The evolution of gene expression in seasonal environments

Shuichi N. Kudo[1], Yuka Ikezaki[2], Junko Kusumi[3], Hideki Hirakawa[4,5], Sachiko Isobe[4,6], Akiko Satake[2]*

1. Graduate School of Systems Life Sciences, Kyushu University, Fukuoka 819-0395, Japan
2. Department of Biology, Faculty of Science, Kyushu University, Fukuoka 819-0395, Japan
3. Department of Environmental Changes, Faculty of Social and Cultural Studies, Kyushu University, Fukuoka 819-0395, Japan.
4. Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
5. Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Fukuoka 819-0395, Japan
6. Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, 113-8657, Japan

*Corresponding author: Akiko Satake

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

Abstract

The biological activities of organisms are closely linked to seasonality. Phenology, the temporal orchestration of biological activities, is governed by gene expression, yet the evolutionary dynamics underlying seasonal gene expression remain unclear. To investigate these dynamics, we compared genome-wide expression dynamics (molecular phenology) in four dominant evergreen Fagaceae species in Asia (*Quercus glauca*, *Q. acuta*, *Lithocarpus edulis*, and *L. glaber*), using leaf and bud tissues over two seasonal cycles. We assembled high-quality reference genomes, identifying 11749 single-copy orthologous genes. Seasonal transcriptomic profiling of these orthologous genes revealed highly conserved gene expression across species in winter when temperatures fall below ~10°C. Rhythmic gene expression with significant periodic oscillations was more prevalent in buds (51.9%) than in leaves (40.6%), with most rhythmic genes (78.4–92.0%) exhibiting annual periodicity, while a smaller fraction (1.2–11.9%) followed half-annual cycles. The seasonal peaks of rhythmic genes were highly synchronized across species in winter but diverged during the growing season, reflecting species-specific timing of leaf flushing and flowering. These findings suggest that the four species share a common molecular calendar in winter, which constrains the evolution of gene expression under seasonal environments.

Introduction

Fundamental biological processes, such as growth, mortality, and reproduction, rarely remain constant over time; instead, they exhibit pronounced variation in response to seasonal environmental changes (Fretwell, 1972). This temporal organization of a series of biological activities, known as phenology, plays a critical role in the evolution of life histories in response to seasonal fluctuating environments (Forrest and Miller-Rushing, 2010; Morellato et al., 2013; Piao et al., 2019; Schwartz, 2013). Phenology is regulated by the coordinated expression of genes that enable organisms to exhibit essential functions at the appropriate times (Kudoh, 2016; Satake et al., 2024a, 2022). However, gene expression dynamics underpinning seasonal responses in organisms remain not fully understood.

To explore the evolution of gene expression, transcriptome analyses are widely employed to compare transcript profiles across organs, sexes, species, and developmental stages within a given environment. Such studies have been conducted in mammals (Brawand et al., 2011; Cardoso-Moreira et al., 2019; Naqvi et al., 2019), *Drosophila* (Lemos et al., 2005; Nuzhdin et al., 2004), and Tanzanian cichlids (El Taher et al., 2021), providing key insights into the evolutionary dynamics of gene expression and its central role in shaping phenotypic diversity. These studies have revealed that gene expression profiles tend to be more conserved across species within the same tissue, whereas the evolutionary rate of gene expression varies among organs. In mammals, for instance, nervous tissue exhibits a slower rate of gene expression change compared to the testis (Brawand et al., 2011; Cardoso-Moreira et al., 2019; Naqvi et al., 2019), suggesting the

presence of tissue-specific functional constraints that influence the evolution of gene expression.

In addition to tissue-dependent constraints, seasonal environmental changes may play a crucial role in shaping the evolution of gene expression, particularly by influencing the temporal coordination of gene expression under natural conditions. Gene expression responses to certain seasons may be more resistant to change, whereas in other seasons, divergence in gene expression timing across species may occur, contributing to temporal niche differentiation. Despite its potential importance, the influence of seasonal environmental fluctuations on the gene expression evolution remains poorly understood, largely because most transcriptomic analyses have been conducted under tightly controlled laboratory conditions.

To investigate the evolution of gene expression in seasonal environments, we performed a molecular phenology analysis on Fagaceae tree species, which are widely distributed across the Northern Hemisphere, with its center of diversity in subtropical Southeast Asia (Hipp et al., 2020; Kremer et al., 2012; Manos et al., 2001; Zhou et al., 2022). Molecular phenology refers to transcriptome analysis conducted under natural fluctuating conditions (Komoto et al., 2024; Kudoh, 2016; Nagano et al., 2019; Satake et al., 2023) and provides a valuable framework for studying gene–environment interactions and their role in orchestrating the temporal biological functions essential for adaptation to environmental fluctuations (Satake et al., 2022). In this study, we selected four evergreen Fagaceae species—two *Quercus* species (*Q. glauca* and *Q. acuta*) and two *Lithocarpus* species (*L. edulis* and *L. glaber*)—that co-occur in the same locality in Japan. *Q. glauca*, *L. edulis* and *L. glaber* inhabit nearly identical environment, while *Q. acuta* is restricted to the high-altitude regions with slightly colder environment compared to the other three species. Leaf flushing and flowering phenology vary across species, with *Q. glauca* and *Q. acuta* flowering in early spring, almost simultaneously with leaf flushing, *L. edulis* flowering in late spring, and *L. glaber* flowering in autumn, approximately one to two months after leaf flushing. This setting enabled us to compare seasonal gene expression profiles across species with diverse phenology, both within and between genera, under nearly identical environmental fluctuations. By assessing and comparing seasonal transcriptomic dynamics across Fagaceae species, our study provides new insights into the evolution of gene expression and adaptation to seasonal natural environments.

Results

Genome assemblies of *Q. glauca* and *L. edulis*. To quantify transcript abundance of non-model Fagaceae species, we first constructed reference genomes for *Q. glauca*, as a representative of genus *Quercus*, and *L. edulis,* as a representative of genus *Lithocarpus* (Fig. 1A). To obtain high-quality chromosome-level reference genomes, we conducted a hybrid genome assembly for the HiFi reads and the high-throughput chromosome conformation capture (Hi-C) reads (Table S1).The resultant scaffolds were 867.7 and 835.7 Mb and consisted of 797 and 750 scaffolds including 12 chromosome-level scaffolds with contig N50 of 71.7

Mb and 68.8 Mb for *Q. glauca* and *L. edulis*, respectively (Table S1). Approximately 94.8% and 97.2% of the final assemblies were assigned to 12 pseudochromosomes of *Q. glauca* and *L. edulis*, respectively (Figure 1B). The number of chromosome-scale scaffolds were identical to the known numbers in other *Quercus* (Zoldos et al., 1999) and *Lithocarpus* species (Liu et al., 2023). The genomes were estimated to contain 39758 and 34059 protein-coding genes (33947 and 30246 genes were annotated within 12 chromosomes), covering 98.3% and 95.0% of complete Benchmarking Universal Single Copy Orthologs (BUSCO) genes (eudicots_odb10) for *Q. glauca* and *L. edulis*, respectively. The genome sizes of *Q. glauca* and *L. edulis* were estimated to be 891.2 Mb and 868.7 Mb, respectively, using KmerGenie (Chikhi and Medvedev, 2014) and GenomeScope2 (Ranallo-Benavidez et al., 2020).

Overall, the abundance and distribution of protein-coding genes, GC content, and repeat sequences (Fig. 1C, D*;* Table S1) and syntenic structure (Fig. 1B) were similar between *Q. glauca* and *L. edulis*. Using OrthoFinder2 (Emms and Kelly, 2019), we clustered genes from the two newly assembled genomes into orthologous groups. We identified 13427 single-copy orthologs, enabling pairwise comparisons between species (Fig. 1E). Single-copy genes exhibited significantly higher average expression levels than multi-copy genes (Steel-Dwass test, $p < 0.05$; Fig. 1F, G). These single-copy orthologous genes were used for further analyses.

Seasonal gene expression dynamics. To characterize the evolution of seasonal gene expressions, we generated a total of 483 transcriptome profiles every four weeks over two years using leaf and bud tissues from *Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber* during 2021–2023. Reads from the two *Quercus* species were mapped to the newly constructed *Q. glauca* reference genome, achieving a mean mapping rate of 87.0 ± 5.6%. Similarly, reads from the two *Lithocarpus* species were mapped to the newly constructed *L. edulis* reference genome, with a mean mapping rate of 90.0 ± 3.0%. The number of reads uniquely mapped on the reference genome was 19.9 million reads per sample, on average. Among the single copy orthologous genes, we excluded genes with low expression levels (mean of Reads Per Kilobase (RPK) across samples for each species < 1) and normalized the transcript count data using the GeTMM method (Smid et al., 2018). This process resulted in a high-quality expression dataset for 11749 genes (Fig. S1A). For further analyses, we used the mean expression levels across three biological replicates (three individuals per species), except for *L. edulis*, where only one individual was available.

To gain an initial overview of gene expression dynamics in a seasonal environment, we assessed the relative magnitude of seasonal variation in expression levels for each gene by calculating the standard deviation divided by the mean for each gene (σ_i/μ_i) and compared it to the overall variation across all genes in the genome (σ_g/μ_g) (Fig. S1B). A prior study using 3025 yeast RNA-seq experiments reported that the dynamic range of individual genes across experiments was considerably smaller than the overall genomic variation under

153 laboratory conditions (Zrimec et al., 2021). Consistently, our results demonstrate
154 that despite substantial seasonal fluctuations in environmental conditions (Fig. S2),
155 individual gene expression remains relatively conserved. On average, the
156 magnitude of genome-wide variation was 15.8 times greater than the seasonal
157 variation observed for individual genes (Fig. S1B).
158 To assess the similarity of gene expression profiles across tissues, species,
159 and seasons, we performed a hierarchical clustering of seasonal transcriptome
160 profiles from leaf and bud tissues across four species and identified five distinct
161 clusters (cluster L1, L2, B1–B3) according to the Elbow method (Fig. 2A; Fig. S3A;
162 Dataset S1). Transcriptional profiles first grouped into leaves (cluster L1 and L2)
163 and buds (cluster B1–B3), irrespective of seasons and species (Fig. 2A), except
164 that newly flushed leaf samples were grouped closely with buds (Fig. 2A; Fig. S4).
165 Within the leaf cluster, transcriptome profiles were first clustered by genus and
166 then by species in *Lithocarpus* (cluster L1; Fig. 2A). In contrast, samples from
167 *Quercus* species formed a mixture of two species (cluster L2; Fig. 2A). Notably,
168 the winter samples of *Q. acuta*, collected during periods of low mean temperatures,
169 were distinctly separated from those collected in other seasons (cluster L2; Fig.
170 2A). Within the bud cluster, transcriptome profiles were predominantly grouped by
171 season—winter (cluster B3; November to February, with the exception of a March
172 sample from *L. edulis*; Dataset S1) versus the other growing seasons (cluster B1
173 and B2; March to November; Dataset S1)—and subsequently organized by genus
174 and species. This suggests that the winter season exerts a stronger influence on
175 gene expression in buds than phylogenetic relationships.
176 Within the bud cluster, transcriptional profiles in the winter cluster were
177 observed under significantly lower temperatures, approximately below 10°C,
178 compared to other seasons (Fig. 2B; Fig. S5A). In contrast, no such temperature-
179 related distinction was evident within the leaf cluster (Fig. 2B). Additionally,
180 photoperiods were markedly shorter in the winter cluster in buds (Fig. S5B). These
181 findings suggest that the buds of the four species exhibit a conserved response to
182 environmental factors characteristic of the winter season. To further investigate the
183 relationships among all samples, we conducted a principal component analysis
184 (PCA). The first principal component (PC1) clearly separated samples by tissue
185 type, while the second (PC2) and the third principal component (PC3) gradually
186 differentiated winter samples from those collected during other seasons (Fig. S6).
187 To investigate seasonal expression dynamics of each gene, we conducted
188 hierarchical clustering on the transcriptional dynamics of 11749 genes and
189 identified five distinct gene sets characterizing different transcriptional modes
190 (gene set I–V; Fig. 2C, D; Fig. S3B; Dataset S2). Gene set I exhibited predominant
191 expression in buds, with seasonal peaks during summer, whereas gene set II was
192 primarily expressed in leaves, highlighting clear tissue-specificity (Fig. 2C, D).
193 Gene sets III and V demonstrated genus-specific expression patterns, with higher
194 expression levels observed in *Lithocarpus* and *Quercus*, respectively (Fig. 2C, D).

In contrast, gene set IV was expressed during winter across both leaf and bud tissues (Fig. 2C, D).

Detection of rhythmic genes. To identify genes revealing seasonal expression dynamics and determine the period and seasonal peaks in gene expression, we applied the RAIN algorithm (Thaben and Westermark, 2014). This approach enabled us to detect genes with rhythmic expression (hereafter rhythmic genes) and analyze their periods and phases (peak months). Our analysis revealed that 44.6–58.8% of genes in bud tissues and 33.9–45.9% in leaf tissues exhibited significant rhythmic expression across species (Fig. S7A; Dataset S3). Notably, bud tissues demonstrated a higher proportion of rhythmic genes, with 51.9% across four species on average, compared to 40.6% in leaf tissues (Fig. S7A).

Among the rhythmic genes, the majority (78.4–92.0%) exhibited annual periodicity, with a period of 12 ± 1 month observed across species and tissues (Fig. 3A, B, Dataset S3). Additionally, hundreds of genes (1.2–11.9%) displayed half-annual periodicity (6 ± 1 month) with these genes being most frequently observed in the bud samples of *L. edulis* (Fig. 3A, B). The fraction of genes exhibiting other periodicities was relatively small (5.7–14.1%; Fig. 3A, B). Based on the observed period distributions, we classified genes with periods within the range of 8–16 months as exhibiting annual periodicity (96.0%) and those with periods shorter than 8 months as having half-annual periodicity (4.0%), accounting for detection noise (Fig. S7A). Rhythmic genes with longer periods (>17 months) were also identified, although their periodicity may be less reliable due to the limited number of time points. Many genes with annual periodicity were shared across all species, including 1280 genes in leaves and 1894 genes in buds (Fig. S7B). In contrast, most genes exhibiting half-annual periodicity were species-specific, with no genes shared among all the species (Fig. S7C).

Peak month distribution of rhythmic genes and intra-genus and inter-genera comparison. To determine the seasonal timing of gene expression peaks, we focused on genes with annual or half-annual periodicity and visualized the density and direction of these peaks throughout the year using a circular histogram (Fig. 3C, D; Dataset S3). On average, 34.9% and 41.3% of annual rhythmic genes peaked in winter The biological activities of organisms are closely linked to seasonality. Phenology, the temporal orchestration of biological activities, is governed by gene expression, yet the evolutionary dynamics underlying seasonal gene expression remain unclear. To investigate these dynamics, we compared genome-wide expression dynamics (molecular phenology) in four dominant evergreen Fagaceae species in Asia (*Quercus glauca*, *Q. acuta*, *Lithocarpus edulis*, and *L. glaber*), using leaf and bud tissues over two seasonal cycles. We assembled high-quality reference genomes, identifying 11749 single-copy orthologous genes. Seasonal transcriptomic profiling of these orthologous genes revealed highly conserved gene expression across species in winter when temperatures fall below ~10°C. Rhythmic gene expression with significant periodic oscillations was more prevalent in buds (51.9%) than in leaves (40.6%), with most

rhythmic genes (78.4–92.0%) exhibiting annual periodicity, while a smaller fraction (1.2–11.9%) followed half-annual cycles. The seasonal peaks of rhythmic genes were highly synchronized across species in winter but diverged during the growing season, reflecting species-specific timing of leaf flushing and flowering. These findings suggest that the four species share a common molecular calendar in winter, which constrains the evolution of gene expression under seasonal environments.

(December–February) in leaf and bud samples, respectively. Additionally, another peak corresponding to or occurring slightly after leaf flushing and flowering events was observed, with its timing varying by species: April in *Q. glauca* (15.6%), July in *Q. acuta* (15.3%), June in *L. edulis* (16.0%), September in *L. glaber* (18.1%), for leaf tissues (Fig. 3C). Similar peaks corresponding to leaf flushing or flowering were also detected in bud samples, occurring approximately one to two months earlier than in leaf samples (Fig. 3D). For genes with half-annual periodicity, the majority (63.2 ± 4.8%) exhibited expression peaks both in spring (March–May) and autumn (September–November) in both tissues (Fig. S7D; Dataset S3), with the exception of *L. glaber*, an autumn flowering species. Interestingly, *WRKY22*, a transcription factor regulating ethylene biosynthesis and senescence (Zhu et al., 2024), exhibited half-annual periodicity in *L. glaber* (Fig. S7*E*). In addition, *SNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT GAMMA 1* (*KINγ1*), a subunit of SNF1-related kinase (*SnRK*) involved in stress response such as starvation (Emanuelle et al., 2016), showed two distinct peaks in *Q. glauca* and *L. edulis*, occurring in both summer (July) and winter (January) (Fig. S7E). This result suggests that both summer and winter pose stressful conditions for these species.

To compare the peak month of gene expression with annual periodicity across species, we quantified the number of genes within each peak combination for both intra-genus (Fig. 4A) and inter-genera species pairs in buds (Fig. 4B) and leaves (Fig. S8A, B). During winter, a large number of genes exhibited highly synchronized expression peaks across species, with high pairwise Pearson's correlation coefficients observed (Fig. 4A, B). In contrast, the peak timing of gene expression from spring to autumn varied across species, aligning with differences in leaf flushing and flowering times (Fig. 4A, B). Some genes exhibited pronounced divergence, showing negative Pearson's correlation coefficients (< -0.3) between species. Notably, the number of such genes was larger in inter-genera comparisons (11–77 genes; Fig. 4B) than in intra-genus comparisons (4–12 genes; Fig. 4A). A similar pattern was observed in leaf samples (Fig. S8A, B). When comparing seasonal expression peaks between tissues, the majority of genes were synchronously activated in winter in both tissues (Fig. S8C). During spring to summer, slight variations in seasonal expression peaks were observed among tissues (Fig. S8C).

We quantified the degree of divergence in seasonal gene expression patterns across species by calculating the peak month differences between two species, where one species exhibits an expression peak in the given month. We defined the molecular phenology divergence index (*D*) for each month by quantifying the proportion of genes with peak month differences greater than 2

months between two species, relative to the total number of genes with expression peaks in that month. From spring to autumn, the mean of *D* over all species pairs in buds ranged from 0.47 to 0.66, but it declined sharply during winter, reaching a minimum of 0.12 in January in buds (Nemenyi test, $p < 0.05$; Fig. 4C). A similar trend was observed in leaf samples, with another trough in summer (Fig. S8D). The values of *D* were not consistently smaller in inter-genera pairs compared to intra-genus pairs. These findings suggest that there is a seasonal constraint in gene expression in which the degree of molecular phenology divergence is not constant but varies significantly across seasons, with the least divergence occurring in winter.

Phylogenetic constraints in the evolution of seasonal gene expression. We next investigated whether phylogenetic relationship influences the evolution of seasonal gene expression by calculating pairwise Pearson's correlations of seasonal expression for each gene across all species pairs (Dataset S4). The median of Pearson's correlation coefficients across all genes were significantly higher between species within the same genus compared to those between species from different genera in both leaves (Steel-Dwass test, $p < 0.05$; Fig. S9A) and buds (Steel-Dwass test, $p < 0.05$; Fig. 5A). The intra-genus comparison in genus *Lithocarpus* exhibited the highest correlation in their seasonal gene expression, despite their notable differences in flowering phenology, with *L. edulis* flowering in spring and *L. glaber* in autumn. In contrast, the intra-genus comparison in genus *Quercus* resulted in a relatively low correlation in their gene expression regardless of relatively similar phenology in leaf flushing and flowering. This difference may be attributed to environmental factors, as *Q. acuta* is the only species found in the high-altitude study site with colder climates compared to the other three species.

Rhythmic genes exhibited more conserved expression patterns than arrhythmic genes in both tissues (Fig. 5*A*; Fig. S9A). In particular, genes with seasonal expression peaks in winter showed high correlation across species in both tissues (Fig. 5B; Fig. S9B). The proportion of rhythmic genes with annual periodicity declined as the mean correlation coefficient across species decreased, with a more rapid decline observed for genes peaking in winter compared to those with peak expression during the growing seasons (Fig. 5B; Fig. S9B). Additionally, the mean absolute difference in seasonal gene expression dynamics was significantly smaller between species within the same genus than between species from different genera in both tissues (Fig. S10A, B). The relationship between the mean absolute difference and the proportion of rhythmic genes was less clear compared to that of the mean correlation coefficient (Fig. S10C, D). These findings suggest that, in addition to seasonal constraints, the overall dynamics of seasonal gene expression are also shaped by phylogenetic constraints.

Functional annotation of genes with conserved or divergent expression across species. To functionally annotate genes with conserved or divergent expression across species, we extracted genes with the top and bottom 5% based

330 on the distribution of mean correlation (Fig. 5B; Fig. S9B) and conducted Gene
331 Ontology (GO) enrichment analysis. GO analysis revealed that the top 5% of genes
332 in leaves were associated with photosynthesis, while those in buds were linked to
333 cell fate determination (Fig. S11A; Dataset S5). The gene with the highest
334 correlation in leaves was *STRESS RESPONSE SUPPRESSOR1* (*STRS1*; mean
335 of pairwise correlation = 0.91; Fig. 5C), a DEAD-box RNA helicase previously
336 reported to enhance tolerance to abiotic stresses such as cold (Kant et al., 2007).
337 In buds, the most highly correlated gene was the histone deacetylase *HDA19*
338 (mean of pairwise correlation = 0.93; Fig. 5C), which plays a key role in
339 environmental stress responses through histone modification and chromatin
340 remodeling (Chen and Wu, 2010). These findings suggest that genes associated
341 with photosynthesis during the growing season and stress responses during winter
342 exhibit highly conserved seasonal expression patterns across species.
343 The bottom 5% of genes were predominantly associated with DNA
344 replication and cell division in leaves (Fig. S11B), aligning with the divergent leaf
345 phenology observed (Fig. 3B). In contrast, GO terms related to immune response
346 were weakly enriched in buds (Fig. S11B). Notably, we identified MADS-box genes
347 linked to floral transition among the bottom 5% of genes: *BROTHER OF FT AND*
348 *TFL1* (*BFT*) in leaves (mean of pairwise correlation = -0.094) and *FRUITFULL*
349 (*FUL*) in buds (mean of pairwise correlation = -0.069). *BFT*, a floral repressor (Yoo
350 et al., 2010), was expressed in autumn in the spring-flowering species *L. edulis*
351 and in spring in the autumn-flowering species *L. glaber* leaves (Fig. 5D).
352 Meanwhile, *FRUITFULL* (*FUL*), a positive regulator of floral transition (Smaczniak
353 et al., 2012), was expressed in spring in *L. edulis* and in autumn in *L. glaber* (Fig.
354 5D), mirroring the timing of flowering. In *Q. glauca* and *Q. acuta*, *FUL* was
355 expressed in winter (Fig. 5D), nearly out of phase with their expressions in
356 *Lithocarpus* species. These results suggest that genes involved in reproductive
357 processes exhibit species-specific divergence in expression patterns in response
358 to seasonal environmental changes.
359
360 **Relating seasonal gene expression divergence to sequence divergence.** If
361 divergence of seasonal timing of gene expression is related to selective constraint,
362 we would expect a correlation between sequence divergence patterns and the
363 divergence in seasonal expression peaks. To explore the relationship between the
364 divergence of seasonal gene expression and protein sequence evolution, we
365 calculated synonymous (dS) and non-synonymous substitution rates (dN) and their
366 ratio (dN/dS) for each gene using coding sequences from newly assembled
367 genomes of *Q. glauca* and *L. edulis*. We first confirmed that the highly expressed
368 genes evolve more slowly by identifying a significant negative correlation between
369 evolutionary rate (dN/dS) and mean of the gene expression (r = -0.21, p = 1.8 ×
370 10^{-118} for dN/dS; Fig. S12A; Dataset S6). This finding aligns with previous studies
371 in yeasts, mammals, and plants (Drummond et al., 2005; Gaut et al., 2011; Liao
372 and Zhang, 2006; Yang and Gaut, 2011). By targeting the genes that exhibited
373 annual periodicity across both species and tissues, we examined the relationship
374 between these evolutionary measures and peak month differences in seasonal

gene expression between species ($\Delta\varphi$) in leaves and bud tissues using linear regression. Assessment of the significance of the regression coefficient using a *t*-test revealed that genes with smaller peak month differences between species exhibited lower values of dN/dS in buds (r = 0.072, p =2.6 × 10^{-2}) and marginally lower values of dN/dS in leaves (r = 0.062, p = 5.7 × 10^{-2}). The lower values were also found in dN in both leaves (r = 0.086, p = 8.1 × 10^{-3}) and buds (r = 0.086, p = 7.8 × 10^{-3}), whereas no significant trend was observed for dS in both tissues (Fig. 5E; Fig. S12B, C; Dataset S6). This finding suggests that genes with conserved seasonal expression patterns tend to evolve at a slower rate. We further investigated whether the association between the divergence in seasonal expression peaks and sequence divergence patterns varies between genes expressed in winter and in the growing season. We classified genes that exhibited annual periodicity across both species and tissues into four groups based on their expression peak timing (winter or growing seasons) and tissue specificity (leaves or buds) and compared the extent of protein sequence divergence across these groups. We found no significant differences in dN/dS (Kruskal Wallis test, p = 2.8 × 10^{-1}; Fig. S12*D*; Dataset S6) between genes expressed in winter in both tissues and those in growing seasons. The same result was observed even when focusing only on genes with conserved peaks (peak month difference < 2), with no significant differences in dN/dS between these gene sets (Kruskal Wallis test, p = 3.5 × 10^{-1}; Fig. S12E; Dataset S6). These results suggest that, regardless of the season, genes with conserved expression peaks tend to evolve at slower rates.

Discussion

A central goal in biology is to unravel the molecular basis of phenotypic evolution, with gene expression serving as a fundamental driver of this process. In this study, we investigated the impact of seasonal environmental fluctuations on the evolution of gene expression and its relationship with phenological traits in four dominant *Fagaceae* tree species in forest ecosystems. We found that gene expression profiles were more resistant to change during winter compared to the growing season, indicating the presence of a seasonal constraint on the evolution of gene expression. This constraint becomes particularly pronounced in buds during winter when temperatures fall below approximately 10°C (Fig. 2B). This temperature threshold aligns with previous findings, where winter-specific transcript profiles were reported at temperatures below 8.7–11.3°C in cherry trees (Miyawaki - Kuwakado et al., 2024), 11.3°C in ever-green *Fagaceae* (Satake et al., 2023), and 10.5°C in perennial *Arabidopsis* (Aikawa et al., 2010). These results suggest the presence of a conserved mechanism for gene expression regulation under low-temperature environments across different species. The greater conservation of gene expression patterns in buds compared to leaves across species may be attributed to their functional role protecting essential meristematic tissues, including the shoot apex and leaf and flower primordia. Additionally, buds likely play a crucial role in maintaining common physiological functions necessary for the resumption of growth in spring. This functional consistency may impose

evolutionary constraints on gene expression, contributing to its conservation across species.

It is intriguing to compare the seasonal constraints identified in this study with developmental constraints in gene expression evolution, where evolutionary conservation and variation across developmental stages during the life cycle have been observed, such as in the early conservation model and the developmental hourglass model (Fig. S13A)(Irie and Kuratani, 2011; Lotharukpong et al., 2024; Quint et al., 2012). Previous studies have demonstrated that certain developmental stages exhibit remarkably similar gene expression patterns (Drost et al., 2016; Kalinka et al., 2010; Levin et al., 2016). Our findings highlight that such highly similar gene expression patterns are observed during the winter season (Fig. S13B), suggesting that a common physiological state is essential for enduring harsh winter conditions. This seasonal constraint on gene expression evolution may have the potential to limit temporal niche partitioning at the phenotypic level and slow species divergence rates in seasonal environments (Cleland and Wolkovich, 2024; Kubo and Iwasa, 1996). Although completely disentangling the effects of developmental stages and seasonal changes on gene expression is challenging when using molecular phenology data collected from natural environments, the observation that species with different phenology—i.e., different developmental timing in seasonal environments—exhibit convergent gene expression patterns during winter underscores the profound influence of winter environmental conditions on the evolution of gene expression. Future research is needed to elucidate the mechanisms driving the establishment of such conserved gene expression patterns, including the dynamic reorganization of chromatin architecture in response to cold (Fischl et al., 2020; Nishio et al., 2020; Zhang et al., 2023).

We found that the evolution of coding-sequences was weakly correlated with the divergence of seasonal peaks in gene expression across species. This suggests that weak functional constraints exist for specific proteins that are expressed in certain seasons, such as photosynthesis in summer and stress response to winter. However, among genes with conserved seasonal peaks, the evolutionary rate of coding-sequences was similarly slow regardless of whether they were expressed in winter or during the growing season. Therefore, to understand the evolutionary mechanism underlying the seasonal constraint on gene expression that results in similar expression patterns during winter, it is essential to consider evolution of non-coding sequences that modulate gene expression levels in response to seasonal environmental changes. Regulatory variation has been recognized as a major driver of evolutionary innovation (Prud'homme et al., 2007; True and Haag, 2001). Recent deep learning studies utilizing genome and transcriptome data from yeast and humans have highlighted the significance of co-evolution between coding and non-coding sequences in shaping gene expression levels (Zrimec et al., 2021, 2020). Understanding how seasonal environments drive the co-evolution of coding and regulatory sequences, as well as the divergence of molecular phenology and its phenological outputs, will require expanding analyses to a broader range of species. Such efforts will

464 facilitate the development of predictive models for evolutionary processes
465 influenced by seasonal dynamics. Comparison of molecular phenology across
466 wide climatic zones, including both tropical and temperate zones, will also be
467 effective to identify key genes for the phenological response to on-going climate
468 change.
469

470 Materials and Methods

471 Sample collection for genome sequencing

472 To construct genomes of *Q. glauca* and *L. edulis*, we collected approximately 5g
473 of fresh leaves from one individual in spring, respectively. We shipped these
474 samples to Dovetail Genomics, LLC (Scotts Valley, CA, USA) where Dovetail Staff
475 performed DNA extraction, library preparation, sequencing, and assembly steps.
476

477 Genome sequencing and assembly

478 DNA extraction was performed using the QIAamp DNA Mini Kit (Qiagen,
479 Germantown, MD, USA) using the manufacturer's recommended protocol. Mean
480 fragment length of the extracted DNA was 60 kbp for *Q. glauca* and 65 kbp for *L.*
481 *edulis*. DNA samples were quantified using Qubit 2.0 Fluorometer (Life
482 Technologies, Carlsbad, CA, USA). The library preparation, sequencing and
483 scaffolding were carried out by Dovetail Genomics (California, USA) according to
484 their standard genome assembly pipeline (https://dovetailgenomics.com/). The
485 PacBio SMRTbell library (~20kb) for PacBio Sequel was constructed using
486 SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) using
487 the manufacturer's recommended protocol. The library was bound to polymerase
488 using the Sequel II Binding Kit 2.0 (PacBio) and loaded onto PacBio Sequel II.
489 Sequencing was performed on PacBio Sequel II 8M SMRT cells. PacBio CCS
490 reads were used as an input to Hifiasm1 v0.15.4-r347 with default parameters. For
491 *Q. glauca*, we also generated sequence data using PacBio Revio. BLASTn results
492 of the contigs obtained by Hifiasm assembly against the NCBI's NT database were
493 used as input for blobtools2 v1.1.1 and contigs identified as possible contamination
494 were removed from the assembly. Finally, purge_dups3 v1.2.5 was used to
495 remove haplotigs and contig overlaps.
496

497 Assembly scaffolding with HiRise

498 A proximity ligation library was generated by the Dovetail's Omni-C technique,
499 followed by sequencing on an Illumina HiSeqX platform. Chromatin was fixed in
500 place in the nucleus with formaldehyde before extraction
501 (https://dovetailgenomics.com/wp-content/uploads/2021/09/Omni-C-Tech-
502 Note.pdf). Fixed chromatin was digested with DNase I, fragmented chromatin ends
503 were repaired and biotinylated to adapters followed by proximity ligation.
504 Crosslinks were then reversed, the DNA purified, and the biotin subsequently
505 removed. The DNA library was prepared and sequenced to produce 2 × 150 bp
506 paired-end reads (around 30× coverage). The Omni-C technology uses a
507 sequence-independent endonuclease which provides even, unbiased genome
508 coverage. The input *de novo* assembly and Omni-C library reads were used as

input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al., 2016). The Omni-C library sequences were aligned to the contigs assembled by Hifiasm using bwa (Li and Durbin, 2009). The separations of Omni-C reads from paired-end mapped within the contigs were analyzed by HiRise to produce a likelihood model for genomic distance between paired-end reads, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

Genome size estimation

The genome size of *Q. glauca* was estimated by GenomeScope2 v2.0 (Ranallo-Benavidez et al., 2020) using *k*-mer frequency plot obtained by kmc v3.2.4 (Kokot et al., 2017) with *k*-mer size 61 against the HiFi reads. The paired-end reads of four individuals of *L. edulis* were merged and trimmed by fastp v0.23.2 (Chen et al., 2018). The genome size of *L. edulis* was estimated by KmerGenie (Chikhi and Medvedev, 2014) using *k*-mer size 55 against the trimmed paired-end reads.

Sample collection for RNA sequencing

To monitor the genome-wide gene expression profiles, we collected leaf, bud, and flower samples from four Fagaceae species, *Q. glauca*, *Q. acuta*, *L. edulis* and *L. glaber*, every four weeks spanning from May 2019 to June 2023 across Fukuoka and Saga prefectures in Kyushu, Japan. 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. glauca* is the biodiversity reserve on the Ito campus of Kyushu University (33º35′ N, 130º12′ E, 20 to 57 m a.s.l). *Q. acuta* was monitored in Sefuri mountains (33º26′ N, 130º22′ E, 970 to 974 m a.s.l). *L. edulis* and *L. glaber* were studied at the Imajuku Field Activity Center (33º33′ N, 130º16′ E, 84 to 111 m a.s.l). Environmental data from Ito Campus and the Imajuku Field Activity Center were obtained from the Japan Meteorological Agency, while data from Sefuri Mountain were collected using a logger installed at the sampling site.

Samples for RNA-seq were taken from the sun-exposed crown (approximately 2–5 m from the ground) using long pruning shears from 11:30 to 12:30 h. For each sample, 0.2–0.4 g of tissue was preserved in a 2 ml microtube containing 1.5 ml of RNA-stabilizing reagent (RNAlater; Thermo Fisher Scientific, Waltham, MA, USA) immediately after harvesting. The samples were transferred to the laboratory within 2 hr after sampling, stored at 4°C overnight and then stored at −80°C until RNA extraction. During transport to the laboratory, the samples were kept in a cooler box with ice to maintain a low temperature.

RNA extraction and RNA-seq analysis

RNA was extracted independently from bud and leaf samples of each tree. Total RNA of the leaf sample was extracted in accordance with the method described in a previous study (Miyazaki et al., 2014). Total RNA of the bud sample was

554 extracted using the PureLink™ Plant RNA Reagent (Thermo Fisher Scientific,
555 Waltham, MA, USA) and subsequently processed with DNase using the TURBO
556 DNA-free kit™ (Thermo Fisher Scientific) according to their protocols and purified
557 using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to its RNA
558 cleanup protocol. RNA integrity was examined using the Agilent RNA 6000 Nano
559 kit on a 2100 Bioanalyzer (Agilent Technologies), and the RNA yield was
560 determined using a NanoDrop ND-2000 spectrophotometer (Thermo Fisher
561 Scientific). Five to six micrograms of RNA extracted from each sample were sent
562 to Hangzhou Veritas Genetics Medical Institute Co., Ltd., where a cDNA library
563 was prepared with a NEBNext Ultra II RNA Library Prep Kit for Illumina, and 150
564 base paired-end transcriptome sequencing of each sample was conducted using
565 an Illumina NovaSeq6000 sequencer (Illumina).
566 The quantification of gene expression followed a three-step process. First,
567 quality filtering was performed using the fastp v0.23.2 with default options: phred
568 quality ≥15, the maximum number of N base in one read = 5 (Chen et al., 2018).
569 Second, RNA-seq reads were mapped to the reference genome using STAR
570 v2.7.10b (Dobin et al., 2013). Third, RNA-seq reads were quantified using RSEM
571 v1.3.1 (Li and Dewey, 2011). We excluded the genes with low expression level
572 using the threshold, the value of mean Reads Per Kilobase (RPK) < 1 for each
573 species. Then, we further normalized the expression level using the value of Gene
574 length corrected Trimmed Mean of M-values (GeTMM) (Smid et al., 2018) using
575 an R package "edgeR" v3.42.4. For the downstream analysis we calculated the
576 mean expression levels of GeTMM of each gene across multiple individuals and
577 transformed them to $\log_2$(GeTMM+1).
578
579 **Genome annotation**
580 The RNA-seq reads of *Q. glauca* (8 samples obtained by a previous study (20)
581 collected from May to December 2017; GEO's accession number: GSE211384)
582 and *L. edulis* (20 samples covering leaf, bud and flower tissues collected in
583 January, April, July, and November 2021) were respectively mapped against the
584 scaffolds of *Q. glauca* and *L. edulis* obtained by HiRise with HISAT2 v2.1.0 (Kim
585 et al., 2019). The bam files generated by the mappings and amino acid sequences
586 of the 9 species obtained from NCBI (*Castanea mollissima*: GCA_000763605.2
587 (NCBI's accession number), *Q. lobata*: GCF_001633185.1, *Q. robur*:
588 GCF_932294415.1, *Q. suber*: GCF_002906115.1, *Carpinus fangiana*:
589 GCA_006937295.1, *Carya illinoinensis*: GCF_018687715.1, *Juglans macrocarpa*
590 × *J. regia*: GCF_004785595.1, *Juglans regia*: GCF_001411555.2, *Morella rubra*:
591 GCA_003952965.2) were used for ab initio gene prediction by Braker2 v2.1.6
592 (Brůna et al., 2021), and the genes with highest-score were selected to exclude
593 splicing variants. In parallel, the gene models were respectively constructed in *Q.*
594 *glauca* and *L. edulis* by GeMoMa v1.7.1 (Keilwagen et al., 2019) with RNA-seq
595 reads and amino acid sequences used for the gene prediction in Braker2, and the
596 genes with highest-score were selected to exclude splicing variants. The
597 duplicated genes between Braker2 and GeMoMa were excluded by GFFcompare
598 v0.11.2 (Pertea and Pertea, 2020). The remaining genes were searched against

UniProtKB (Bateman et al., 2025) by DIAMOND v2.0.5 (Buchfink et al., 2014) in 'more-sensitive' option with E-value ≤10^{-20} and identity ≥25%. Next, gene expression analysis was performed by Salmon v1.4.0 (Patro et al., 2017), and the genes with TPM values >0.1 were considered as expressed. The orthologous genes were searched against EggNOG (Hernández-Plaza et al., 2023) using eggNOG-mapper v2.1.8 (Cantalapiedra et al., 2021) with E-value ≤10^{-10}. The genes were searched against protein sequences of *Castanea mollissima* (Accession: GCA_000763605.2), *Quercus lobata* (GCF_001633185.2), and *Arabidopsis thaliana* (Araport11 (Krishnakumar et al., 2015)) by BLASTp v2.13.0 (Camacho et al., 2009) with E-value ≤10^{-20}. According to these similarity searches, the genes whose product names contain keywords related to transposable elements (TEs) were categorized into TE. The genes having similarity hits and being expressed by Salmon were classified into high-confidence (HC) genes, while remaining genes were categorized into low-confidence (LC) genes. The genes categorized into HC were used for further analyses.

Genome synteny analysis

To investigate the syntenic relationship between *Q. glauca* and *L. edulis*, we first performed an all-vs-all BLASTp search of the protein sequences, using a cut-off E-value of 10^{-5}. Collinearity blocks within and between the genomes were then identified using MCScanX (Wang et al., 2012) based on the results of the BLASTp searches. Finally, the collinearity blocks were visualized using SynVisio (https://synvisio.github.io).

Illumina short-read sequencing and library preparation

To draw molecular phylogeny, we also generated short-read sequencing of *Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber.* DNA extraction was performed using a modified version of the method described previously (Satake et al., 2024b). The DNA samples were sent to Macrogen Inc. (Republic of Korea) for sequencing on the Illumina HiseqX platform (Illumina, San Diego, CA, USA). DNA was sheared to around 500 bp fragments in size using dsDNA fragmentase (New England BioLabs, Ipswich, MA, USA). Library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit (New England BioLabs) according to the manufacturer's protocol, and the libraries were individually indexed with the NEBNext Multiplex Oligos for Illumina (New England BioLabs) by PCR. The quality and quantity of each amplified library were analyzed using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), the High Sensitivity DNA kit (Agilent Technologies), and the NEBNext Library Quant Kit for Illumina (New England BioLabs).

Phylogenetic analysis

We constructed a phylogenetic tree using an alignment-free method (Fan et al., 2015) based on short-read data from *Q. acuta* and *L. glaber* and genome assembly data from *Q. glauca* and *L. edulis*. Additionally, genomic data for *Fagus sylvatica* was obtained from beechgenome.net (Mishra et al., 2022) and included as an

644 outgroup. Due to the large volume of short-read data for *Q. acuta* and *L. glaber*
645 (18 Gb each), we randomly subsampled half of the reads. Bootstrap values were
646 estimated through 100 resampling of the k-mer table.
647
648 **Identification of orthologs**
649 The prediction of orthogroups was based on a BLASTp all-against-all comparison
650 of the protein sequences (E-value < 10^{-5}) of 5 Fagaceae species: *L. edulis*, *Q.*
651 *glauca*, *Q. robur*, *Fagus sylvatica* and *Juglans regia*. Clustering was conducted
652 with OrthoFinder2, using version 2.2.6 for BLAST data generation and version
653 2.5.5 for orthogroup determination (Emms and Kelly, 2019). The genomic data of
654 *Q. robur* and *J. regia* were downloaded from NCBI (dhQueRobu3.1 and Walnut
655 2.0), while *F. sylvatica* data were sourced from beechgenome.net (Mishra et al.,
656 2022). Since the *F. sylvatica* dataset contained numerous predicted transposable
657 elements (TEs), we re-annotated genes using three databases—EggNOG5
658 (EggNOG-mapper), Pfam35.0 (HMMER v3.3.2), and UniRef90_2023_0628
659 (DIAMOND v2.1.8.162)—and excluded genes identified as TEs in at least one
660 database. In total, we identified 26692 orthogroups and focused on single-copy
661 orthogroups after filtering out genes mapped outside the 12 chromosomes. This
662 resulted in a final dataset of 11749 genes for further analysis (Data S2).
663
664 **Hierarchical clustering and principal component analysis (PCA)**
665 To assess the similarity of the genome-wide transcriptional profiles across
666 orthologous genes and 483 samples collected in different seasons, we performed
667 hierarchical clustering using the monthly time series data of 11749 orthologous
668 genes from October 2019 to April 2023. For each orthologous gene, there were 26
669 or 27 time points, with three individuals each for *Q. glauca*, *Q. acuta*, *L. glaber* and
670 one individual for *L. edulis*. We calculated the mean expression levels of each
671 orthologous gene across three individuals in each species and subsequently
672 normalized the values by adjusting the mean to zero and the standard deviation to
673 one. We performed hierarchical clustering using the Ward method and the
674 Euclidian distance using the pheatmap function in R (v 1.0.12). To determine the
675 number of the clusters, we calculated Within-Cluster-Sum of Squared Error (WSS)
676 and determined the number of clusters as an elbow point of it (Fig. S3). To assess
677 the seasonal expression dynamics of 11749 orthologous genes, we also
678 performed PCA of the gene expression profiles from all samples using the prcomp
679 function in R (v 4.3.1).
680
681 **Identification of genes with rhythmic expression**
682 To identify genes with rhythmic expression and determine the period and peak
683 month of gene expression, we applied the RAIN algorithm (Thaben and
684 Westermark, 2014) using the rain function in R (v 1.34.0). We set the following
685 settings: deltat = 1, period = 12, period.delta = 11, peak.border = (0,1), nr.series =
686 1, method= "independent", na.rm= T , adjp.method = "ABH". Based on the results
687 of the RAIN analysis, we identified genes with significant periodicity using a
688 Benjamini-Hochberg (BH) adjusted *q*-value threshold of < 0.01. Since the period

and peak month in the RAIN output are expressed in terms of the number of time points, we converted them into months using the sampling span (four weeks) and sampling dates. Subsequently, we categorized the rhythmic genes into three categories based on the period length T (in months): "half-annual" ($0 \leqq T < 8$), "annual" ($8 \leqq T \leqq 16$), and "long" ($T > 16$).

Calculation of Pearson's correlation coefficient across species

To quantify the similarity of seasonal gene expression pattern across species, we calculated pairwise Pearson's correlations of seasonal expression for each gene across all the species pairs. For species within the same genus, the monitoring period was identical, allowing direct comparison. On the other hand, the monitoring periods differed between *Quercus* and *Lithocarpus* species: *Q. glauca* and *Q. acuta* were monitored from May 2021 to April 2023, while *L. edulis* and *L. glaber* were monitored from October 2019 to October 2021 (Fig. S2). To ensure comparable time series data for inter-genera comparisons, we extracted the time series from *Q. glauca* and *Q. acuta* spanning May 19, 2021, to November 2, 2022, and from *L. edulis* and *L. glaber* spanning May 13, 2020, to October 20, 2021, for correlation analysis. The mean absolute differences for each gene across all the species pairs were also calculated in a similar manner.

Molecular phenology divergence index (*D*)

To quantify the seasonal difference of peak month divergence across species, we defined the molecular phenology divergence index (D) for each month t and species pair A-B as the proportion of genes that peak in month t in the reference species A but exhibit a peak month difference ($\Delta\varphi$) greater than two months in species B using following formula:

$$D := \frac{\#\ \text{Genes with peak month } t \text{ in species A and } t \pm \Delta\phi(\geq 2) \text{ in species B}}{\#\ \text{Genes with peak month } t \text{ in species A}}.$$

Gene Ontology (GO) enrichment analysis

First, we assigned GO terms for each gene in *Lithocarpus* and *Quercus* genomes based on the annotation data of *Arabidopsis thaliana* orthologs downloaded from PLAZA Database (http://bioinformatics.psb.ugent.be/plaza/). Then, GO enrichment analysis was conducted using the GOstats (Falcon and Gentleman, 2007) function in R (v 2.66.0) with the following settings: ontology = BP, pvalueCutoff = 0.05, conditional = TRUE, and testDirection = over. The universe gene set was defined as genes associated with GO terms common to both of the genomes. Genes without annotated GO terms were removed from the universe set, and those with differing GO term annotations between the two genomes were also excluded. The BH adjusted q-values were calculated from the p-values generated by GOstats to account for multiple testing.

Phenology measurement: leaf unfolding and flowering

To compare seasonal changes in genome-wide gene expression profiles with phenology, we conducted observations of leaf unfolding and flowering phenology on the three terminal branches of each of the three individuals used for gene expression monitoring. Leaf unfolding was defined as the stage where both the entire leaf blade and the leaf stalk were visible (Fig. S4). Flowering was defined as the stage where either male or female flowers were visible (Fig. S4). We subsequently plotted the proportion of branches exhibiting leaf unfolding and flowering. The data are provided in Dataset S2.

Estimation of protein-coding sequence divergence

Pairwise distances for non-synonymous (dN) and synonymous (dS) substitutions, as well as the dN/dS ratio, were estimated for each orthologous gene pair between *L. edulis* and *Q. glauca* using codeml (PAML 4.0)(Yang, 2007). The analysis was performed with the following settings: seqtype = 1, CodonFreq = 2, Runmode = -2, and the transition/transversion ratio estimated from the data. To ensure reliability, genes exhibiting signs of saturated divergence or underestimated dS values were excluded, as codeml produces robust results only within a moderate range of sequence divergence. Specifically, we removed 43 orthologs with dN/dS $\geqq$ 99 and 483 genes with dS > 10, dS < 0.001, or dN < 0.0003. The thresholds of dN and dS were determined by examining the relationship between dN and dS and excluding outliers that deviated from the main distribution. After filtering, the final dataset comprised 11223 orthologous genes.

Linear regression analyses were performed to examine the relationship between dN, dS, dN/dS and peak month differences ($\Delta\varphi$) in gene expression between *Q. glauca* and *L. edulis*, as well as mean expression levels. These analyses were conducted using the lm function in R software (v 4.3.1). Specifically, we modeled the evolutionary rates (dN, dS, and dN/dS) as the response variable and the peak month difference ($\Delta\varphi$) or the mean levels in gene expression as the predictor variable. To account for the skewed distributions of evolutionary rates, logarithm transformations with base 2 were applied.

Data deposition and code availability

The genome sequences of Q. glauca and L. edulis determined in this paper are available from the National Center for Biotechnology Information (NCBI). The sequence reads used for genome assembly and expression analyses are available from BioProject PRJNA1242478. All codes used in the analysis are available at https://github.com/ku-biomath/Evolution_of_gene_expression_in_seasonal_envrionment.

Acknowledgments

We thank Atsuko Miyawaki-Kuwakado for her valuable technical advice on RNA-seq analysis. We also appreciate the technical support provided by Kayoko Ohta. Additionally, we thank Koharu Yamaguchi, Yuta Aoyagi, and Ryosuke Imai for their help with sample collection.

Additional information

Funding

This work was supported by JSPS KAKENHI Grant Numbers 23H04965 and 23H04966.

Author Contributions

S.N.K. and A.S. designed research; S.N.K., Y.I., J.K., H.H., S.I., and A.S. performed research; S.N.K., Y.I., J.K., H.H., S.I., and A.S. analyzed data; S.N.K. and A.S. wrote the paper with inputs from all the authors.

Competing Interest Statement

The authors declare no conflict of interest.

References

Aikawa S, Kobayashi MJ, Satake A, Shimizu KK, Kudoh H. 2010. Robust control of the seasonal expression of the *Arabidopsis* FLC gene in a fluctuating environment. *Proc. Natl. Acad. Sci. U. S. A.* **107**:11632–11637. doi:10.1073/pnas.0914293107

Bateman A, Martin M-J, Orchard S, Magrane M, Adesina A, Ahmad S, Bowler-Barnett EH, Bye-A-Jee H, Carpentier D, Denny P, Fan J, Garmiri P, Gonzales LJ da C, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Joshi V, Jyothi D, Kandasaamy S, Lock A, Luciani A, Luo J, Lussi Y, Marin JSM, Raposo P, Rice DL, Santos R, Speretta E, Stephenson J, Totoo P, Tyagi N, Urakova N, Vasudev P, Warner K, Wijerathne S, Yu CW-H, Zaru R, Bridge AJ, Aimo L, Argoud-Puy G, Auchincloss AH, Axelsen KB, Bansal P, Baratin D, Batista Neto TM, Blatter M-C, Bolleman JT, Boutet E, Breuza L, Gil BC, Casals-Casas C, Echioukh KC, Coudert E, Cuche B, de Castro E, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gaudet P, Gehant S, Gerritsen V, Gos A, Gruaz N, Hulo C, Hyka-Nouspikel N, Jungo F, Kerhornou A, Mercier P Le, Lieberherr D, Masson P, Morgat A, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Poux S, Pozzato M, Pruess M, Redaschi N, Rivoire C, Sigrist CJA, Sonesson K, Sundaram S, Sveshnikova A, Wu CH, Arighi CN, Chen C, Chen Y, Huang H, Laiho K, Lehvaslaiho M, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Y, Zhang J. 2025. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **53**:D609–D617. doi:10.1093/nar/gkae1010

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**:343–348. doi:10.1038/nature10532

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**:1–11. doi:10.1093/nargab/lqaa108

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*. doi:10.1038/nmeth.3176

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* **10**. doi:10.1186/1471-2105-10-421

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**:5825–5829. doi:10.1093/molbev/msab293

Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascenção K, Rummel C, Ovchinnikova S, Mazin P V., Xenarios I, Harshman K, Mort M, Cooper DN, Sandi C, Soares MJ, Ferreira PG, Afonso S, Carneiro M, Turner JMA, VandeBerg JL, Fallahshahroudi A, Jensen P, Behr R, Lisgo S, Lindsay S, Khaitovich P, Huber W, Baker J, Anders S, Zhang YE, Kaessmann H. 2019. Gene expression across mammalian organ development. *Nature* **571**:505–509. doi:10.1038/s41586-019-1338-5

Chen LT, Wu K. 2010. Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response. *Plant Signal. Behav.* **5**:1318–1320. doi:10.4161/psb.5.10.13168

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890. doi:10.1093/bioinformatics/bty560

Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**:31–37. doi:10.1093/bioinformatics/btt310

Cleland EE, Wolkovich EM. 2024. Effects of phenology on plant community assembly and structure. *Annu. Rev. Ecol. Evol. Syst.* **55**:471–492. doi:10.1146/annurev-ecolsys-102722-011653

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635

Drost HG, Bellstädt J, Ó'Maoiléidigh DS, Silva AT, Gabel A, Weinholdt C, Ryan PT, Dekkers BJW, Bentsink L, Hilhorst HWM, Ligterink W, Wellmer F, Grosse I, Quint M. 2016. Post-embryonic hourglass patterns mark ontogenetic transitions in plant development. *Mol. Biol. Evol.* **33**:1158–1163. doi:10.1093/molbev/msw039

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci.* **102**:14338–14343. doi:10.1073/pnas.0504070102

El Taher A, Böhne A, Boileau N, Ronco F, Indermaur A, Widmer L, Salzburger W. 2021. Gene expression dynamics during rapid organismal diversification in African cichlid fishes. *Nat. Ecol. Evol.* **5**:243–250. doi:10.1038/s41559-020-01354-3

Emanuelle S, Doblin MS, Stapleton DI, Bacic A, Gooley PR. 2016. Molecular insights into the enigmatic metabolic regulator, SnRK1. *Trends Plant Sci.* **21**:341–353. doi:10.1016/j.tplants.2015.11.001

858 Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative
859 genomics. *Genome Biol.* **20**. doi:10.1186/s13059-019-1832-y

860 Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association.
861 *Bioinformatics* **23**:257–258. doi:10.1093/bioinformatics/btl567

862 Fan H, Ives AR, Surget-Groba Y, Cannon CH. 2015. An assembly and alignment-free
863 method of phylogeny reconstruction from next-generation sequencing data. *BMC*
864 *Genomics* **16**. doi:10.1186/s12864-015-1647-5

865 Fischl H, McManus D, Oldenkamp R, Schermelleh L, Mellor J, Jagannath A, Furger A.
866 2020. Cold-induced chromatin compaction and nuclear retention of clock mRNAs
867 resets the circadian rhythm. *EMBO J.* **39**. doi:10.15252/embj.2020105604

868 Forrest J, Miller-Rushing AJ. 2010. Toward a synthetic understanding of the role of
869 phenology in ecology and evolution. *Philos. Trans. R. Soc. B* **365**:3101–3112.
870 doi:10.1098/rstb.2010.0145

871 Fretwell SD. 1972. Populations in a seasonal environment. *Monogr. Popul. Biol.* **5**:1–217.

872 Gaut B, Yang L, Takuno S, Eguiarte LE. 2011. The patterns and causes of variation in
873 plant nucleotide substitution rates. *Annu. Rev. Ecol. Evol. Syst.* **42**:245–266.
874 doi:10.1146/annurev-ecolsys-102710-145119

875 Hernández-Plaza A, Szklarczyk D, Botas J, Cantalapiedra CP, Giner-Lamia J, Mende DR,
876 Kirsch R, Rattei T, Letunic I, Jensen LJ, Bork P, von Mering C, Huerta-Cepas J. 2023.
877 eggNOG 6.0: enabling comparative genomics across 12 535 organisms. *Nucleic*
878 *Acids Res.* **51**:D389–D394. doi:10.1093/nar/gkac1022

879 Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J, Crowl AA, Deng
880 M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-Rodríguez
881 A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa
882 H, Schulze ED, Simeone MC, Sork VL, Valencia-Avalos S. 2020. Genomic landscape
883 of the global oak phylogeny. *New Phytol.* **226**:1198–1212. doi:10.1111/nph.16162

884 Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic
885 period during organogenesis. *Nat. Commun.* **2**. doi:10.1038/ncomms1248

886 Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U,
887 Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the
888 developmental hourglass model. *Nature* **468**:811–816. doi:10.1038/nature09634

889 Kant P, Kant S, Gordon M, Shaked R, Barak S. 2007. Stress Response Suppressor1 and
890 Stress Response Suppressor2, two Dead-box RNA helicases that attenuate
891 Arabidopsis responses to multiple abiotic stresses. *Plant Physiol.* **145**:814–830.
892 doi:10.1104/pp.107.099895

893 Keilwagen J, Hartung F, Grau J. 2019. GeMoMa: Homology-based gene prediction
894 utilizing intron position conservation and RNA-seq dataMethods in Molecular Biology.
895 Humana Press Inc. pp. 161–177. doi:10.1007/978-1-4939-9173-0_9

896 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment
897 and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**:907–915.
898 doi:10.1038/s41587-019-0201-4

899 Kokot M, Dlugosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer
900 statistics. *Bioinformatics* **33**:2759–2761. doi:10.1093/bioinformatics/btx304

901 Komoto H, Nagahama A, Miyawaki-Kuwakado A, Hata Y, Kyozuka J, Kajita Y, Toyama H,
902 Satake A. 2024. The transcriptional changes underlying the flowering phenology shift
903 of *Arabidopsis halleri* in response to climate warming. *Plant Cell Environ.* **47**:174–
904 186. doi:10.1111/pce.14716

905 Kremer A, Abbott AG, Carlson JE, Manos PS, Plomion C, Sisco P, Staton ME, Ueno S,
906 Vendramin GG. 2012. Genomics of Fagaceae. *Tree Genet. Genomes* **8**:583–610.
907 doi:10.1007/s11295-012-0498-3

908 Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M, Rosen BD,
909 Cheng CY, Moreira W, Mock SA, Stubbs J, Sullivan JM, Krampis K, Miller JR,
910 Micklem G, Vaughn M, Town CD. 2015. Araport: the Arabidopsis information portal.
911 *Nucleic Acids Res.* **43**:D1003–D1009. doi:10.1093/nar/gku1200

912 Kubo T, Iwasa Y. 1996. Phenological pattern of tree regeneration in a model for forest
913 species diversity. *Theor. Popul. Biol.* **49**:90–117. doi:10.1006/tpbi.1996.0004

914 Kudoh H. 2016. Molecular phenology in plants: *in natura* systems biology for the
915 comprehensive understanding of seasonal responses under natural environments.
916 *New Phytol.* **210**:399–412. doi:10.1111/nph.13733

917 Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins and gene
918 expression levels are coupled in *Drosophila* and are independently associated with
919 mRNA abundance, protein length, and number of protein-protein interactions. *Mol.*
920 *Biol. Evol.* **22**:1345–1354. doi:10.1093/molbev/msi122

921 Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver
922 DH, Feder M, Fernandez-Valverde SL, Nakanishi N, Simmons D, Simakov O,
923 Larsson T, Liu SY, Jerafi-Vider A, Yaniv K, Ryan JF, Martindale MQ, Rink JC, Arendt
924 D, Degnan SM, Degnan BM, Hashimshony T, Yanai I. 2016. The mid-developmental
925 transition and the evolution of animal body plans. *Nature* **531**:637–641.
926 doi:10.1038/nature16994

927 Li B, Dewey CN. 2011. RSEM: Accurate transcript quantification from RNA-Seq data with
928 or without a reference genome. *BMC Bioinformatics* **12**. doi:10.1186/1471-2105-12-
929 323

930 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
931 transform. *Bioinformatics* **25**:1754–1760. doi:10.1093/bioinformatics/btp324

932 Liao BY, Zhang J. 2006. Low rates of expression profile divergence in highly expressed
933 genes and tissue-specific genes during mammalian evolution. *Mol. Biol. Evol.*
934 **23**:1119–1128. doi:10.1093/molbev/msj119

935 Liu H, Zhang R, Zhou BF, Shen Z, Chen XY, Gao J, Wang B. 2023. Chromosome-scale
936 genome assembly of sweet tea (*Lithocarpus polystachyus Rehder*). *Sci. Data* **10**.
937 doi:10.1038/s41597-023-02791-y

938 Lotharukpong JS, Zheng M, Luthringer R, Liesner D, Drost H-G, Coelho SM. 2024. A
939 transcriptomic hourglass in brown algae. *Nature* **635**:129–135. doi:10.1038/s41586-
940 024-08059-8

941 Manos PS, Zhou Z, Cannon CH. 2001. Systematics of Fagaceae: phylogenetic tests of
942 reproductive trait evolution. *Int. J. Plant Sci.* **162**:1361–1379. doi:10.1086/322949

943 Mishra B, Ulaszewski B, Meger J, Aury JM, Bodénès C, Lesur-Kupin I, Pfenninger M, Da
944 Silva C, Gupta DK, Guichoux E, Heer K, Lalanne C, Labadie K, Opgenoorth L, Ploch
945 S, Le Provost G, Salse J, Scotti I, Wötzel S, Plomion C, Burczyk J, Thines M. 2022.
946 A Chromosome-level genome assembly of the European beech (*Fagus sylvatica*)
947 reveals anomalies for organelle DNA integration, repeat content and distribution of
948 SNPs. *Front. Genet.* **12**. doi:10.3389/fgene.2021.691058

949 Miyawaki-Kuwakado A, Han Q, Kitamura K, Satake A. 2024. Impacts of climate change
950 on the transcriptional dynamics and timing of bud dormancy release in Yoshino-
951 cherry tree. *Plants, People, Planet* **6**:1505–1521. doi:10.1002/ppp3.10548

952 Miyazaki Y, Maruyama Y, Chiba Y, Kobayashi MJ, Joseph B, Shimizu KK, Mochida K,
953 Hiura T, Kon H, Satake A. 2014. Nitrogen as a key regulator of flowering in Fagus
954 crenata: understanding the physiological mechanism of masting by gene expression
955 analysis. *Ecol. Lett.* **17**:1299–1309. doi:10.1111/ele.12338

956 Morellato LPC, Camargo MGG, Gressler E. 2013. “A review of plant phenology in South
957 and Central America.” in Phenology: An Integrative Environmental Science. Springer,
958 91-113.

959 Nagano AJ, Kawagoe T, Sugisaka J, Honjo MN, Iwayama K, Kudoh H. 2019. Annual
960 transcriptome dynamics in natural environments reveals plant seasonal adaptation.
961 *Nat. Plants* **5**:74–83. doi:10.1038/s41477-018-0338-z

962 Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, Page DC. 2019.
963 Conservation, acquisition, and functional impact of sex-biased gene expression in
964 mammals. *Science (1979)* **365**. doi:10.1126/science.aaw7317

965 Nishio H, Nagano AJ, Ito T, Suzuki Y, Kudoh H. 2020. Seasonal plasticity and diel stability
966 of H3K27me3 in natural fluctuating environments. *Nat. Plants* **6**:1091–1097.
967 doi:10.1038/s41477-020-00757-1

968 Nuzhdin S V., Wayne ML, Harmon KL, McIntyre LM. 2004. Common pattern of evolution
969 of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.*
970 **21**:1308–1317. doi:10.1093/molbev/msh128

971 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and
972 bias-aware quantification of transcript expression. *Nat. Methods* **14**:417–419.
973 doi:10.1038/nmeth.4197

974 Pertea M, Pertea G. 2020. GFF Utilities: GffRead and GffCompare. *F1000Res.* **9**.
975 doi:10.12688/f1000research.23297.1

976 Piao S, Liu Q, Chen A, Janssens IA, Fu Y, Dai J, Liu L, Lian X, Shen M, Zhu X. 2019.
977 Plant phenology and global climate change: current progresses and challenges. *Glob.*
978 *Chang. Biol.* **25**:1922–1940. doi:10.1111/gcb.14619

979 Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution.
980 *Proc. Natl. Acad. Sci.* **104**:8605–8612. doi:10.1073/pnas.0700488104

981 Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A,
982 Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. 2016. Chromosome-
983 scale shotgun assembly using an in vitro method for long-range linkage. *Genome*
984 *Res.* **26**:342–350. doi:10.1101/gr.193474.115

985 Quint M, Drost HG, Gabel A, Ullrich KK, Bönn M, Grosse I. 2012. A transcriptomic
986 hourglass in plant embryogenesis. *Nature* **490**:98–101. doi:10.1038/nature11394

987 Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot
988 for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**.
989 doi:10.1038/s41467-020-14998-3

990 Satake A, Hagiwara T, Nagano AJ, Yamaguchi N, Sekimoto K, Shiojiri K, Sudo K. 2024a.
991 Plant molecular phenology and climate feedbacks mediated by BVOCs. *Annu. Rev.*
992 *Plant Biol.* **75**:605–627. doi:10.1146/annurev-arplant-060223-032108

993 Satake A, Imai R, Fujino T, Tomimoto S, Ohta K, Na'iem M, Indrioko S, Widiyatno W,
994 Purnomo S, Morales AM, Nizhynska V, Tani N, Suyama Y, Sasaki E, Kasahara M.
995 2024b. Somatic mutation rates scale with time not growth rate in long-lived tropical
996 trees. *Elife* **12**. doi:10.7554/eLife.88456

997 Satake A, Nagahama A, Sasaki E. 2022. A cross-scale approach to unravel the molecular
998 basis of plant phenology in temperate and tropical climates. *New Phytol.* **233**:2340–
999 2353. doi:10.1111/nph.17897

1000 Satake A, Ohta K, Takeda-Kamiya N, Toyooka K, Kusumi J. 2023. Seasonal gene
1001 expression signatures of delayed fertilization in Fagaceae. *Mol. Ecol.* **32**:4801–4813.
1002 doi:10.1111/mec.17079

1003 Schwartz MD. 2013. Phenology: An Integrative Environmental Science, 2nd ed. Springer
1004 Dordrecht.

1005 Smaczniak C, Immink RGH, Muiño JM, Blanvillain R, Busscher M, Busscher-Lange J,
1006 Dinh QD, Liu S, Westphal AH, Boeren S, Parcy F, Xu L, Carles CC, Angenent GC,
1007 Kaufmann K. 2012. Characterization of MADS-domain transcription factor complexes
1008 in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U. S. A.* **109**:1560–1565.
1009 doi:10.1073/pnas.1112871109

1010 Smid M, Coebergh van den Braak RRJ, van de Werken HJG, van Riet J, van Galen A, de
1011 Weerd V, van der Vlugt-Daane M, Bril SI, Lalmahomed ZS, Kloosterman WP, Wilting
1012 SM, Foekens JA, IJzermans JNM, Martens JWM, Sieuwerts AM. 2018. Gene length
1013 corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data

1014 performs similarly in intersample analyses while improving intrasample comparisons.
1015 *BMC Bioinformatics* **19**. doi:10.1186/s12859-018-2246-7

1016 Thaben PF, Westermark PO. 2014. Detecting rhythms in time series with rain. *J Biol*
1017 *Rhythms* **29**:391–400. doi:10.1177/0748730414553029

1018 True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary
1019 trajectories. *Evol. Dev.* **3**:109–119. doi:10.1046/j.1525-142x.2001.003002109.x

1020 Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H,
1021 Kissinger JC, Paterson AH. 2012. MCScanX: A toolkit for detection and evolutionary
1022 analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**.
1023 doi:10.1093/nar/gkr1293

1024 Yang L, Gaut BS. 2011. Factors that contribute to variation in evolutionary rate among
1025 arabidopsis genes. *Mol. Biol. Evol.* **28**:2359–2369. doi:10.1093/molbev/msr058

1026 Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*
1027 **24**:1586–1591. doi:10.1093/molbev/msm088

1028 Yoo SJ, Chung KS, Jung SH, Yoo SY, Lee JS, Ahn JH. 2010. BROTHER of FT and TFL1
1029 (BFT) has TFL1-like activity and functions redundantly with TFL1 in inflorescence
1030 meristem development in *Arabidopsis*. *Plant J.* **63**:241–253. doi:10.1111/j.1365-
1031 313X.2010.04234.x

1032 Zhang X, Yu G, Dai Y, Zhang H, Wang K, Han J. 2023. High-resolution Hi-C maps highlight
1033 multiscale chromatin architecture reorganization during cold stress in Brachypodium
1034 distachyon. *BMC Plant Biol.* **23**. doi:10.1186/s12870-023-04269-w

1035 Zhou BF, Yuan S, Crowl AA, Liang YY, Shi Y, Chen XY, An QQ, Kang M, Manos PS,
1036 Wang B. 2022. Phylogenomic analyses highlight innovation and introgression in the
1037 continental radiations of Fagaceae across the Northern Hemisphere. *Nat. Commun.*
1038 **13**. doi:10.1038/s41467-022-28917-1

1039 Zhu GQ, Qu L, Xue HW. 2024. Casein kinase 1 AELs promote senescence by enhancing
1040 ethylene biosynthesis through phosphorylating WRKY22 transcription factor. *New*
1041 *Phytol.* **244**:116–130. doi:10.1111/nph.19785

1042 Zoldos V, Papes D, Cerbah M, Panaud O, Besendorfer V, Siljak-Yakovlev S. 1999.
1043 Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal
1044 conserved genome organization among 11 *Quercus* species. *Theor. Appl. Genet.*
1045 **99**:969–977. doi:10.1007/s001220051404

1046 Zrimec J, Börlin CS, Buric F, Muhammad AS, Chen R, Siewers V, Verendel V, Nielsen J,
1047 Töpel M, Zelezniak A. 2020. Deep learning suggests that gene expression is encoded
1048 in all parts of a co-evolving interacting gene regulatory structure. *Nat. Commun.* **11**.
1049 doi:10.1038/s41467-020-19921-4

1050 Zrimec J, Buric F, Kokina M, Garcia V, Zelezniak A. 2021. Learning the regulatory code of
1051 gene expression. *Front. Mol. Biosci.* **8**. doi:10.3389/fmolb.2021.673363

1052

1053

1054 **Figures and Tables**

1055 **Figure 1.** Synteny and distribution of genomic features of *Q. glauca* and *L. edulis*
1056 genomes. (A) A phylogenetic tree of five Fagaceae species from three genera,
1057 *Quercus*, *Lithocarpus*, and *Fagus*. Bootstrap values for all branches are 100. (B)
1058 Synteny relationship between the *Q. glauca* and *L. edulis* genomes. The collinear
1059 blocks within the genomes were displayed by lines. Black lines show reverse
1060 alignment and gray lines represent regular. Numbers correspond to the 12
1061 chromosomes. (C–D) Circos plots of the *Q. glauca* (C) and *L. edulis* genomes (D).
1062 Lanes depict circular representation of chromosome length (Mbp) (a), gene density
1063 (b), GC density (c), and repeat (d). Lines in the inner circle represent links between
1064 synteny-selected paralogs. (E) Gene copy-number variations between the *Q.*
1065 *glauca* and *L. edulis* genomes. The bar plot illustrates the number of genes in gene
1066 families categorized into 1:1 orthologs, N:N (many-to-many), 1:N (one-to-many),
1067 N:1 (many-to-one), and species-specific unique genes. (F–G) Mean gene
1068 expression levels in each copy number variation category for *Q. glauca* (F) and *L.*

edulis (G). Different letters denote statistically significant differences ($p < 0.05$, Steel-Dwass test).

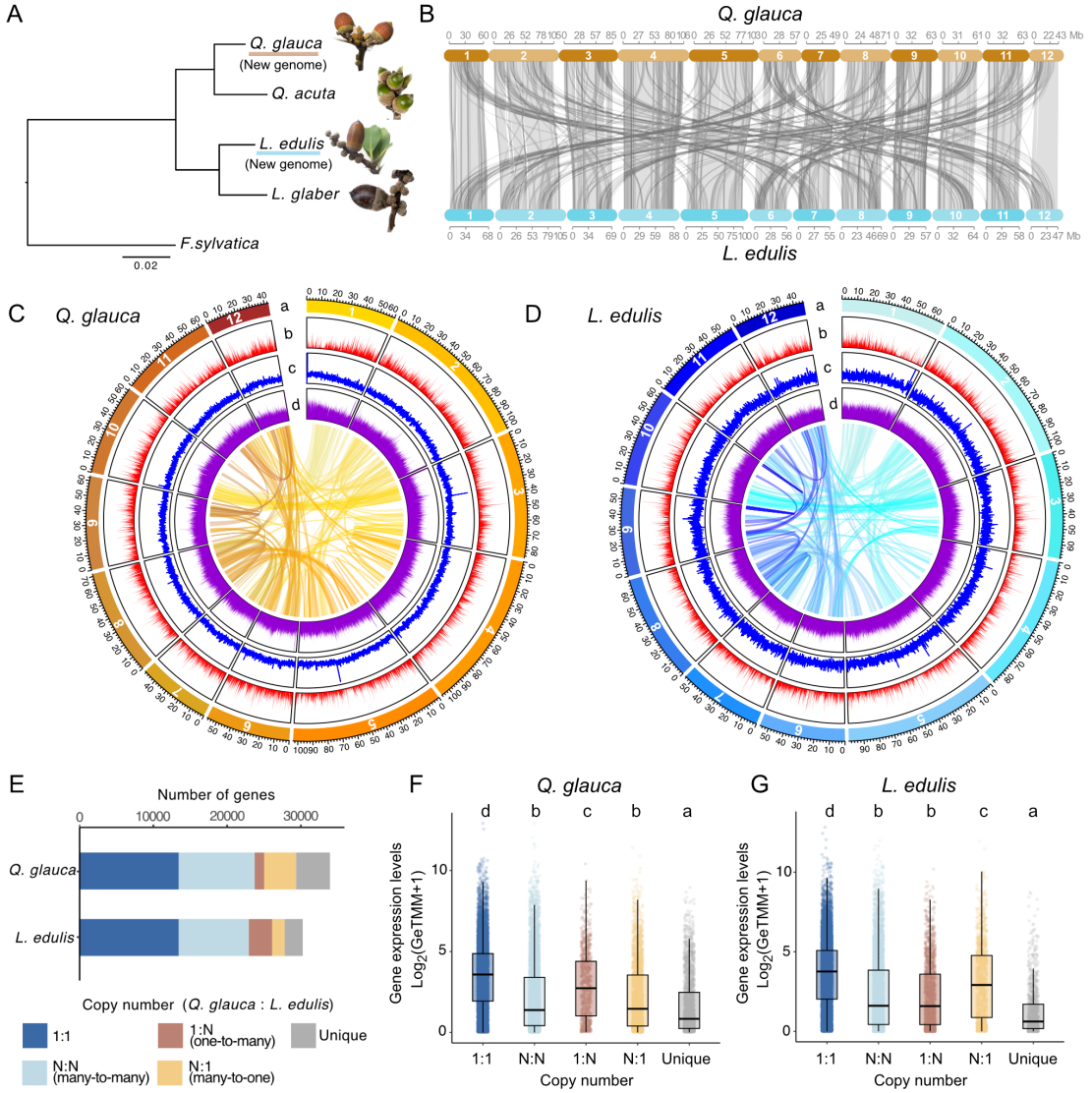


Figure 2. Seasonal gene expression dynamics. (A) Hierarchical clustering of 213 gene expression profiles encompassing two tissues (leaf and bud) and four species (*Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber*), collected over two-year. (B) Distribution of daily mean temperature on each sampling date across clusters, from L1 (top) to B3 (bottom). The Mann-Whitney U test was used to compare daily mean temperature between cluster L1 and L2 (p = 0.136). The Steel-Dwass test was applied for multiple comparisons across clusters B1–B3 ($p < 0.05$). Different letters denote statistically significant differences ($p < 0.05$). (C–D) Hierarchical clustering of 11749 genes based on the similarity of gene expression dynamics. Daily mean temperature, flowering and leaf flushing phenology (proportion) are shown in the upper three panels of the heatmaps.

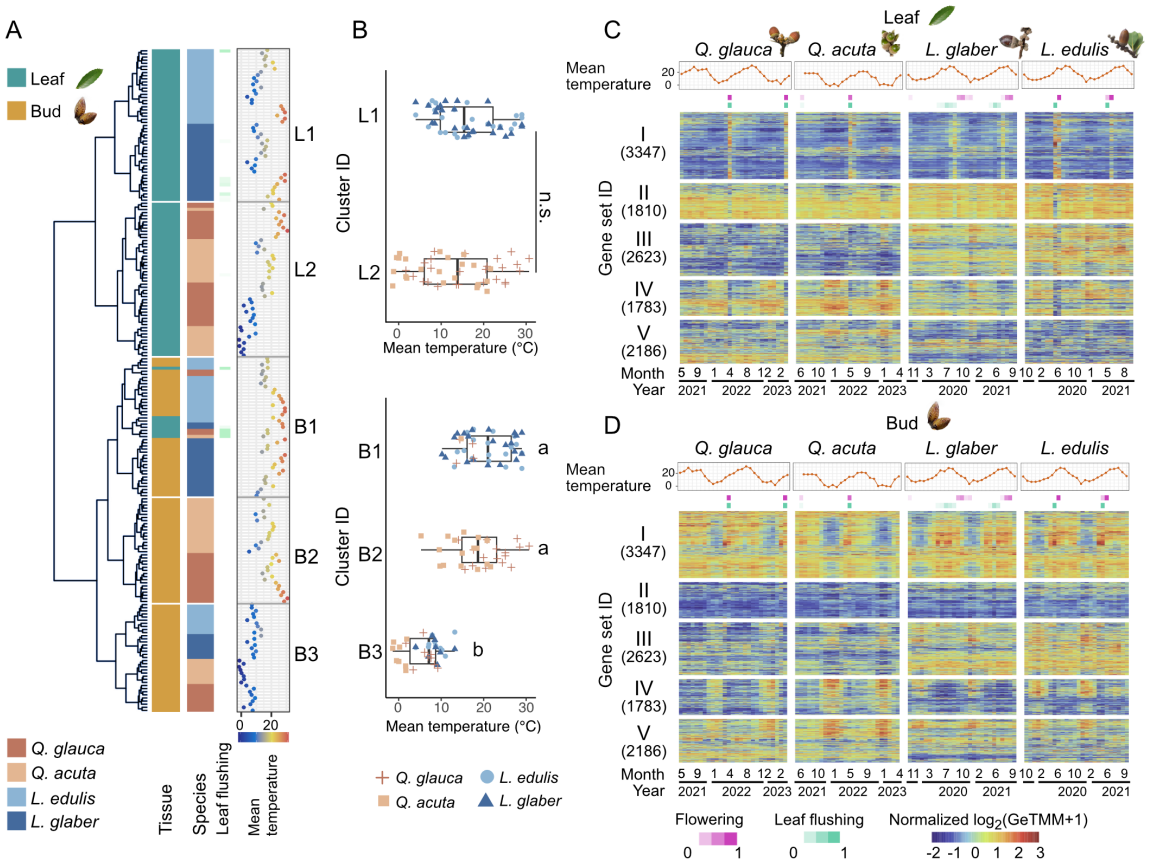


Figure 3. Distribution of period and peak month in rhythmic gene expression. (A–B) Distribution of period length (month) among the genes with significant rhythmicity calculated by RAIN for leaf (A) and bud (B) samples of each species. (C–D) Distribution of seasonal peaks among the genes with annual periodicity. The angle of each bar represents the peak month, while the bar height indicates the number of genes. Purple and green squares around each rose plot denote the observed phenology of flowering and leaf flushing (proportion), respectively.

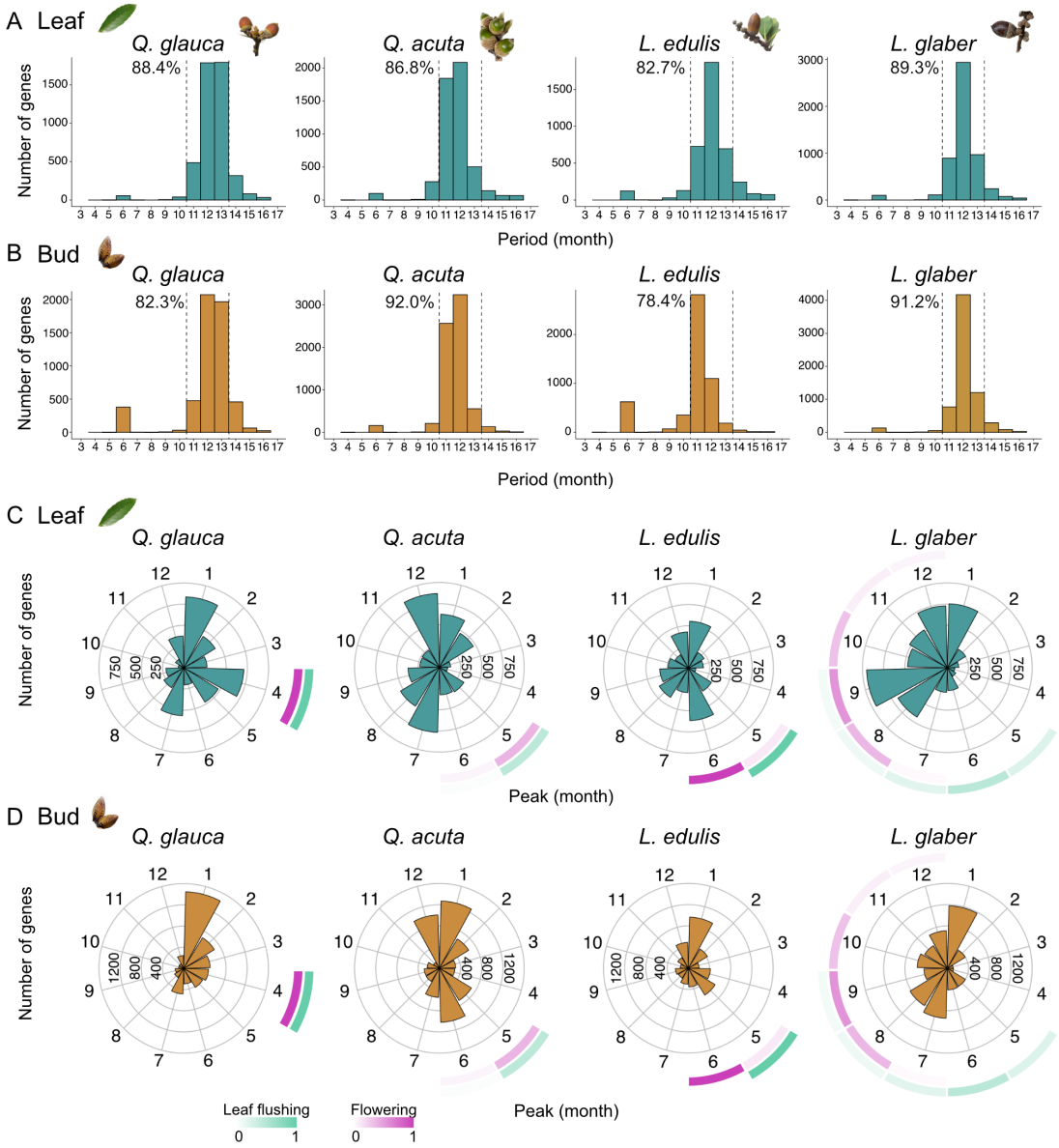


Figure 4. Comparison of peak months in seasonal gene expression across species in buds. (A) Intra-genus and (B) inter-genera comparisons of peak months in seasonal gene expression. The size and color of the circles represent the number of genes and the mean of Pearson's correlation, respectively. Genes with negative correlations lower than -0.3 are highlighted in blue, while the month of leaf flushing is highlighted in green. (C) Plot of molecular phenology divergence index *D* across months. Triangles and circles indicate the values of *D* calculated from intra-genus and inter-genera comparison, respectively. The line and shaded envelope indicate the mean and standard deviation (s.d.). Different letters denote statistically significant differences (Nemenyi test, $p < 0.05$).

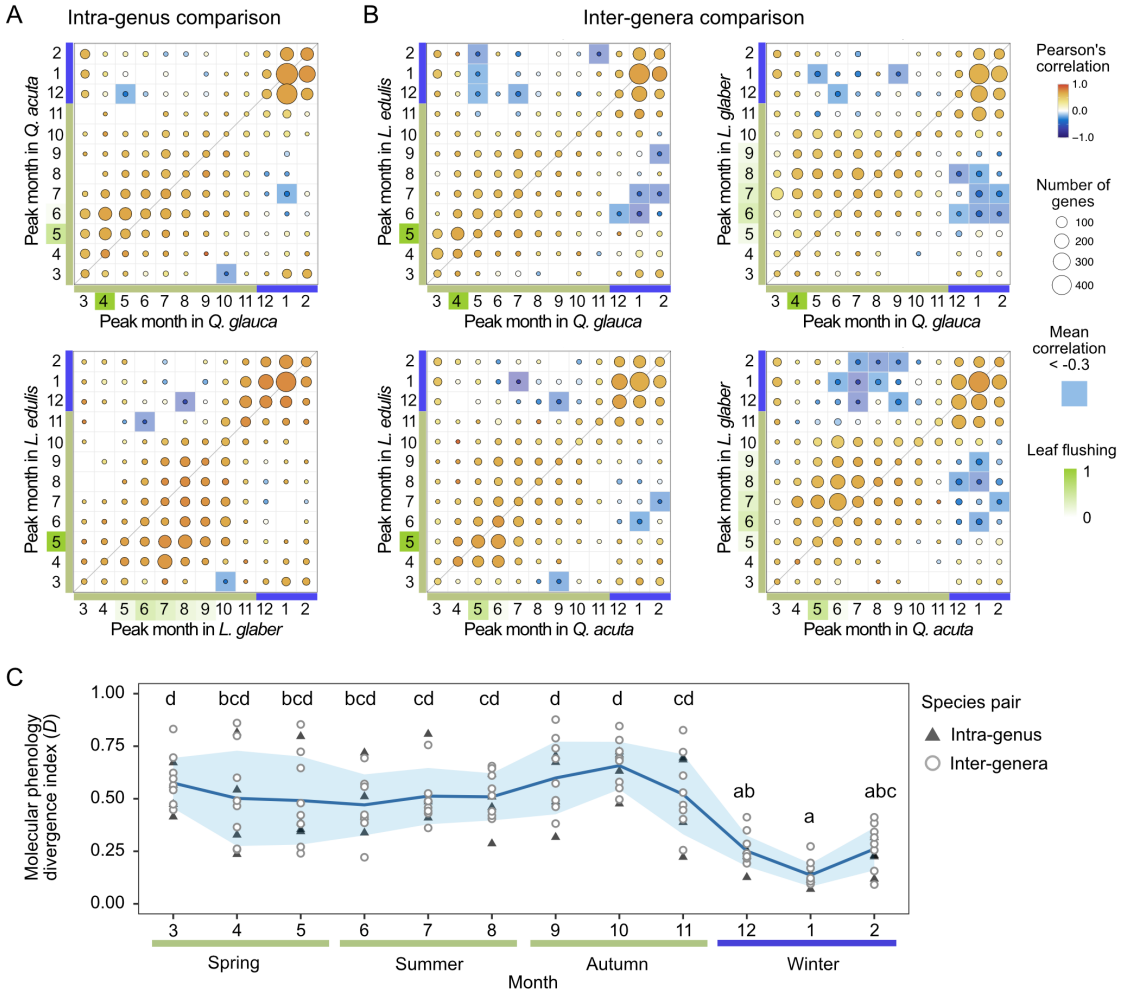
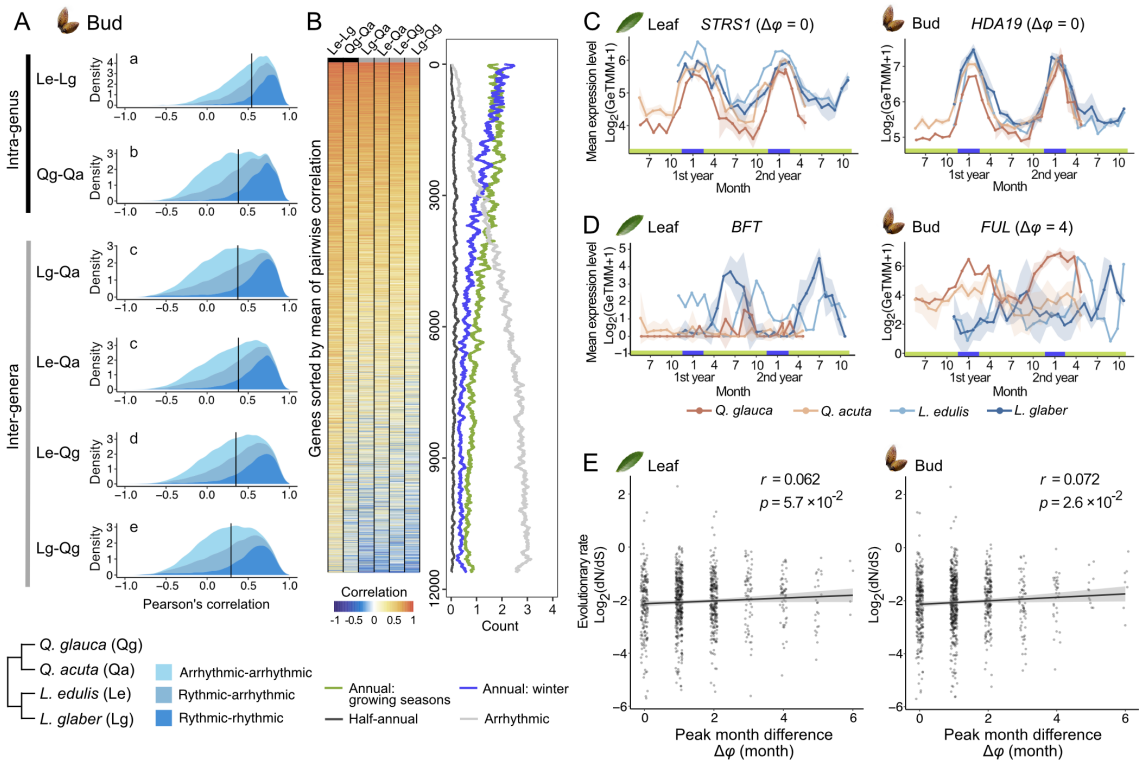


Figure 5. Phylogenetic constraints in the evolution of seasonal gene expression and the relationship between seasonal gene expression divergence and sequence evolution. (A) Distribution of pairwise Pearson's correlation coefficients for gene expression across species (Le: *L. edulis*, Lg: *L. glaber*, Qg: *Q. glauca*, Qa: *Q. acuta*; n = 11749). Colors indicate rhythmicity categories. Black bars show median correlation values, and different letters indicate significant differences (Nemenyi test, $p < 0.05$). (B) Heatmap of pairwise correlation coefficients for each gene, with mean gene counts in four seasonal expression categories: (1) annual rhythmic genes peaking in the growing season (March–November), (2) annual rhythmic genes peaking in winter (December–February), (3) half-annual rhythmic genes, and (4) arrhythmic genes. Genes are ordered by mean correlation across species pairs. (C–D) Conserved (C) and diverged (D) gene expression patterns, based on the top and bottom 5% of pairwise correlations. Solid lines and error bars indicate mean ± SD. Numbers in parentheses show peak month differences between *Q. glauca* and *L. edulis*. Peak month difference for *BFT* was not calculated due to arrhythmic expression in *Q. glauca*. (E) Relationship between evolutionary rate (dN/dS) and peak month difference ($\Delta\varphi$) for genes with annual periodicity in *Q. glauca* and *L. edulis* (n = 951). Black lines and transparent bands represent regression lines and 95% confidence intervals.



Additional files

Supplementary file 1

Supplementary figures (Fig. S1 to Fig. S13) and a supplementary table (Table S1) were included in this file.

Supplementary Dataset

Dataset S1

Dataset S2

Dataset S3

Dataset S4

Dataset S5

Dataset S6