628 Wang C, Tang K, Bao M, Wang J, Yang S, Yan H, Ning G (2024) **Evolution of the**
629 **biosynthetic pathways of terpene scent compounds in roses.** *Current Biology*
630 **34**:3550-3563.e8. doi:10.1016/j.cub.2024.06.075.
631 Sharkey TD, Gray DW, Pell HK, Breneman SR, Topper L (2013) **Isoprene synthase**
632 **genes form a monophyletic clade of acyclic terpene synthases in the Tps-b**
633 **terpene synthase family.** *Evolution (N Y)* **67**:1026–1040. doi:10.1111/evo.12013.

634 Silver GM, Fall R (1995) **Characterization of aspen isoprene** synthase, an enzyme
635 responsible for leaf isoprene emission to the atmosphere. *Journal of Biological*
636 *Chemistry* **270**:13010-13016. doi:10.1074/jbc.270.22.13010.
637 Tani A, Masui N, Chang TW, Okumura M, Kokubu Y (2024) **Basal emission rates of**
638 **isoprene and monoterpenes from major tree species in Japan: interspecies and**
639 **intraspecies variabilities.** *Prog Earth Planet Sci* **11**:42. doi:10.1186/s40645-024-
640 00645-8.
641 Tingey DT, Evans RC, Bates EH, Gumpertz ML (1987) **Isoprene emissions and**
642 **photosynthesis in three ferns – The influence of light and temperature.** *Physiol*
643 *Plant* **69**:609-616. doi:10.1111/j.1399-3054.1987.tb01974.x.
644 Trapp SC, Croteau RB (2001) **Genomic Organization of Plant Terpene Synthases and**
645 **Molecular Evolutionary Implications.** *Genetics* **158**:811-832. doi:
646 10.1093/genetics/158.2.811.
647 Yang Z (2007) **PAML 4: Phylogenetic analysis by maximum likelihood.** *Mol Biol*
648 *Evol* **24**:1586–1591. doi:10.1093/molbev/msm088.
649 Zhang J, Nielsen R, Yang Z (2005) **Evaluation of an improved branch-site likelihood**
650 **method for detecting positive selection at the molecular level.** *Mol Biol Evol*
651 **22**:2472-2479. doi:10.1093/molbev/msi237.
652 Zhou BF, Yuan S, Crowl AA, Liang YY, Shi Y, Chen XY, An QQ, Kang M, Manos PS,
653 Wang B (2022) **Phylogenomic analyses highlight innovation and introgression in**
654 **the continental radiations of Fagaceae across the Northern Hemisphere.** *Nat*
655 *Commun* **13**:1320. doi:10.1038/s41467-022-28917-1.
656 Zhou F, Pichersky E (2020) **The complete functional characterisation of the terpene**
657 **synthase family in tomato.** *New Phytologist* **226**:1341–1360.
658 doi:10.1111/nph.16431.
659 Zuo Z, Weraduwage SM, Huang T, Sharkey TD (2025) **How volatile isoprenoids**
660 **improve plant thermotolerance.** *Trends Plant Sci* in press.
661 doi:10.1016/J.TPLANTS.2025.05.004.
662
663

Figures

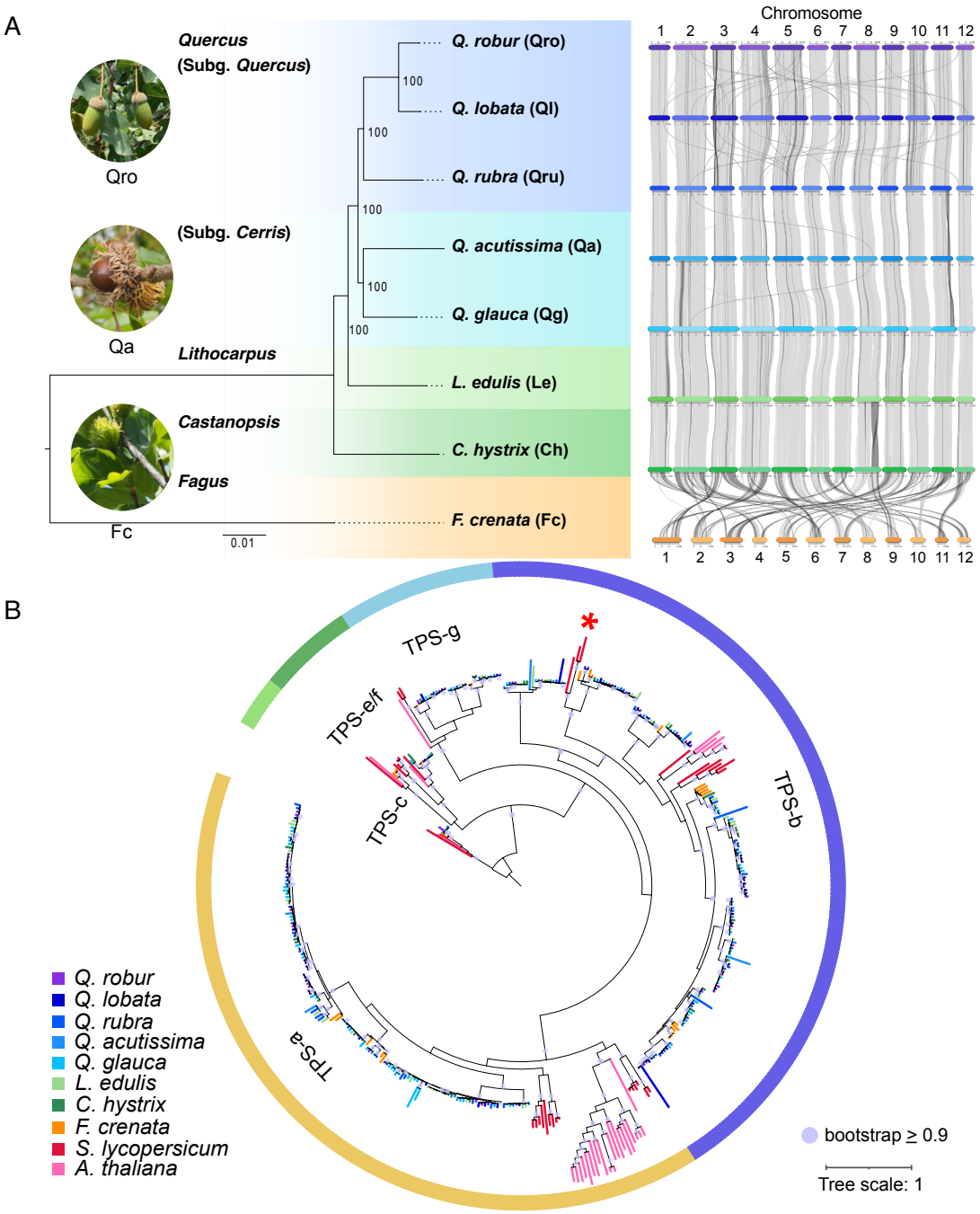


Figure 1. Syntenic relationships of Fagaceae genomes and the phylogenetic analysis of TPS genes. (*A*) Syntenic relationships among eight Fagaceae genomes. The phylogenetic tree was constructed based on whole-genome sequences from eight Fagaceae species, with bootstrap values shown at each branch. Chromosomes of each species are arranged along the corresponding branches of the phylogenetic tree. Syntenic relationships are illustrated

671 with black and gray lines, where black indicates reverse alignments and gray represents
672 standard (collinear) alignments. Chromosome numbers are labeled accordingly. (*B*)
673 Molecular phylogenetic tree of TPS genes in the eight Fagaceae species. TPS genes from
674 *S. lycopersicum* and *A. thaliana* were used as references to define five TPS subclades (TPS-
675 a, -b, -c, -e/f, and -g). The three TPS-b genes in tomato including isoprene synthase were
676 shown by red asterisk. The *ent*-copalyl diphosphate (*ent*-CPP) synthase gene from *Picea*
677 *sitchensis* was used as the outgroup. The branches whose bootstrap values were 0.9 or more
678 were shown by purple circles.
679

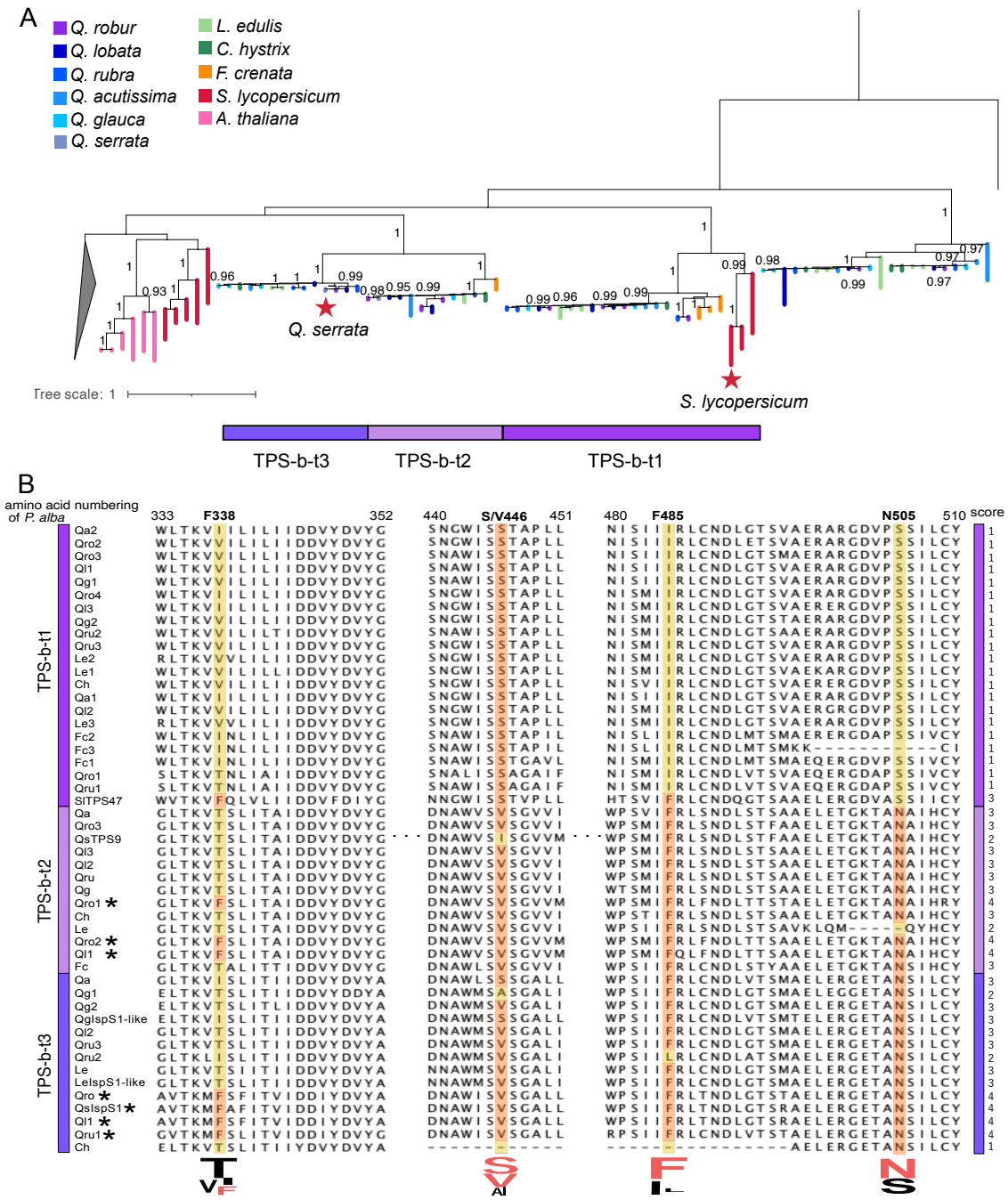


Figure 2. Phylogenetic tree and multiple sequence alignment of TPS genes in the TPS-b Subclade. (*A*) Molecular phylogenetic tree of TPS genes classified within the TPS-b subclade. The phylogeny was inferred with RAxML v8.2.11 under the GAMMA JTT substitution model, using 500 bootstrap replicates to 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

687 in *Q. serrata*—are marked with red stars. (*B*) Multiple sequence alignment of TPS genes
688 from the three TPS-b subclades: TPS-b-t1, TPS-b-t2, and TPS-b-t3. Four conserved amino
689 acid positions used to compute isoprene scores are highlighted by red and yellow shading,
690 red means conserved amino acid at this position and yellow means another. Genes with the
691 highest isoprene score (score 4) are indicated by asterisks.
692

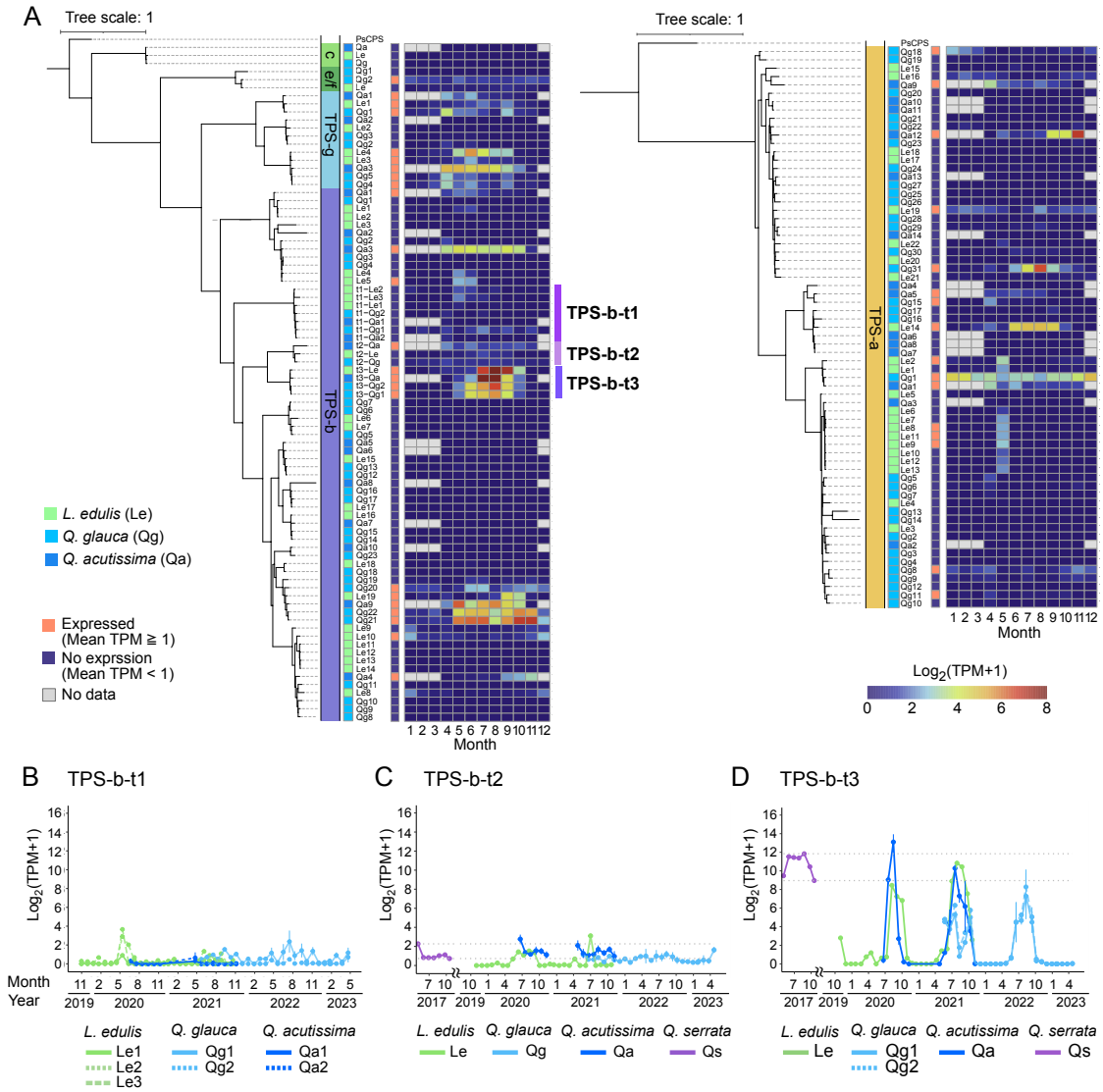
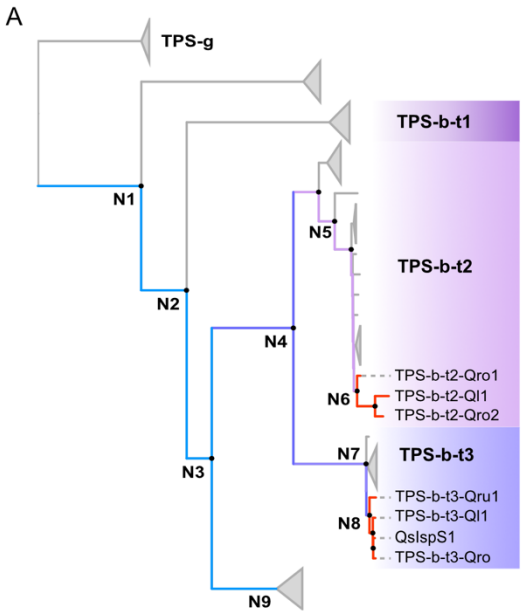
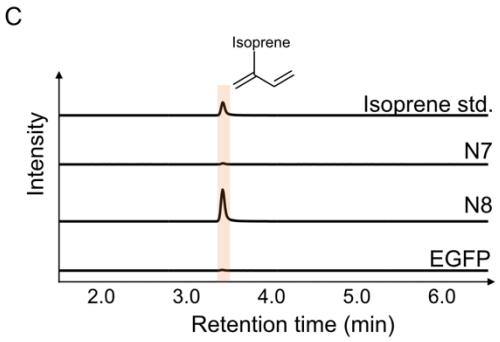


Figure 3. Seasonal expression patterns of TPS genes in leaves. (*A*) Heatmap showing the seasonal expression patterns of TPS genes in three Fagaceae species: *Q. acutissima* (blue), *Q. glauca* (light blue), and *L. edulis* (green). TPS-c genes are not included in the heatmap, as no TPS-c genes were 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



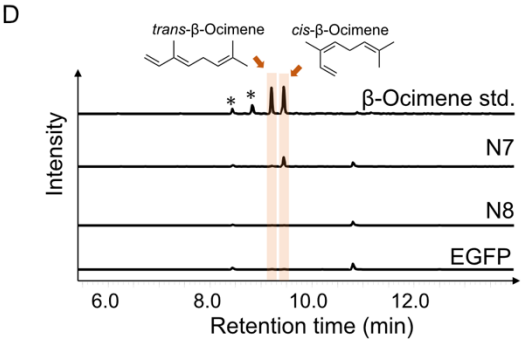


Figure 4. Results of ancestral sequence reconstruction of Fagaceae TPS genes using PAML. (*A*) Maximum-likelihood phylogeny of the TPS gene set used for ancestral reconstructions. The phylogeny was inferred with RAxML v8.2.11 under the GAMMA GTR substitution model, using 1,000 bootstrap replicates to 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 =

717 0.970). N8 – F338: ACT (Thr, PP = 0.996) → TTT (Phe, PP = 0.990). (*C*) and (*D*) Enzyme
718 assays of the ancestral sequence products using DMAPP and GPP as substrates. The
719 reaction products were detected at *m/z* 67 for isoprene (*C*) and *m/z* 93 for monoterpenes
720 (*D*) as the substrate. The *cis* or *trans* configurations of β-ocimene were identified based on
721 spectral annotation. The chromatograms, except for the β-ocimene standard, are plotted on
722 the same scale. Asterisks represent peaks of impurities.
723