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

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.

Keywords: Phenology, Gene expression evolution, Seasonal constraints

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 shaping seasonal gene expression remain unclear. To investigate the evolutionary patterns of gene expression in seasonal environments, 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 for *Q. glauca* and *L. edulis*, 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 synchronized across species in winter but diverged during the growing season, aligning with phenological events such as leaf flushing and flowering. Genes with conserved seasonal peaks exhibited slower evolutionary rates, suggesting weak selective pressures maintaining seasonal gene expression. However, the evolutionary rates were similarly slow regardless of whether genes were expressed in winter or the growing season. These findings suggest that winter gene expression is more conserved than that in the growing season, imposing a seasonal constraint on gene expression evolution. This constraint may limit temporal niche partitioning and slow species divergence rates in 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 (1). 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 (2–5). Phenology is regulated by the coordinated expression of genes that enable organisms to exhibit essential functions at the appropriate times (6–8). However, gene expression dynamics underpinning seasonal responses in organisms remain complex and 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 (9–11), *Drosophila* (12, 13), and Tanzanian cichlids (14), 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 (9–11), 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 (15–18). Molecular phenology refers to transcriptome analysis conducted under natural fluctuating conditions (6, 19–21) 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 (7). 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. 1*A*). To obtain high-quality chromosome-level reference genomes, we conducted a hybrid genome assembly for the HiFi reads (total length of subreads: 273.6 Gb, 165× coverage) and the high-throughput chromosome conformation capture (Hi-C) reads (total length: 48.8 Gb, 54.8× coverage). 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 1*B*). The number of chromosome-scale scaffolds were identical to the known numbers in other *Quercus* (22) and *Lithocarpus* species (23). 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 (Table S1). The genome

sizes of *Q. glauca* and *L. edulis* were estimated 891.2 Mb and 868.7 Mb by KmerGenie (24) and GenomeScope2 (25), respectively.

Overall, the abundance and distribution of protein-coding genes, GC content, and repeat sequences (Fig. 1*C*, *D;* Table S1) and syntenic structure (Fig. 1*B*) were similar between *Q. glauca* and *L. edulis*. Using OrthoFinder2 (26), we clustered genes from the two newly assembled genomes into orthologous groups. We identified 13427 single-copy orthologs, enabling pairwise comparisons between species (Fig. 1*E*). Single-copy genes exhibited significantly higher average expression levels than multi-copy genes (Steel-Dwass test, $p < 0.05$, Fig. 1*F*, *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 (27). This process resulted in a high-quality expression dataset for 11749 genes (Fig. S1*A*). 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. S1*B*). 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 laboratory conditions (28). Consistently, our results demonstrate that despite substantial seasonal fluctuations in environmental conditions (Fig. S2), individual gene expression remains relatively conserved. On average, the magnitude of genome-wide variation was 15.8 times greater than the seasonal variation observed for individual genes (Fig. S1*B*).

To assess the similarity of gene expression profiles across tissues, species, and seasons, we performed a hierarchical clustering of seasonal transcriptome profiles from leaf and bud tissues across four species and identified five distinct clusters (cluster L1, L2, B1–B3) according to the Elbow method (Fig. 2*A*; Fig. S3*A*; Datasets S1). Transcriptional profiles first grouped into leaves (cluster L1 and L2) and buds (cluster B1–B3), irrespective of seasons and species (Fig. 2*A*), except

that newly flushed leaf samples were grouped closely with buds (Fig. 2*A*; Fig. S4). Within the leaf cluster, transcriptome profiles were first clustered by genus and then by species in *Lithocarpus* (cluster L1; Fig. 2*A*). In contrast, samples from *Quercus* species formed a mixture of two species (cluster L2; Fig. 2*A*). Notably, the winter samples of *Q. acuta*, collected during periods of low mean temperatures, were distinctly separated from those collected in other seasons (cluster L2; Fig. 2*A*). Within the bud cluster, transcriptome profiles were predominantly grouped by season—winter (cluster B3; November to February, with the exception of a March sample from *L. edulis*; Datasets S1) versus the other growing seasons (cluster B1 and B2; March to November; Datasets S1)—and subsequently organized by genus and species. This suggests that the winter season exerts a stronger influence on gene expression in buds than phylogenetic relationships.

Within the bud cluster, transcriptional profiles in the winter cluster were observed under significantly lower temperatures, approximately below 10°C, compared to other seasons (Fig. 2*B*; Fig. S5*A*). In contrast, no such temperature-related distinction was evident within the leaf cluster (Fig. 2*B*). Additionally, photoperiods were markedly shorter in the winter cluster in buds (Fig. S5*B*). These findings suggest that the buds of the four species exhibit a conserved response to environmental factors characteristic of the winter season. To further investigate the relationships among all samples, we conducted a principal component analysis (PCA). The first principal component (PC1) clearly separated samples by tissue type, while the second (PC2) and the third principal component (PC3) gradually differentiated winter samples from those collected during other seasons (Fig. S6).

To investigate seasonal expression dynamics of each gene, we conducted hierarchical clustering on the transcriptional dynamics of 11749 genes and identified five distinct gene sets characterizing different transcriptional modes (gene set I–V; Fig. 2*C, D*; Fig. S3*B*; Datasets S2). Gene set I exhibited predominant expression in buds, with seasonal peaks during summer, whereas gene set II was primarily expressed in leaves, highlighting clear tissue-specificity (Fig. 2*C*, *D*). Gene sets III and V demonstrated genus-specific expression patterns, with higher expression levels observed in *Lithocarpus* and *Quercus*, respectively (Fig. 2*C*, *D*). In contrast, gene set IV was expressed during winter across both leaf and bud tissues (Fig. 2*C*, *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 (29). 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. S7*A*; Datasets 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. S7*A*).

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. 3*A*-*B*, Datasets 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. 3*A*, *B*). The fraction of genes exhibiting other periodicities was relatively small (5.7–14.1%; Fig. 3*A*-*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. S7*A*). 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. S7*B*). In contrast, most genes exhibiting half-annual periodicity were species-specific, with no genes shared among all the species (Fig. S7*C*).

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. 3*C*, *D*; Datasets S3). On average, 34.9% and 41.3% of annual rhythmic genes peaked in winter (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. 3*C*). 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. 3*D*). 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. S7*D*; Dataset S3), with the exception of *L. glaber*, an autumn flowering species. Interestingly, *WRKY22*, a transcription factor regulating ethylene biosynthesis and senescence (30), 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 (31), showed two distinct peaks in *Q. glauca* and *L. edulis*, occurring in both summer (July) and winter (January) (Fig. S7*E*). 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. 4*A*) and inter-genera species pairs in buds (Fig. 4*B*) and leaves (Fig. S8*A*, *B*). During winter, a large number of genes exhibited highly synchronized expression peaks across species, with high pairwise Pearson's correlation coefficients observed (Fig. 4*A*, *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. 4*A*, *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. 4*B*) than in intra-genus comparisons (4–12 genes; Fig. 4*A*). A similar pattern was observed in leaf samples (Fig. S8*A, B*). When comparing seasonal expression peaks between tissues, the majority of genes were synchronously activated in winter in both tissues (Fig. S8*C*). During spring to summer, slight variations in seasonal expression peaks were observed among tissues (Fig. S8*C*).

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. 4*C*). A similar trend was observed in leaf samples, with another trough in summer (Fig. S8*D*). 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 (Datasets 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. S9*A*) and buds (Steel-Dwass test, $p < 0.05$; Fig. 5*A*). 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. S9*A*). In particular, genes with seasonal expression peaks in winter showed high correlation across species in both tissues (Fig. 5*B*; Fig. S9*B*). 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. 5*B*; Fig. S9*B*). 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. S10*A*, *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. S10*C*, *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 on the distribution of mean correlation (Fig. 5*B*; Fig. S9*B*) and conducted Gene Ontology (GO) enrichment analysis. GO analysis revealed that the top 5% of genes in leaves were associated with photosynthesis, while those in buds were linked to cell fate determination (Fig. S11*A*; Datasets S5). The gene with the highest correlation in leaves was *STRESS RESPONSE SUPPRESSOR1* (*STRS1*; mean of pairwise correlation = 0.91; Fig. 5*C*), a DEAD-box RNA helicase previously reported to enhance tolerance to abiotic stresses such as cold (32). In buds, the most highly correlated gene was the histone deacetylase *HDA19* (mean of pairwise correlation = 0.93; Fig. 5*C*), which plays a key role in environmental stress responses through histone modification and chromatin remodeling (33). These findings suggest that genes associated with photosynthesis during the growing season and stress responses during winter exhibit highly conserved seasonal expression patterns across species.

The bottom 5% of genes were predominantly associated with DNA replication and cell division in leaves (Fig. S11*B*), aligning with the divergent leaf phenology observed (Fig. 3*B*). In contrast, GO terms related to immune response were weakly enriched in buds (Fig. S11*B*). Notably, we identified MADS-box genes linked to floral transition among the bottom 5% of genes: *BROTHER OF FT AND TFL1* (*BFT*) in leaves (mean of pairwise correlation = -0.094) and *FRUITFULL* (*FUL*) in buds (mean of pairwise correlation = -0.069). *BFT*, a floral repressor (34), was expressed in autumn in the spring-flowering species *L. edulis*, and in spring in the autumn-flowering species *L. glaber* leaves (Fig. 5*D*). Meanwhile, *FRUITFULL* (*FUL*), a positive regulator of floral transition (35), was expressed in spring in *L. edulis* and in autumn in *L. glaber* (Fig. 5*D*), mirroring the timing of flowering. In *Q. glauca* and *Q. acuta*, *FUL* was expressed in winter (Fig. 5*D*), nearly out of phase with their expressions in *Lithocarpus* species. These results suggest that genes involved in reproductive processes exhibit species-specific divergence in expression patterns in response to seasonal environmental changes.

Relating seasonal gene expression divergence to sequence divergence. If divergence of seasonal timing of gene expression is related to selective constraint, we would expect a correlation between sequence divergence patterns and the divergence in seasonal expression peaks. To explore the relationship between the

divergence of seasonal gene expression and protein sequence evolution, we calculated synonymous (dS) and non-synonymous substitution rates (dN) and their ratio (dN/dS) for each gene using coding sequences from newly assembled genomes of *Q. glauca* and *L. edulis*. We first confirmed that the highly expressed genes evolve more slowly by identifying a significant negative correlation between evolutionary rate (dN/dS) and mean of the gene expression (r = -0.21, p = 1.8 × 10^{-118} for dN/dS; Fig. S12*A*; Datasets S6). This finding aligns with previous studies in yeasts, mammals, and plants (36–39). By targeting the genes that exhibited annual periodicity across both species and tissues, we examined the relationship 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. 5*E*; Fig. S12*B*,*C*; Datasets 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*; Datasets 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. S12*E*; Datasets 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. 2*B*). 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 (40), 11.3°C in

395 ever-green *Fagaceae* (20), and 10.5°C in perennial *Arabidopsis* (41). These
396 results suggest the presence of a conserved mechanism for gene expression
397 regulation under low-temperature environments across different species. The
398 greater conservation of gene expression patterns in buds compared to leaves
399 across species may be attributed to their functional role protecting essential
400 meristematic tissues, including the shoot apex and leaf and flower primordia.
401 Additionally, buds likely play a crucial role in maintaining common physiological
402 functions necessary for the resumption of growth in spring. This functional
403 consistency may impose evolutionary constraints on gene expression, contributing
404 to its conservation across species.
405 It is intriguing to compare the seasonal constraints identified in this study
406 with developmental constraints in gene expression evolution, where evolutionary
407 conservation and variation across developmental stages during the life cycle have
408 been observed, such as in the early conservation model and the developmental
409 hourglass model (Fig. S13*A*)(42–44). Previous studies have demonstrated that
410 certain developmental stages exhibit remarkably similar gene expression patterns
411 (45–47). Our findings highlight that such highly similar gene expression patterns
412 are observed during the winter season (Fig. S13*B*), suggesting that a common
413 physiological state is essential for enduring harsh winter conditions. This seasonal
414 constraint on gene expression evolution may have the potential to limit temporal
415 niche partitioning at the phenotypic level and slow species divergence rates in
416 seasonal environments (48, 49). Although completely disentangling the effects of
417 developmental stages and seasonal changes on gene expression is challenging
418 when using molecular phenology data collected from natural environments, the
419 observation that species with different phenology—i.e., different developmental
420 timing in seasonal environments—exhibit convergent gene expression patterns
421 during winter underscores the profound influence of winter environmental
422 conditions on the evolution of gene expression. Future research is needed to
423 elucidate the mechanisms driving the establishment of such conserved gene
424 expression patterns, including the dynamic reorganization of chromatin
425 architecture in response to cold (50–52).
426 We found that the evolution of coding-sequences was weakly correlated
427 with the divergence of seasonal peaks in gene expression across species. This
428 suggests that weak functional constraints exist for specific proteins that are
429 expressed in certain seasons, such as photosynthesis in summer and stress
430 response to winter. However, among genes with conserved seasonal peaks, the
431 evolutionary rate of coding-sequences was similarly slow regardless of whether
432 they were expressed in winter or during the growing season. Therefore, to
433 understand the evolutionary mechanism underlying the seasonal constraint on
434 gene expression that results in similar expression patterns during winter, it is
435 essential to consider evolution of non-coding sequences that modulate gene
436 expression levels in response to seasonal environmental changes. Regulatory
437 variation has been recognized as a major driver of evolutionary innovation (53, 54).
438 Recent deep learning studies utilizing genome and transcriptome data from yeast
439 and humans have highlighted the significance of co-evolution between coding and

non-coding sequences in shaping gene expression levels (28, 55). 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 facilitate the development of predictive models for evolutionary processes influenced by seasonal dynamics. Comparison of molecular phenology across wide climatic zones, including both tropical and temperate zones, will also be effective to identify key genes for the phenological response to on-going climate change.

Materials and Methods

Sample collection for genome sequencing

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

Genome sequencing and assembly

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

Assembly scaffolding with HiRise

A proximity ligation library was generated by the Dovetail's Omni-C technique, followed by sequencing on an Illumina HiSeqX platform. Chromatin was fixed in place in the nucleus with formaldehyde before extraction (https://dovetailgenomics.com/wp-content/uploads/2021/09/Omni-C-Tech-Note.pdf). Fixed chromatin was digested with DNase I, fragmented chromatin ends were repaired and biotinylated to adapters followed by proximity ligation. Crosslinks were then reversed, the DNA purified, and the biotin subsequently

removed. The DNA library was prepared and sequenced to produce 2 × 150 bp paired-end reads (around 30× coverage). The Omni-C technology uses a sequence-independent endonuclease which provides even, unbiased genome 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 (56). The Omni-C library sequences were aligned to the contigs assembled by Hifiasm using bwa (57). The separations of Omni-C reads from paired-end reads 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 (25) using *k*-mer frequency plot obtained by kmc v3.2.4 (58) 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 (59). The genome size of *L. edulis* was estimated by KmerGenie (24) 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 (20), 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

529 RNA was extracted independently from bud and leaf samples of each tree. Total
530 RNA of the leaf sample was extracted in accordance with the method described in
531 a previous study (60). Total RNA of the bud sample was extracted using the
532 PureLink™ Plant RNA Reagent (Thermo Fisher Scientific, Waltham, MA, USA)
533 and subsequently processed with DNase using the TURBO DNA-free kit™
534 (Thermo Fisher Scientific) according to their protocols and purified using the
535 RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to its RNA cleanup
536 protocol. RNA integrity was examined using the Agilent RNA 6000 Nano kit on a
537 2100 Bioanalyzer (Agilent Technologies), and the RNA yield was determined using
538 a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific). Five to six
539 micrograms of RNA extracted from each sample were sent to Hangzhou Veritas
540 Genetics Medical Institute Co., Ltd., where a cDNA library was prepared with a
541 NEBNext Ultra II RNA Library Prep Kit for Illumina, and 150 base paired-end
542 transcriptome sequencing of each sample was conducted using an Illumina
543 NovaSeq6000 sequencer (Illumina).
544 The quantification of gene expression followed a three-step process. First,
545 quality filtering was performed using the fastp v0.23.2 with default options: phred
546 quality ≥15, the maximum number of N base in one read = 5 (59). Second, RNA-
547 seq reads were mapped to the reference genome using STAR v2.7.10b (61). Third,
548 RNA-seq reads were quantified using RSEM v1.3.1 (62). We excluded the genes
549 with low expression level using the threshold, the value of mean Reads Per
550 Kilobase (RPK) < 1 for each species. Then, we further normalized the expression
551 level using the value of Gene length corrected Trimmed Mean of M-values
552 (GeTMM) (27) using an R package "edgeR" v3.42.4. For the downstream analysis
553 we calculated the mean expression levels of GeTMM of each gene across multiple
554 individuals and transformed them to $\log_2$(GeTMM+1).
555
556 **Genome annotation**
557 The RNA-seq reads of *Q. glauca* (8 samples obtained by a previous study (20)
558 collected from May to December 2017; GEO's accession number: GSE211384)
559 and *L. edulis* (20 samples covering leaf, bud and flower tissues collected in
560 January, April, July, and November 2021) were respectively mapped against the
561 scaffolds of *Q. glauca* and *L. edulis* obtained by HiRise with HISAT2 v2.1.0 (63).
562 The bam files generated by the mappings and amino acid sequences of the 9
563 species obtained from NCBI (*Castanea mollissima*: GCA_000763605.2 (NCBI's
564 accession number), *Q. lobata*: GCF_001633185.1, *Q. robur*: GCF_932294415.1,
565 *Q. suber*: GCF_002906115.1, *Carpinus fangiana*: GCA_006937295.1, *Carya*
566 *illinoinensis*: GCF_018687715.1, *Juglans macrocarpa* × *J. regia*:
567 GCF_004785595.1, *Juglans regia*: GCF_001411555.2, *Morella rubra*:
568 GCA_003952965.2) were used for ab initio gene prediction by Braker2 v2.1.6 (64),
569 and the genes with highest-score were selected to exclude splicing variants. In
570 parallel, the gene models were respectively constructed in *Q. glauca* and *L. edulis*
571 by GeMoMa v1.7.1 (65) with RNA-seq reads and amino acid sequences used for
572 the gene prediction in Braker2, and the genes with highest-score were selected to
573 exclude splicing variants. The duplicated genes between Braker2 and GeMoMa

were excluded by GFFcompare v0.11.2 (66). The remaining genes were searched against UniProtKB (67) by DIAMOND v2.0.5 (68) in 'more-sensitive' option with E-value ≤10^{-20} and identity ≥25%. Next, gene expression analysis was performed by Salmon v1.4.0 (69), and the genes with TPM values >0.1 were considered as expressed. The orthologous genes were searched against EggNOG (70) using eggNOG-mapper v2.1.8 (71) 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 (72)) by BLASTp v2.13.0 (73) 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 (74) 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 (75). 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 (76) 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 (77) and included as an outgroup. Due to the large volume of short-read data for *Q. acuta* and *L. glaber* (18 Gb each), we randomly

subsampled half of the reads. Bootstrap values were estimated through 100 resampling of the k-mer table.

Identification of orthologs

The prediction of orthogroups was based on a BLASTp all-against-all comparison of the protein sequences (E-value < 10^{-5}) of 5 Fagaceae species: *L. edulis*, *Q. glauca*, *Q. robur*, *Fagus sylvatica* and *Juglans regia*. Clustering was conducted with OrthoFinder2, using version 2.2.6 for BLAST data generation and version 2.5.5 for orthogroup determination (26). The genomic data of *Q. robur* and *J. regia* were downloaded from NCBI (dhQueRobu3.1 and Walnut 2.0), while *F. sylvatica* data were sourced from beechgenome.net (77). Since the *F. sylvatica* dataset contained numerous predicted transposable elements (TEs), we re-annotated genes using three databases—EggNOG5 (EggNOG-mapper), Pfam35.0 (HMMER v3.3.2), and UniRef90_2023_0628 (DIAMOND v2.1.8.162)—and excluded genes identified as TEs in at least one database. In total, we identified 26692 orthogroups and focused on single-copy orthogroups after filtering out genes mapped outside the 12 chromosomes. This resulted in a final dataset of 11749 genes for further analysis (Dataset S2).

Hierarchical clustering and principal component analysis (PCA)

To assess the similarity of the genome-wide transcriptional profiles across orthologous genes and 483 samples collected in different seasons, we performed hierarchical clustering using the monthly time series data of 11749 orthologous genes from October 2019 to April 2023. For each orthologous gene, there were 26 or 27 time points, with three individuals each for *Q. glauca*, *Q. acuta*, *L. glaber* and one individual for *L. edulis*. We calculated the mean expression levels of each orthologous gene across three individuals in each species and subsequently normalized the values by adjusting the mean to zero and the standard deviation to one. We performed hierarchical clustering using the Ward method and the Euclidian distance using the pheatmap function in R (v 1.0.12). To determine the number of the clusters, we calculated Within-Cluster-Sum of Squared Error (WSS) and determined the number of clusters as an elbow point of it (Fig. S3). To assess the seasonal expression dynamics of 11749 orthologous genes, we also performed PCA of the gene expression profiles from all samples using the prcomp function in R (v 4.3.1).

Identification of genes with rhythmic expression

To identify genes with rhythmic expression and determine the period and peak month of gene expression, we applied the RAIN algorithm (29) using the rain function in R (v 1.34.0). We set the following settings: deltat = 1, period = 12, period.delta = 11, peak.border = (0,1), nr.series = 1, method= "independent", na.rm= T , adjp.method = "ABH". Based on the results of the RAIN analysis, we identified genes with significant periodicity using a 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 (78) 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)(79). 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.

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.

References

1. SD. Fretwell, Populations in a seasonal environment. *Monogr. Popul. Biol.* 5, 1–217 (1972).

2. J. Forrest, A. J. Miller-Rushing, Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philos. Trans. R. Soc.* 365, 3101–3112 (2010).

3. L. P. C. Morellato, M. G. G. Camargo, E. Gressler, “A review of plant phenology in South and Central America.” in Phenology: An Integrative Environmental Science, M. D. Schwartz Eds. (Springer, 2013), 91-113.

751 4. M. D. Schwartz (Eds.), Phenology: An Integrative Environmental Science, 2nd Ed.
752 (Springer 2013).

753 5. S. Piao, *et al.*, Plant phenology and global climate change: Current progresses and
754 challenges. *Glob. Chang. Biol.* 25, 1922–1940 (2019).

755 6. H. Kudoh, Molecular phenology in plants: *in natura* systems biology for the
756 comprehensive understanding of seasonal responses under natural environments.
757 *New Phytol.* 210, 399–412 (2016).

758 7. A. Satake, A. Nagahama, E. Sasaki, A cross-scale approach to unravel the
759 molecular basis of plant phenology in temperate and tropical climates. *New Phytol.*
760 233, 2340–2353 (2022).

761 8. A. Satake, *et al.*, Plant molecular phenology and climate feedbacks mediated by
762 BVOCs. *Annu. Rev. Plant Biol.* 75, 605–627 (2024).

763 9. D. Brawand, *et al.*, The evolution of gene expression levels in mammalian organs.
764 *Nature* 478, 343–348 (2011).

765 10. S. Naqvi, *et al.*, Conservation, acquisition, and functional impact of sex-biased gene
766 expression in mammals. *Science* 365 (2019).

767 11. M. Cardoso-Moreira, *et al.*, Gene expression across mammalian organ
768 development. *Nature* 571, 505–509 (2019).

769 12. S. V. Nuzhdin, M. L. Wayne, K. L. Harmon, L. M. McIntyre, Common pattern of
770 evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol.*
771 *Evol.* 21, 1308–1317 (2004).

772 13. B. Lemos, B. R. Bettencourt, C. D. Meiklejohn, D. L. Hartl, Evolution of proteins and
773 gene expression levels are coupled in *Drosophila* and are independently associated
774 with mRNA abundance, protein length, and number of protein-protein interactions.
775 *Mol. Biol. Evol.* 22, 1345–1354 (2005).

776 14. A. El Taher, *et al.*, Gene expression dynamics during rapid organismal
777 diversification in African cichlid fishes. *Nat. Ecol. Evol.* 5, 243–250 (2021).

778 15. A. Kremer, *et al.*, Genomics of Fagaceae. *Tree Genet. Genomes* 8, 583–610 (2012).

779 16. B. F. Zhou, *et al.*, Phylogenomic analyses highlight innovation and introgression in
780 the continental radiations of Fagaceae across the Northern Hemisphere. *Nat.*
781 *Commun.* 13 (2022).

782 17. P. S. Manos, Z. Zhou, C. H. Cannon, Systematics of Fagaceae: phylogenetic tests
783 of reproductive trait evolution. *Int. J. Plant Sci.* 162, 1361–1379 (2001).

784 18. A. L. Hipp, *et al.*, Genomic landscape of the global oak phylogeny. *New Phytol.* 226,
785 1198–1212 (2020).

786 19. H. Komoto, *et al.*, The transcriptional changes underlying the flowering phenology
787 shift of Arabidopsis halleri in response to climate warming. *Plant Cell Environ.* 47,
788 174–186 (2024).

789 20. A. Satake, K. Ohta, N. Takeda-Kamiya, K. Toyooka, J. Kusumi, Seasonal gene
790 expression signatures of delayed fertilization in Fagaceae. *Mol. Ecol.* 32, 4801–
791 4813 (2023).

792 21. A. J. Nagano, *et al.*, Annual transcriptome dynamics in natural environments
793 reveals plant seasonal adaptation. *Nat. Plants* 5, 74–83 (2019).

794 22. V. Zoldos, *et al.*, Molecular-cytogenetic studies of ribosomal genes and
795 heterochromatin reveal conserved genome organization among 11 Quercus
796 species. *Theor. Appl. Genet.* 99, 969–977 (1999).

797 23. H. Liu, *et al.*, Chromosome-scale genome assembly of sweet tea (Lithocarpus
798 polystachyus Rehder). *Sci Data* 10 (2023).

799 24. R. Chikhi, P. Medvedev, Informed and automated k-mer size selection for genome
800 assembly. *Bioinformatics* 30, 31–37 (2014).

801 25. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and
802 Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11
803 (2020).

804 26. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for
805 comparative genomics. *Genome Biol.* 20 (2019).

806 27. M. Smid, *et al.*, Gene length corrected trimmed mean of M-values (GeTMM)
807 processing of RNA-seq data performs similarly in intersample analyses while
808 improving intrasample comparisons. *BMC Bioinformatics* 19 (2018).

809 28. J. Zrimec, F. Buric, M. Kokina, V. Garcia, A. Zelezniak, Learning the regulatory code
810 of gene expression. *Front. Mol. Biosci.* 8 (2021).

811 29. P. F. Thaben, P. O. Westermark, Detecting rhythms in time series with rain. *J. Biol.*
812 *Rhythms* 29, 391–400 (2014).

813 30. G. Q. Zhu, L. Qu, H. W. Xue, Casein kinase 1 AELs promote senescence by
814 enhancing ethylene biosynthesis through phosphorylating WRKY22 transcription
815 factor. *New Phytol.* 244, 116–130 (2024).

816 31. S. Emanuelle, M. S. Doblin, D. I. Stapleton, A. Bacic, P. R. Gooley, Molecular
817 insights into the enigmatic metabolic regulator, SnRK1. *Trends Plant Sci.* 21, 341–
818 353 (2016).

819 32. P. Kant, S. Kant, M. Gordon, R. Shaked, S. Barak, Stress Response Suppressor1
820 and Stress Response Suppressor2, two Dead-box RNA helicases that attenuate
821 *Arabidopsis* responses to multiple abiotic stresses. *Plant Physiol.* 145, 814–830
822 (2007).

33. L. T. Chen, K. Wu, Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response. *Plant Signal. Behav.* 5, 1318–1320 (2010).

34. S. J. Yoo, *et al.*, BROTHER of FT and TFL1 (BFT) has TFL1-like activity and functions redundantly with TFL1 in inflorescence meristem development in *Arabidopsis*. *Plant J.* 63, 241–253 (2010).

35. C. Smaczniak, *et al.*, Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1560–1565 (2012).

36. D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, F. H. Arnold, Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14338–14343 (2005).

37. B. Y. Liao, J. Zhang, Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol. Biol. Evol.* 23, 1119–1128 (2006).

38. B. S. Gaut, The patterns and causes of variation in plant nucleotide substitution rates. *Annu. Rev. Ecol. Evol. Syst.* 42, 245-266 (2011).

39. L. Yang, B. S. Gaut, Factors that contribute to variation in evolutionary rate among *Arabidopsis* genes. *Mol. Biol. Evol.* 28, 2359–2369 (2011).

40. A. Miyawaki-Kuwakado, Q. Han, K. Kitamura, A. Satake, Impacts of climate change on the transcriptional dynamics and timing of bud dormancy release in Yoshino-cherry tree. *Plants People Planet* 6, 1505–1521 (2024).

41. S. Aikawa, M. J. Kobayashi, A. Satake, K. K. Shimizu, H. Kudoh, 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 (2010).

42. N. Irie, S. Kuratani, Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* 2 (2011).

43. M. Quint, *et al.*, A transcriptomic hourglass in plant embryogenesis. *Nature* 490, 98–101 (2012).

44. J. S. Lotharukpong, *et al.*, A transcriptomic hourglass in brown algae. *Nature* 635, 129–135 (2024).

45. A. T. Kalinka, *et al.*, Gene expression divergence recapitulates the developmental hourglass model. *Nature* 468, 811–816 (2010).

46. H. G. Drost, *et al.*, Post-embryonic hourglass patterns mark ontogenetic transitions in plant development. *Mol. Biol. Evol.* 33, 1158–1163 (2016).

47. M. Levin, *et al.*, The mid-developmental transition and the evolution of animal body plans. *Nature* 531, 637–641 (2016).

859 48. T. Kubo, Y. Iwasa, Phenological pattern of tree regeneration in a model for forest
860 species diversity. *Theor. Popul. Biol.* 49, 90-117 (1996).

861 49. E. E. Cleland, E. M. Wolkovich, Effects of phenology on plant community assembly
862 and structure. *Annu. Rev Ecol. Evol. Syst.* 55, 471-492 (2025).

863 50. H. Nishio, A. J. Nagano, T. Ito, Y. Suzuki, H. Kudoh, Seasonal plasticity and diel
864 stability of H3K27me3 in natural fluctuating environments. *Nat. Plants* 6, 1091–1097
865 (2020).

866 51. H. Fischl, *et al.*, Cold-induced chromatin compaction and nuclear retention of clock
867 mRNAs resets the circadian rhythm. *EMBO J.* 39 (2020).

868 52. X. Zhang, *et al.*, High-resolution Hi-C maps highlight multiscale chromatin
869 architecture reorganization during cold stress in *Brachypodium distachyon*. *BMC*
870 *Plant Biol.* 23 (2023).

871 53. J. R. True, E. S. Haag, Developmental system drift and flexibility in evolutionary
872 trajectories. *Evol. Dev.* 3, 109–119 (2001).

873 54. B. Prud'homme, N. Gompel, S. B. Carroll, Emerging principles of regulatory
874 evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8605–8612 (2007).

875 55. J. Zrimec, *et al.*, Deep learning suggests that gene expression is encoded in all
876 parts of a co-evolving interacting gene regulatory structure. *Nat. Commun.* 11
877 (2020).

878 56. N. H. Putnam, *et al.*, Chromosome-scale shotgun assembly using an in vitro method
879 for long-range linkage. *Genome Res.* 26, 342–350 (2016).

880 57. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler
881 transform. *Bioinformatics* 25, 1754–1760 (2009).

882 58. M. Kokot, M. Dlugosz, S. Deorowicz, KMC 3: counting and manipulating k-mer
883 statistics. *Bioinformatics* 33, 2759–2761 (2017).

884 59. S. Chen, Y. Zhou, Y. Chen, J. Gu, "Fastp: An ultra-fast all-in-one FASTQ
885 preprocessor" in *Bioinformatics*, (Oxford University Press, 2018), pp. i884–i890.

886 60. Y. Miyazaki, *et al.*, Nitrogen as a key regulator of flowering in *Fagus crenata*:
887 understanding the physiological mechanism of masting by gene expression
888 analysis. *Ecol. Lett.* 17, 1299–1309 (2014).

889 61. A. Dobin, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–
890 21 (2013).

891 62. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data
892 with or without a reference genome. *BMC Bioinformatics* 12 (2011).

893 63. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome
894 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37,
895 907–915 (2019).

896 64. T. Brůna, K. J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2:
897 Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS
898 supported by a protein database. *NAR Genom. Bioinform.* 3, 1–11 (2021).

899 65. J. Keilwagen, F. Hartung, J. Grau, "GeMoMa: Homology-based gene prediction
900 utilizing intron position conservation and RNA-seq data" in *Methods in Molecular*
901 *Biology*, (Humana Press Inc., 2019), pp. 161–177.

902 66. M. Pertea, G. Pertea, GFF Utilities: GffRead and GffCompare. *F1000Res.* 9 (2020).

903 67. A. Bateman, *et al.*, UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic*
904 *Acids Res.* 53, D609–D617 (2025).

905 68. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using
906 DIAMOND. *Nat. Methods* 12, 59–60 (2014).

907 69. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast
908 and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419
909 (2017).

910 70. A. Hernández-Plaza, *et al.*, eggNOG 6.0: enabling comparative genomics across
911 12 535 organisms. *Nucleic Acids Res.* 51, D389–D394 (2023).

912 71. C. P. Cantalapiedra, A. Herṇandez-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas,
913 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
914 Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* 38, 5825–5829 (2021).

915 72. V. Krishnakumar, *et al.*, Araport: The Arabidopsis information portal. *Nucleic Acids*
916 *Res.* 43, D1003–D1009 (2015).

917 73. C. Camacho, *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* 10
918 (2009).

919 74. Y. Wang, *et al.*, MCScanX: A toolkit for detection and evolutionary analysis of gene
920 synteny and collinearity. *Nucleic Acids Res.* 40 (2012).

921 75. A. Satake, *et al.*, Somatic mutation rates scale with time not growth rate in long-
922 lived tropical trees. *Elife* 12 (2024).

923 76. H. Fan, A. R. Ives, Y. Surget-Groba, C. H. Cannon, An assembly and alignment-
924 free method of phylogeny reconstruction from next-generation sequencing data.
925 *BMC Genomics* 16 (2015).

926 77. B. Mishra, *et al.*, A chromosome-Level genome assembly of the European beech
927 (*Fagus sylvatica*) reveals anomalies for organelle DNA integration, repeat Content
928 and distribution of SNPs. *Front. Genet.* 12 (2022).

929 78. S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association.
930 *Bioinformatics* 23, 257–258 (2007).

931 79. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24,
932 1586–1591 (2007).

933
934

935 **Figures and Tables**

936 **Figure 1.** Synteny and distribution of genomic features of *Q. glauca* and *L. edulis*
937 genomes. (*A*) A phylogenetic tree of five Fagaceae species from three genera,
938 *Quercus*, *Lithocarpus*, and *Fagus*. Bootstrap values for all branches are 100. (*B*)
939 Synteny relationship between the *Q. glauca* and *L. edulis* genomes. The collinear
940 blocks within the genomes were displayed by lines. Black lines show reverse
941 alignment and gray lines represent regular. Numbers correspond to the 12
942 chromosomes. (*C*–*D*) Circos plots of the *Q. glauca* (*C*) and *L. edulis* genomes (*D*).
943 Lanes depict circular representation of chromosome length (Mb) (a), gene density
944 (b), GC density (c), and repeat (d). Lines in the inner circle represent links between
945 synteny-selected paralogs. (*E*) Gene copy-number variations between the *Q.*
946 *glauca* and *L. edulis* genomes. The bar plot illustrates the number of genes in gene
947 families categorized into 1:1 orthologs, N:N (many-to-many), 1:N (one-to-many),
948 N:1 (many-to-one), and species-specific unique genes. (F–G) Mean gene
949 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).

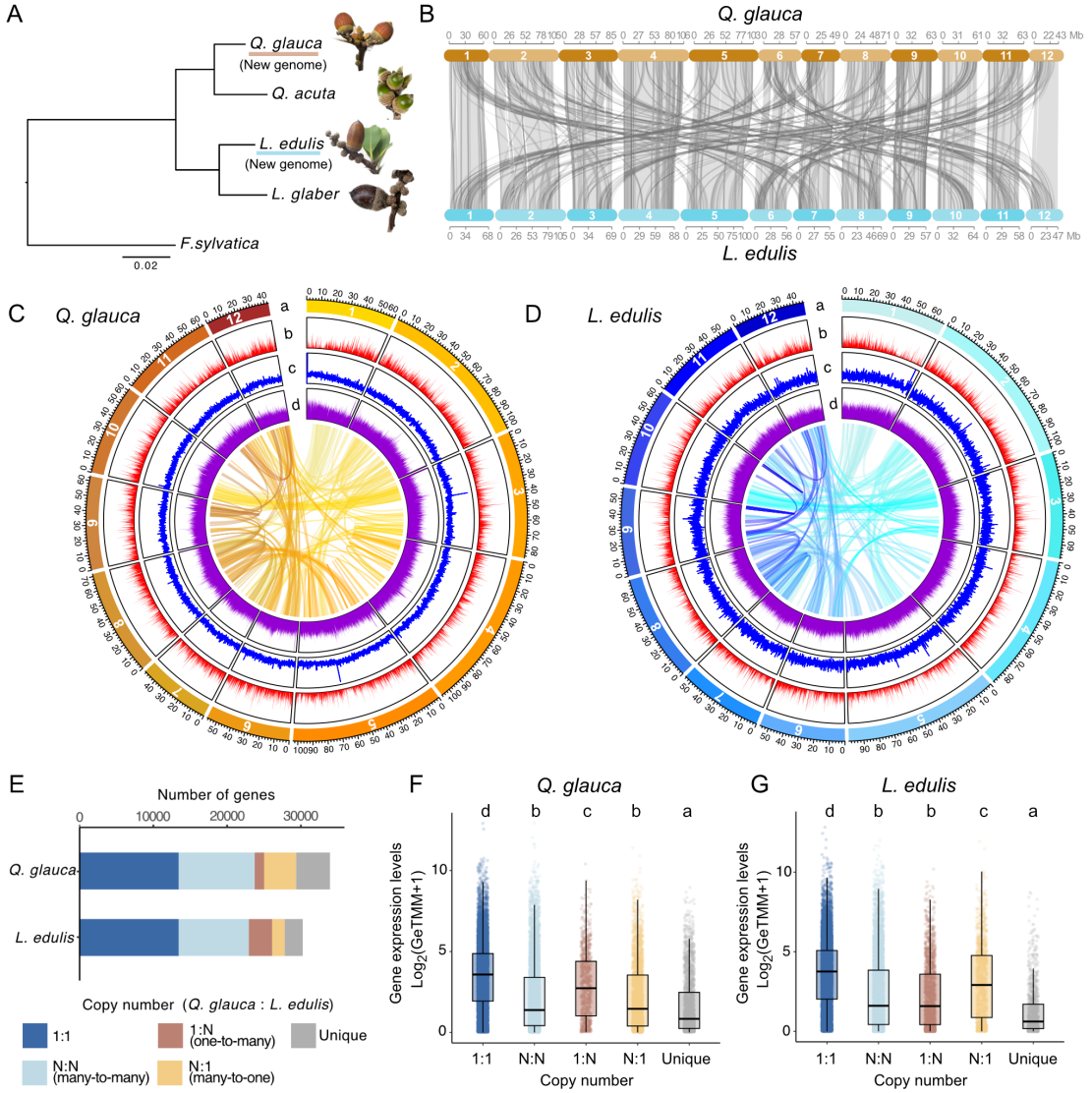

Figure 2. Seasonal gene expression dynamics. (*A*) Hierarchical clustering of 213 gene expression profiles encompassing two tissues (leaf and bud) and four species (*Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber*), collected over two years. (*B*) Distribution of daily mean temperature on each sampling 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.

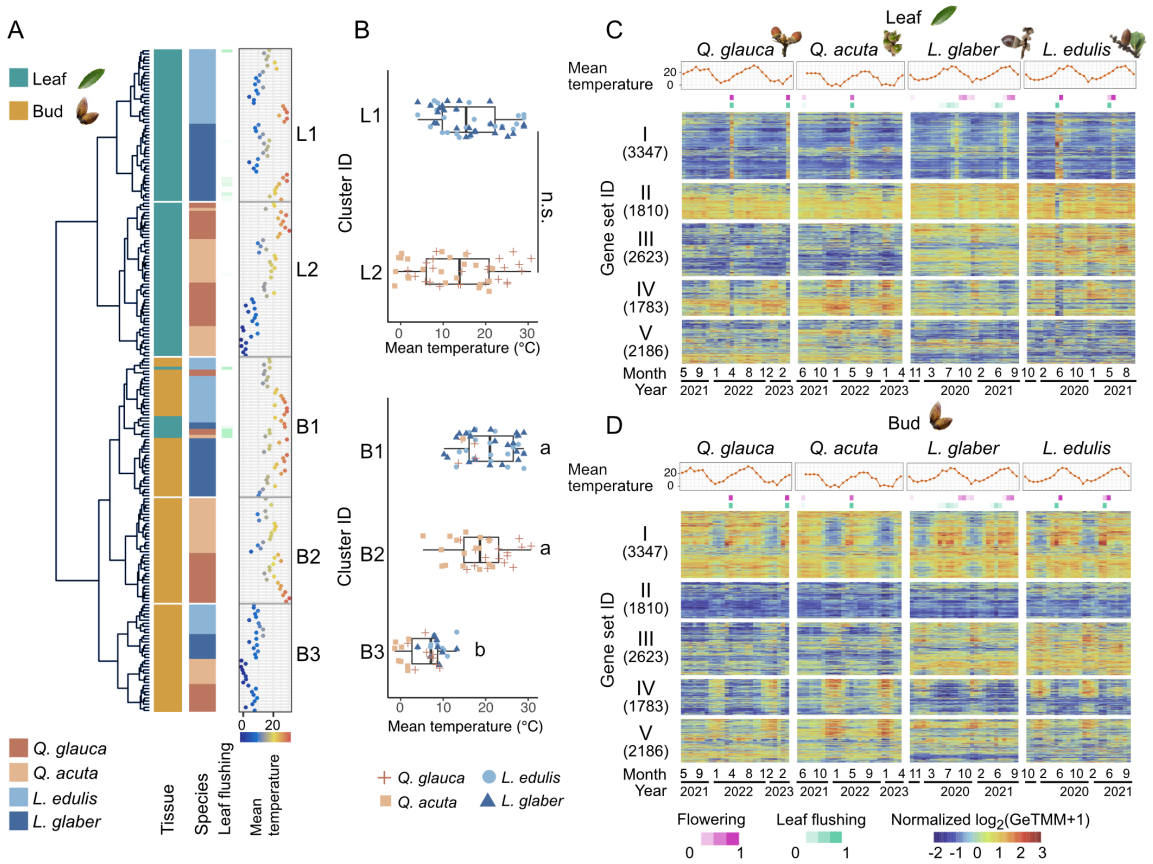

Figure 3. Distribution of period and peak month in rhythmic gene expression. (*A*–*B*) Distribution of period length (month) among the genes with significant rhythmicity calculated by RAIN for leaf (*A*) and bud (*B*) samples of each species. (*C*–*D*) Distribution of seasonal peaks among the genes with 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.

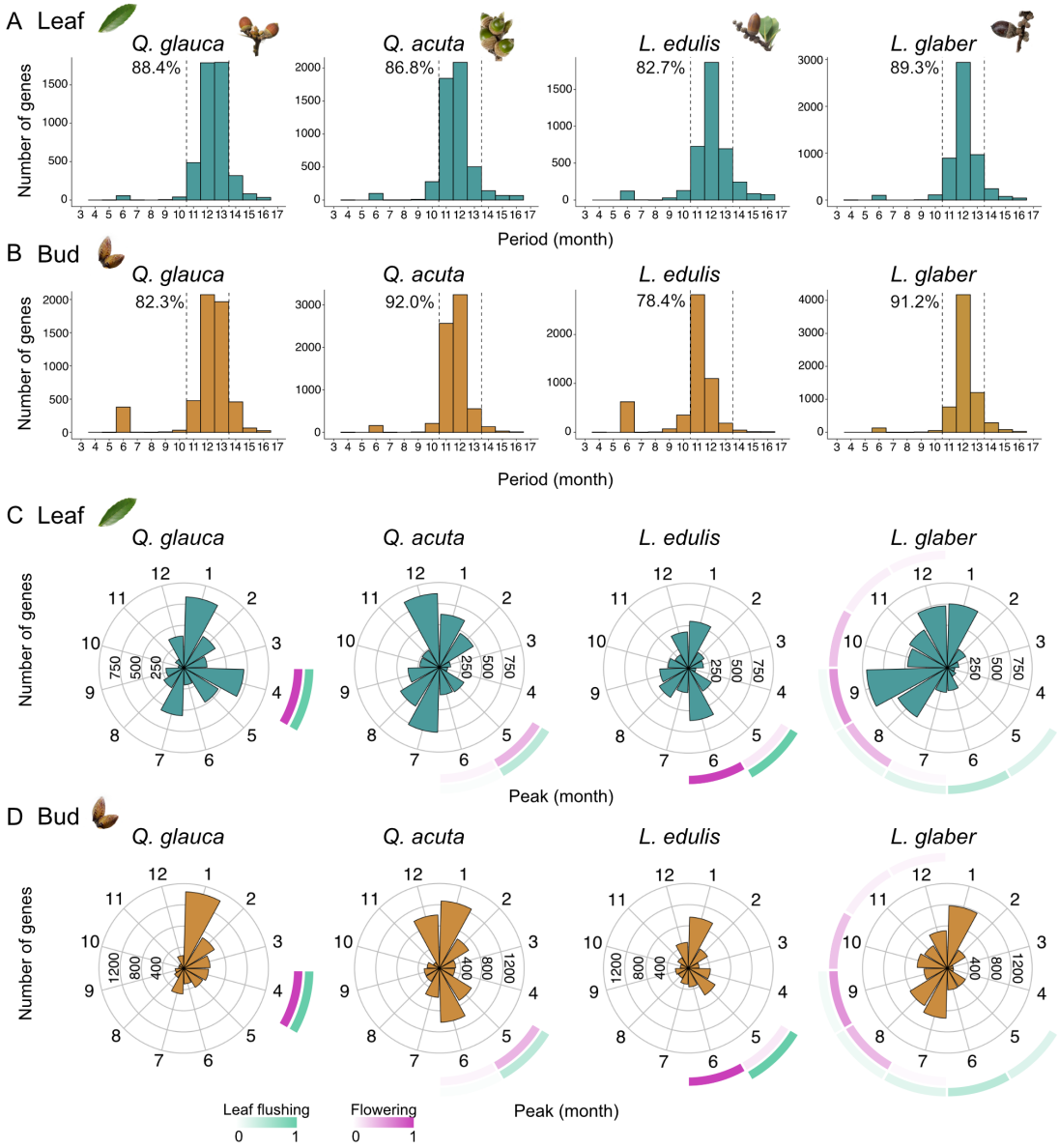

Figure 4. Comparison of peak months in seasonal gene expression across species in buds. (*A*) Intra-genus and (*B*) inter-genera comparisons of peak months in seasonal gene expression. The size and color of the circles represent the number of genes and the mean of Pearson's correlation, respectively. 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, $p < 0.05$).

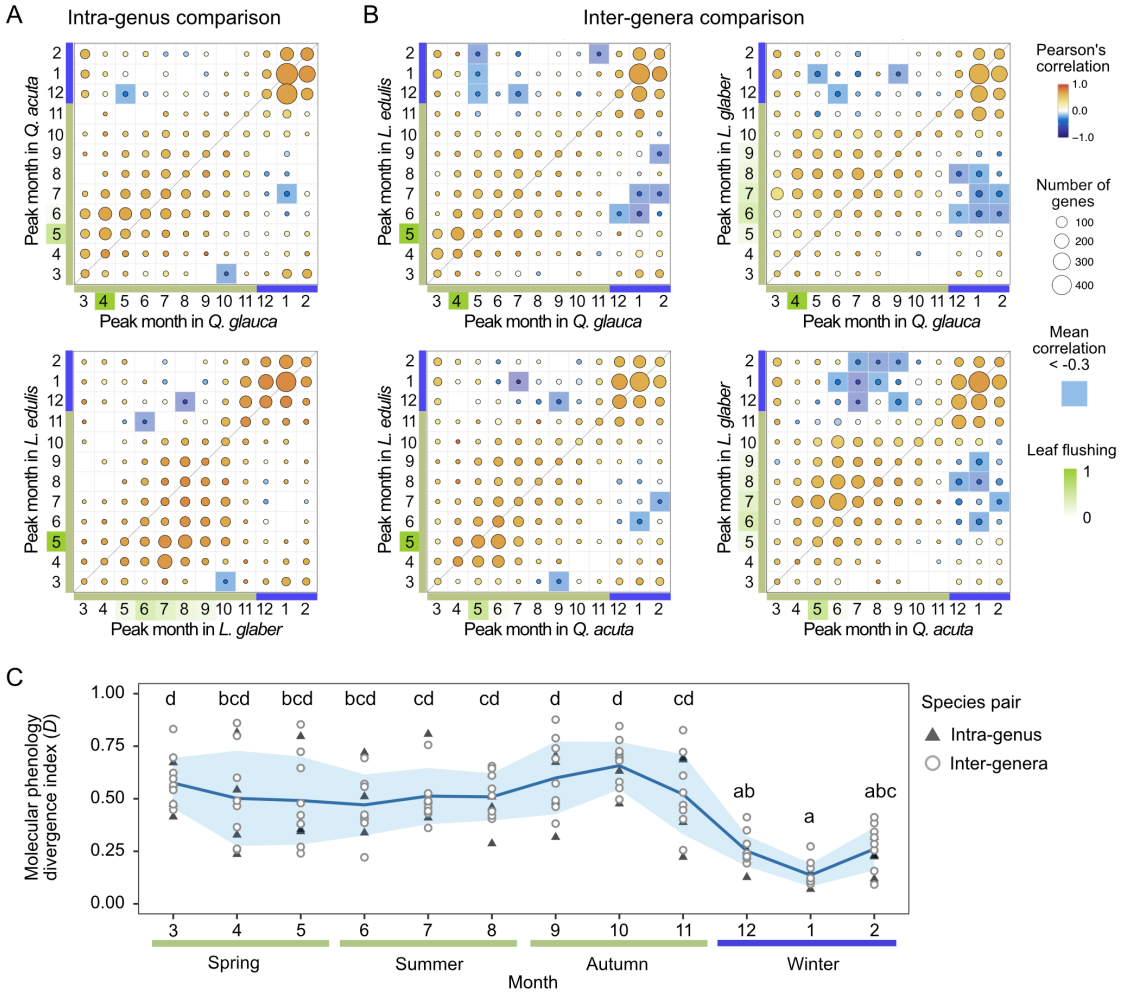
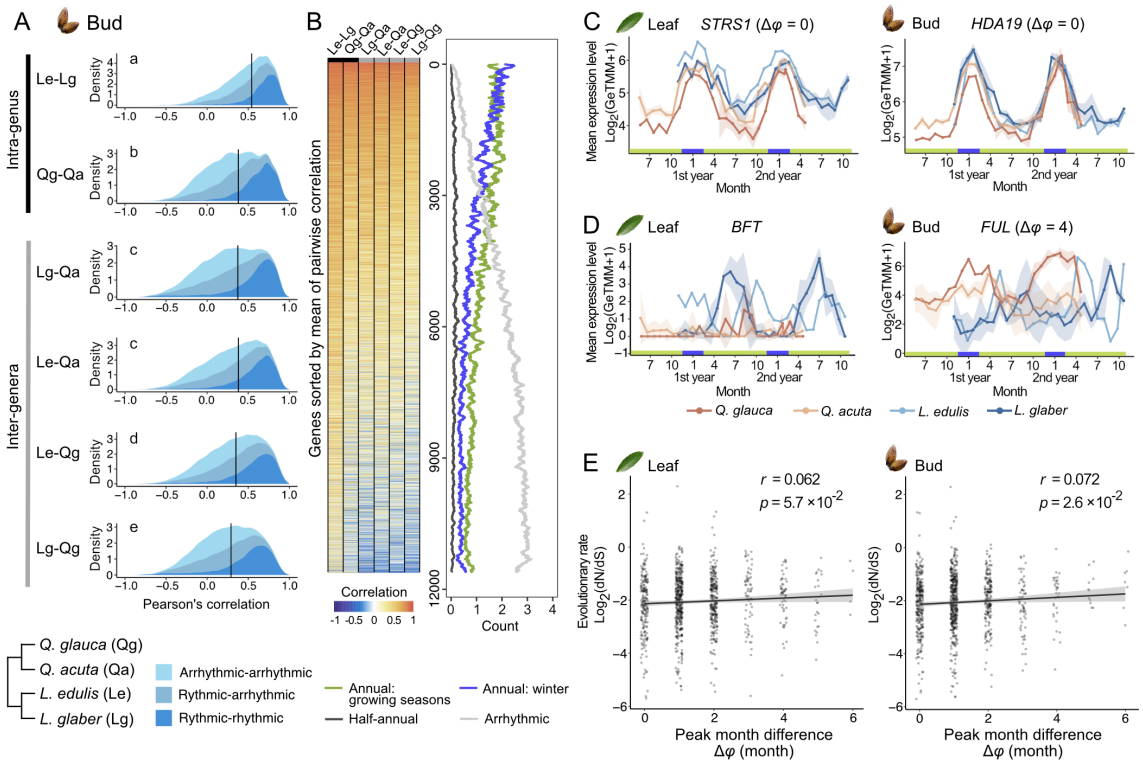
Figure 5. Phylogenetic constraints in the evolution of seasonal gene expression and the relationship between seasonal gene expression divergence and sequence evolution. (*A*) Distribution of pairwise Pearson's correlation coefficients for gene expression across species (Le: *L. edulis*, Lg: *L. glaber*, 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.