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.

Impact Statement

A comparative analysis of genome-wide seasonal gene expression dynamics across four forest tree species revealed that a shared "molecular calendar" emerges in winter, constraining gene expression evolution and potentially limiting temporal niche partitioning and species divergence 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 (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 and phenological observation of leaf flushing and flowering in 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* (Figure 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 (Supplementary file 1).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 (Supplementary file 1). 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 (Figure 1C, D*;* Supplementary file 1) and syntenic structure (Figure 1B) were similar between *Q. glauca* and *L. edulis*. This pattern of synteny conservation appears to be broadly shared across Fagaceae, except for the genus *Fagus*, which shows a distinct genome structure (Ikezaki et al., 2025). The homology between different chromosomes within the genome suggests that small-scale segmental duplications likely occurred, as previously proposed (Chen et al., 2014). 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 (Figure 1E).

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 92.6 ± 2.2% and 88.6 ± 2.3% for *Q. glauca* and *Q. acuta*, respectively. Similarly, reads from the two *Lithocarpus* species were mapped to the newly constructed *L. edulis* reference genome, with a mean mapping rate of 89.3 ± 2.7% and 84.3 ± 5.4% for *L. edulis* and *L. glaber*, respectively. The number of reads uniquely mapped on the reference genome was 19.9 million reads per sample, on average. We used single-copy orthologous genes for further analyses because these genes exhibited significantly higher average expression levels than multi-copy genes (Steel-Dwass test, $p < 0.05$; Figure 1F, G). 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 (Figure 2–figure supplement 1A). 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. We confirmed that expression patterns of individual genes were highly consistent across all the three biological replicates (Figure 2–figure supplement 2) in *Q. glauca* (Pearson's correlation r = 0.947), *Q. acuta* (r = 0.948) and *L. glaber* (r = 0.948).

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) (Figure 2–figure supplement 1B). 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 (Zrimec et al., 2021). Consistently, our results demonstrate that despite substantial seasonal fluctuations in environmental conditions (Figure 2–figure supplement 3), 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 (Figure 2–figure supplement 1B).

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 (Figure 2A; Figure 2–figure supplement 4A; Supplementary file 2). Transcriptional profiles first grouped into leaves (cluster L1 and L2) and buds (cluster B1–B3), irrespective of seasons and species (Figure 2A), except that newly flushed leaf samples were grouped closely with buds (Figure 2A; Figure 2–figure supplement 5). Within the leaf cluster, transcriptome profiles were first clustered by genus and then by species in *Lithocarpus* (cluster L1; Figure 2A). In contrast, samples from *Quercus* species formed a mixture of two species (cluster L2; Figure 2A). 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; Figure 2A). 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*; Supplementary file 2) versus the other growing seasons (cluster B1 and B2; March to November; Supplementary file 2)—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. We confirmed the robustness of our findings to differences in sampling periods between *Quercus* and *Lithocarpus* species by reanalyzing expression data limited to common sampling periods (Figure 2–figure supplement 6). Even when using only the data from this shared period, the clustering patterns were consistently reproduced. To assess whether the higher-elevation site of *Q. acuta* introduced confounding environmental effects, we reanalyzed the data after excluding this

species. Hierarchical clustering still revealed that winter bud samples formed a distinct cluster regardless of species identity (Figure 2–figure supplement 7), consistent with our original findings.

To assess the biological relevance of the clusters identified, we tested whether the temperatures and photoperiod at the time of sample collection differed significantly among the clusters. Within the bud cluster, transcriptional profiles in the winter cluster (cluster B3) were associated with significantly lower temperatures—approximately below 10 °C—compared to those in other seasons (Figure 2B; Figure 2–figure supplement 8A). In contrast, no such temperature-related distinction was evident within the leaf cluster (Figure 2B). Additionally, photoperiods were markedly shorter in the winter cluster in buds (Figure 2–figure supplement 8B). These findings suggest that the buds of the four species exhibit a conserved transcriptional response to environmental cues, namely cold temperatures and short photoperiods, 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 (Figure 2–figure supplement 9).

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; Figure 2C, D; Figure 2–figure supplement 4B; Supplementary file 3). 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 (Figure 2C, D). Gene sets III and V demonstrated genus-specific expression patterns, with higher expression levels observed in *Lithocarpus* and *Quercus*, respectively (Figure 2C, D). In contrast, gene set IV was expressed during winter across both leaf and bud tissues (Figure 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 (Figure 3–figure supplement 1A; Supplementary file 4). 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 (Figure 3–figure supplement 1A).

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 (Figure 3A, B, Supplementary file 4). 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* (Figure 3A, B). The fraction of genes exhibiting other periodicities was relatively small (5.7–14.1%; Figure 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 (Figure 3–figure supplement 1A). 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 (Figure 3–figure supplement 1B). In contrast, most genes exhibiting half-annual periodicity were species-specific, with no genes shared among all the species Figure 3–figure supplement 1C).

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 (Figure 3C, D; Supplementary file 4). 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 (Figure 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 (Figure 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 (Figure 3–figure supplement 1D; Supplementary file 4), 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* (Figure 3–figure supplement 1*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) (Figure 3–figure supplement 1E). 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 (Figure 4A) and inter-genera species pairs in buds (Figure 4B) and leaves (Figure 4–figure supplement 1A, B). During winter, a large number of genes exhibited highly synchronized expression peaks across species, with high pairwise Pearson's correlation coefficients observed (Figure 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 (Figure 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; Figure 4B) than in intra-genus comparisons (4–12 genes; Figure 4A). A similar pattern was observed in leaf samples (Figure 4–figure supplement 1A, B). When comparing seasonal expression peaks between tissues, the majority of genes were synchronously activated in winter in both tissues (Figure 4–figure supplement 1C). During spring to summer, slight variations in seasonal expression peaks were observed among tissues (Figure 4–figure supplement 1C).

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, $p < 0.05$; Figure 4C). A similar trend was observed in leaf samples, with another trough in summer (Figure 4–figure supplement 1D). This seasonal trend remained consistent even after excluding *Q. acuta*, which was sampled from a slightly cooler environment than the other three species, from the analysis (Figure 4–figure supplement 2), indicating that the observed patterns are not driven by environmental differences associated with elevation. 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 (Supplementary file 5). 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, $p < 0.05$; Figure 5–figure supplement 1A; Figure 5–figure supplement 2) and buds (Steel-Dwass test, $p < 0.05$; Figure 5A; Figure 5–figure supplement 2). 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 (Figure 5*A*; Figure 5–figure supplement 1A). In particular, genes with seasonal expression peaks in winter showed high correlation across species in both tissues (Figure 5B; Figure 5–figure supplement 1B). 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 (Figure 5B; Figure 5–figure supplement 1B). 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 (Figure 5–figure supplement 3A, 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 (Figure 5–figure supplement 3C, 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 (Figure 5B; Figure 5–figure supplement 1B) 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 (Figure 5–figure supplement 4A; Supplementary file 6). The gene with the highest correlation in leaves was *STRESS RESPONSE SUPPRESSOR1* (*STRS1*; mean of pairwise correlation = 0.91; Figure 5C), a DEAD-box RNA helicase previously reported to enhance tolerance to abiotic stresses such as cold (Kant et al., 2007). In buds, the most highly correlated gene was the histone deacetylase *HDA19* (mean of pairwise correlation = 0.93; Figure 5C), which plays a key role in environmental stress responses through histone modification and chromatin remodeling (Chen and Wu, 2010). 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 (Figure 5–figure supplement 4B), aligning with the divergent leaf phenology observed (Figure 3B). In contrast, GO terms related to immune response were weakly enriched in buds (Figure 5–figure supplement 4B). 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 (Yoo et al., 2010), was expressed in autumn in the spring-flowering species *L. edulis* and in spring in the

379 autumn-flowering species *L. glaber* leaves (Figure 5D). Meanwhile, *FUL*, a positive
380 regulator of floral transition (Smaczniak et al., 2012), was expressed in spring in *L.*
381 *edulis* and in autumn in *L. glaber* (Figure 5D), mirroring the timing of flowering. In
382 *Q. glauca* and *Q. acuta*, *FUL* was expressed in winter (Figure 5D), nearly out of
383 phase with their expressions in *Lithocarpus* species. These results suggest that
384 genes involved in reproductive processes exhibit species-specific divergence in
385 expression patterns in response to seasonal environmental changes.
386
387 **Relating seasonal gene expression divergence to sequence divergence.** If
388 divergence of seasonal timing of gene expression is related to selective constraint,
389 we would expect a correlation between sequence divergence patterns and the
390 divergence in seasonal expression peaks. To explore the relationship between the
391 divergence of seasonal gene expression and protein sequence evolution, we
392 calculated synonymous (dS) and non-synonymous substitution rates (dN) and their
393 ratio (dN/dS) for each gene using coding sequences from newly assembled
394 genomes of *Q. glauca* and *L. edulis*. We first confirmed that the highly expressed
395 genes evolve more slowly by identifying a significant negative correlation between
396 evolutionary rate (dN/dS) and mean of the gene expression (r = -0.21, p = 1.8 ×
397 10^{-118} for dN/dS; Figure 5–figure supplement 5A; Supplementary file 7). This
398 finding aligns with previous studies in yeasts, mammals, and plants (Drummond et
399 al., 2005; Gaut et al., 2011; Liao and Zhang, 2006; Yang and Gaut, 2011). By
400 targeting the genes that exhibited annual periodicity across both species and
401 tissues, we examined the relationship between these evolutionary measures and
402 peak month differences in seasonal gene expression between species ($\Delta\varphi$) in
403 leaves and bud tissues using linear regression. Assessment of the significance of
404 the regression coefficient using a *t*-test revealed that genes with smaller peak
405 month differences between species exhibited lower values of dN/dS in buds (p
406 =2.6 × 10^{-2}) and marginally lower values of dN/dS in leaves (p = 5.7 × 10^{-2}). The
407 lower values were also found in dN in both leaves (p = 8.1 × 10^{-3}) and buds (p =
408 7.8 × 10^{-3}), whereas no significant trend was observed for dS in both tissues
409 (Figure 5E; Figure 5–figure supplement 5B, C; Supplementary file 7). We further
410 investigated whether the association between the divergence in seasonal
411 expression peaks and sequence divergence patterns varies between genes
412 expressed in winter and in the growing season. We classified genes that exhibited
413 annual periodicity across both species and tissues into four groups based on their
414 expression peak timing (winter or growing seasons) and tissue specificity (leaves
415 or buds) and compared the extent of protein sequence divergence across these
416 groups. We found no significant differences in dN/dS (Kruskal Wallis test, p = 2.8
417 × 10^{-1}; Figure 5–figure supplement 5*D*; Supplementary file 7) between genes
418 expressed in winter in both tissues and those in growing seasons. The same result
419 was observed even when focusing only on genes with conserved peaks (peak
420 month difference < 2), with no significant differences in dN/dS between these gene
421 sets (Kruskal Wallis test, p = 3.5 × 10^{-1}; Figure 5–figure supplement 5E;

Supplementary file 7). These results suggest that coding-sequence variation is unlikely to be the primary driver of seasonal gene expression patterns.

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 (Figure 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 (Figure 5–figure supplement 6A)(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 (Figure 5–figure supplement 6B), 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 only weakly correlated with the divergence of seasonal peaks in gene expression across species. To understand the evolutionary mechanism underlying the seasonal constraint on gene expression that results in similar expression patterns during winter, an important future direction is to explore the 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 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

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
biological sample (*Quercus. glauca*)	Leaf	This paper		Freshly isolated from *Q. glauca*
biological sample (*Quercus. glauca*)	Bud	This paper		Freshly isolated from *Q. glauca*

biological sample (*Quercus. acuta*)	Leaf	This paper		Freshly isolated from *Q. acuta*
biological sample (*Quercus. acuta*)	Bud	This paper		Freshly isolated from *Q. acuta*
biological sample (*Lithocar pus. edulis*)	Leaf	This paper		Freshly isolated from L. *edulis*
biological sample (*Lithocar pus. edulis*)	Bud	This paper		Freshly isolated from *L. edulis*
biological sample (*Lithocar pus. glaber*)	Leaf	This paper		Freshly isolated from L. *glaber*
biological sample (*Lithocar pus. glaber*)	Bud	This paper		Freshly isolated from *L. glaber*
commerci al assay or kit	QIAamp DNA Mini Kit	Qiagen	Cat. #: 51304	
commerci al assay or kit	SMRTbell Express Template Prep Kit 2.0	PacBio	Cat. #: 100-938-900	
commerci al assay or kit	Sequel II Binding Kit 2.0	PacBio	Cat. #: 101-789-500	

commercial assay or kit	PureLink™ Plant RNA Reagent	Thermo Fisher Scientific	Cat. #: 12322012	
commercial assay or kit	TURBO DNA-free kit™	Thermo Fisher Scientific	Cat. #: AM1907	
commercial assay or kit	NEBNext Ultra II RNA Library Prep Kit for Illumina	New England BioLabs	Cat. #: E7770	
commercial assay or kit	NEBNext Ultra II DNA Library Prep Kit	New England BioLabs	Cat. #: E7645	
commercial assay or kit	NEBNext Multiplex Oligos for Illumina	New England BioLabs	Cat. #: E7335, E7500, E7710	
commercial assay or kit	RNeasy Plant Mini Kit	Qiagen	Cat. #: 74904	
commercial assay or kit	Agilent RNA 6000 Nano kit	Agilent Technologies	Cat. #: 5067-1511	
commercial assay or kit	High Sensitivity DNA kit	Agilent Technologies	Cat. #: 5067-4626	
commercial assay or kit	NEBNext Library Quant Kit for Illumina	New England BioLabs	Cat. #: E7630	
chemical compound, drug	RNA-stabilizing reagent (RNAlater)	Thermo Fisher Scientific	Cat. #: AM7021	
software, algorithm	Hifiasm1	Hifiasm	Hifiasm1 v0.15.4-r347	

software, algorithm	blobtools2	BlobToolKit	RRID:SCR_023351	
software, algorithm	purge_dups3	purge_dups	RRID:SCR_021173	
software, algorithm	HiRise	Putnam et al., 2016	RRID:SCR_023037	
software, algorithm	fastp	Chen et al., 2018	RRID:SCR_016962	
software, algorithm	STAR	Dobin et al., 2013	RRID:SCR_004463	
software, algorithm	edgeR	edgeR	RRID:SCR_012802	
software, algorithm	Braker2	Brůna et al., 2021	RRID:SCR_018964	
software, algorithm	GeMoMa	Keilwagen et al., 2019	RRID:SCR_017646	
software, algorithm	GFFcompare	Pertea and Pertea, 2020	DOI: 10.12688/f1000research.23297.1	
software, algorithm	DIAMOND	Buchfink et al., 2014	RRID:SCR_016071	
software, algorithm	Salmon	Patro et al., 2017	RRID:SCR_017036	
software, algorithm	eggNOG-mapper	Cantalapiedra et al., 2021	RRID:SCR_021165	

software, algorithm	BLASTp	Camacho et al., 2009	RRID:SCR_001010	
software, algorithm	MCScanX	Wang et al., 2012	RRID:SCR_022067	
software, algorithm	GenomeScope2	Ranallo-Benavidez et al., 2020	DOI:10.1038/s41467-020-14998-3	
software, algorithm	KmerGenie	Chikhi and Medvedev, 2014	DOI:10.1093/bioinformatics/btt310	
software, algorithm	OrthoFinder 2	Emms and Kelly, 2019	RRID:SCR_017118	
software, algorithm	R	R project	RRID:SCR_001905	
software, algorithm	PAML	Yang, 2007	RRID:SCR_014932	

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, and the 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 (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.

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 for two years (October 2019 to October 2021 for the *Lithocarpus* species and April 2021 to April 2023 for the *Quercus species*) 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). *L. edulis* and *L. glaber* were studied at the Imajuku Field Activity Center (33º33′ N, 130º16′ E, 84 to 111 m a.s.l). The Imajuku and Ito Campus study sites, located only 7.3 km apart, experience nearly identical temperature conditions. *Q. acuta* was monitored in Sefuri mountains (33º26′ N, 130º22′ E, 970 to 974 m a.s.l), a higher-altitude, cooler site.

556 Environmental data from Ito Campus and the Imajuku Field Activity Center were
557 obtained from the Japan Meteorological Agency, while data from Sefuri Mountain
558 were collected using a logger installed at the sampling site. Additionally, to assess
559 the potential impact of different sampling periods between the *Lithocarpus* and
560 *Quercus* species on the gene expression analysis, we conducted an additional
561 year of sampling for *L. edulis* and *L. glaber*, collecting leaf and bud samples from
562 November 2021 to November 2022, and used the samples collected within the
563 same period for all the four species (May 2021 to November 2022).
564 Samples for RNA-seq were taken from the sun-exposed crown
565 (approximately 2–5 m from the ground) using long pruning shears from 11:30 to
566 12:30 h. For each sample, 0.2–0.4 g of tissue was preserved in a 2 ml microtube
567 containing 1.5 ml of RNA-stabilizing reagent (RNAlater; Thermo Fisher Scientific,
568 Waltham, MA, USA) immediately after harvesting. The samples were transferred
569 to the laboratory within 2 hr after sampling, stored at 4°C overnight and then stored
570 at −80°C until RNA extraction. During transport to the laboratory, the samples were
571 kept in a cooler box with ice to maintain a low temperature.
572
573 **RNA extraction and RNA-seq analysis**
574 RNA was extracted independently from bud and leaf samples of each tree. Total
575 RNA of the leaf sample was extracted in accordance with the method described in
576 a previous study (Miyazaki et al., 2014). Total RNA of the bud sample was
577 extracted using the PureLink™ Plant RNA Reagent (Thermo Fisher Scientific,
578 Waltham, MA, USA) and subsequently processed with DNase using the TURBO
579 DNA-free kit™ (Thermo Fisher Scientific) according to their protocols and purified
580 using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to its RNA
581 cleanup protocol. RNA integrity was examined using the Agilent RNA 6000 Nano
582 kit on a 2100 Bioanalyzer (Agilent Technologies), and the RNA yield was
583 determined using a NanoDrop ND-2000 spectrophotometer (Thermo Fisher
584 Scientific). Five to six micrograms of RNA extracted from each sample were sent
585 to Hangzhou Veritas Genetics Medical Institute Co., Ltd., where a cDNA library
586 was prepared with a NEBNext Ultra II RNA Library Prep Kit for Illumina, and 150
587 base paired-end transcriptome sequencing of each sample was conducted using
588 an Illumina NovaSeq6000 sequencer (Illumina).
589 The quantification of gene expression followed a three-step process. First,
590 quality filtering was performed using the fastp v0.23.2 with default options: phred
591 quality ≥15, the maximum number of N base in one read = 5 (Chen et al., 2018).
592 Second, RNA-seq reads were mapped to the reference genome using STAR
593 v2.7.10b (Dobin et al., 2013). Third, RNA-seq reads were quantified using RSEM
594 v1.3.1 (Li and Dewey, 2011). We excluded the genes with low expression level
595 using the threshold, the value of mean Reads Per Kilobase (RPK) < 1 for each
596 species. Then, we further normalized the expression level using the value of Gene
597 length corrected Trimmed Mean of M-values (GeTMM) (Smid et al., 2018) using
598 an R package "edgeR" v3.42.4. For the downstream analysis we calculated the
599 mean expression levels of GeTMM of each gene across multiple individuals and
600 transformed them to $\log_2$(GeTMM+1). To confirm that the effect of the difference

601 in sampling period is negligible, we added the RNA-Seq data of *L. edulis* and *L.*
602 *glaber* collected from December 2021 to December 2022. We extracted the same
603 genes in the existing data that curated based on the mean RPK values across two-
604 year sampling periods. Then we further normalized the expression level using
605 GeTMM.
606
607 **Genome annotation**
608 The RNA-seq reads of *Q. glauca* (8 samples obtained by a previous study (20)
609 collected from May to December 2017; GEO's accession number: GSE211384)
610 and *L. edulis* (20 samples covering leaf, bud and flower tissues collected in
611 January, April, July, and November 2021) were respectively mapped against the
612 scaffolds of *Q. glauca* and *L. edulis* obtained by HiRise with HISAT2 v2.1.0 (Kim
613 et al., 2019). The bam files generated by the mappings and amino acid sequences
614 of the 9 species obtained from NCBI (*Castanea mollissima*: GCA_000763605.2
615 (NCBI's accession number), *Q. lobata*: GCF_001633185.1, *Q. robur*:
616 GCF_932294415.1, *Q. suber*: GCF_002906115.1, *Carpinus fangiana*:
617 GCA_006937295.1, *Carya illinoinensis*: GCF_018687715.1, *Juglans macrocarpa*
618 × *J. regia*: GCF_004785595.1, *Juglans regia*: GCF_001411555.2, *Morella rubra*:
619 GCA_003952965.2) were used for ab initio gene prediction by Braker2 v2.1.6
620 (Brůna et al., 2021), and the genes with highest-score were selected to exclude
621 splicing variants. In parallel, the gene models were respectively constructed in *Q.*
622 *glauca* and *L. edulis* by GeMoMa v1.7.1 (Keilwagen et al., 2019) with RNA-seq
623 reads and amino acid sequences used for the gene prediction in Braker2, and the
624 genes with highest-score were selected to exclude splicing variants. The
625 duplicated genes between Braker2 and GeMoMa were excluded by GFFcompare
626 v0.11.2 (Pertea and Pertea, 2020). The remaining genes were searched against
627 UniProtKB (Bateman et al., 2025) by DIAMOND v2.0.5 (Buchfink et al., 2014) in
628 'more-sensitive' option with E-value $\leq 10^{-20}$ and identity ≥25%. Next, gene
629 expression analysis was performed by Salmon v1.4.0 (Patro et al., 2017), and the
630 genes with TPM values >0.1 were considered as expressed. The orthologous
631 genes were searched against EggNOG (Hernández-Plaza et al., 2023) using
632 eggNOG-mapper v2.1.8 (Cantalapiedra et al., 2021) with E-value $\leq 10^{-10}$. The
633 genes were searched against protein sequences of *Castanea mollissima*
634 (Accession: GCA_000763605.2), *Quercus lobata* (GCF_001633185.2), and
635 *Arabidopsis thaliana* (Araport11 (Krishnakumar et al., 2015)) by BLASTp v2.13.0
636 (Camacho et al., 2009) with E-value $\leq 10^{-20}$. According to these similarity searches,
637 the genes whose product names contain keywords related to transposable
638 elements (TEs) were categorized into TE. The genes having similarity hits and
639 being expressed by Salmon were classified into high-confidence (HC) genes, while
640 remaining genes were categorized into low-confidence (LC) genes. The genes
641 categorized into HC were used for further analyses.
642
643 **Genome synteny analysis**
644 To investigate the syntenic relationship between *Q. glauca* and *L. edulis*, we first
645 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

For genome size estimation and phylogenetic reconstruction, 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).

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. We also used pair-end short reads for genome size estimation of *Q. glauca* and confirmed that resulted estimates were not different. For *L. edulis*, paired-end short reads from four individuals were merged and quality-trimmed using fastp v0.23.2 (Chen et al., 2018). The genome size was then estimated using KmerGenie (Chikhi and Medvedev, 2014) with a k-mer size of 55 applied to the trimmed reads.

Phylogenetic reconstruction

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 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

691 2.5.5 for orthogroup determination (Emms and Kelly, 2019). The genomic data of
692 *Q. robur* and *J. regia* were downloaded from NCBI (dhQueRobu3.1 and Walnut
693 2.0), while *F. sylvatica* data were sourced from beechgenome.net (Mishra et al.,
694 2022). Since the *F. sylvatica* dataset contained numerous predicted transposable
695 elements (TEs), we re-annotated genes using three databases—EggNOG5
696 (EggNOG-mapper), Pfam35.0 (HMMER v3.3.2), and UniRef90_2023_0628
697 (DIAMOND v2.1.8.162)—and excluded genes identified as TEs in at least one
698 database. In total, we identified 26692 orthogroups and focused on single-copy
699 orthogroups after filtering out genes mapped outside the 12 chromosomes. This
700 resulted in a final dataset of 11749 genes for further analysis (Data S2).
701
702 **Hierarchical clustering and principal component analysis (PCA)**
703 To assess the similarity of the genome-wide transcriptional profiles across
704 orthologous genes and 483 samples collected in different seasons, we performed
705 hierarchical clustering using the monthly time series data of 11749 orthologous
706 genes from October 2019 to April 2023. For each orthologous gene, there were 26
707 or 27 time points, with three individuals each for *Q. glauca*, *Q. acuta*, *L. glaber* and
708 one individual for *L. edulis*. We calculated the mean expression levels of each
709 orthologous gene across three individuals in each species and subsequently
710 normalized the values by adjusting the mean to zero and the standard deviation to
711 one. We performed hierarchical clustering using the Ward method and the
712 Euclidian distance using the pheatmap function in R (v 1.0.12). To determine the
713 number of the clusters, we calculated Within-Cluster-Sum of Squared Error (WSS)
714 and determined the number of clusters as an elbow point of it (Figure 2–figure
715 supplement 4). To assess the seasonal expression dynamics of 11749 orthologous
716 genes, we also performed PCA of the gene expression profiles from all samples
717 using the prcomp function in R (v 4.3.1).
718
719 **Identification of genes with rhythmic expression**
720 To identify genes with rhythmic expression and determine the period and peak
721 month of gene expression, we applied the RAIN algorithm (Thaben and
722 Westermark, 2014) using the rain function in R (v 1.34.0). We set the following
723 settings: deltat = 1, period = 12, period.delta = 11, peak.border = (0,1), nr.series =
724 1, method= "independent", na.rm= T , adjp.method = "ABH". Based on the results
725 of the RAIN analysis, we identified genes with significant periodicity using a
726 Benjamini-Hochberg (BH) adjusted *q*-value threshold of < 0.01. Since the period
727 length and peak month in the RAIN output are expressed in terms of the number
728 of time points, we converted them into months using the sampling span (four
729 weeks) and sampling dates. Subsequently, we categorized the rhythmic genes into
730 three categories based on the period length T (in months): “half-annual” ($0 \leqq T <$
731 8), “annual” ($8 \leqq T \leqq 16$), and “long” ($T > 16$).
732
733 **Calculation of Pearson’s correlation coefficient across species**
734 To quantify the similarity of seasonal gene expression pattern across species, we
735 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 (Figure 2–figure supplement 3). 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 (Figure 2–figure supplement 5). Flowering was defined as the stage where either male or female flowers were visible (Figure 2–figure supplement 5). We subsequently plotted the proportion of branches exhibiting leaf unfolding and flowering. The data are provided in Supplementary file 3.

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 ≧ 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

All sequencing data generated in this study have been deposited in the DNA Data Bank of Japan (DDBJ) and are available under the BioProject accession number PRJDB20498. RNA-seq reads used for expression analyses are available under the DDBJ Sequence Read Archive (DRA) run accession numbers DRR659185-DRR659374, DRR664882-DRR665873, and DRR724385-DRR724496, while DNA sequences used for genome assembly, genome size estimation and phylogenetic reconstruction are available under DRR659899-DRR659929. The assembled genomes are also available from DDBJ: the *Q. glauca* genome under accession numbers BAAHOF010000001-BAAHOF010000797, and the *L. edulis* genome under BAAHOE010000001-BAAHOE010000750. 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 JP23H04965 and JP23H04966.

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. *PNAS* **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 Research* **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 Genomics and Bioinformatics* **3**:1–11. doi:10.1093/nargab/lqaa108

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

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

872 Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021.
873 eggNOG-mapper v2: functional annotation, orthology assignments, and
874 domain prediction at the metagenomic scale. *Molecular Biology and Evolution*
875 **38**:5825–5829. doi:10.1093/molbev/msab293

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

883 Chen LT, Wu K. 2010. Role of histone deacetylases HDA6 and HDA19 in ABA and
884 abiotic stress response. *Plant Signaling & Behavior* **5**:1318–1320.
885 doi:10.4161/psb.5.10.13168

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

888 Chen SC, Cannon CH, Kua CS, Liu JJ, Galbraith DW. 2014. Genome size variation
889 in the Fagaceae and its implications for trees. *Tree Genetics and Genomes*
890 **10**:977–988. doi:10.1007/s11295-014-0736-y

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

894 Cleland EE, Wolkovich EM. 2024. Effects of phenology on plant community
895 assembly and structure. *Annual Review of Ecology, Evolution, and*
896 *Systematics* **55**:471–492. doi:10.1146/annurev-ecolsys-102722-011653

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

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

905 Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly
906 expressed proteins evolve slowly. *PNAS* **102**:14338–14343.
907 doi:10.1073/pnas.0504070102

908 El Taher A, Böhne A, Boileau N, Ronco F, Indermaur A, Widmer L, Salzburger W.
909 2021. Gene expression dynamics during rapid organismal diversification in
910 African cichlid fishes. *Nature Ecology and Evolution* **5**:243–250.
911 doi:10.1038/s41559-020-01354-3

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

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

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

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

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

926 Forrest J, Miller-Rushing AJ. 2010. Toward a synthetic understanding of the role
927 of phenology in ecology and evolution. *Philosophical Transactions of the*
928 *Royal Society B: Biological Sciences* **365**:3101–3112.
929 doi:10.1098/rstb.2010.0145

930 Fretwell SD. 1972. Populations in a seasonal environment. *Monographs in*
931 *Population Biology* **5**:1–217.

932 Gaut B, Yang L, Takuno S, Eguiarte LE. 2011. The patterns and causes of variation
933 in plant nucleotide substitution rates. *Annual Review of Ecology, Evolution,*
934 *and Systematics* **42**:245–266. doi:10.1146/annurev-ecolsys-102710-145119

935 Hernández-Plaza A, Szklarczyk D, Botas J, Cantalapiedra CP, Giner-Lamia J,
936 Mende DR, Kirsch R, Rattei T, Letunic I, Jensen LJ, Bork P, von Mering C,
937 Huerta-Cepas J. 2023. eggNOG 6.0: enabling comparative genomics across
938 12 535 organisms. *Nucleic Acids Research* **51**:D389–D394.
939 doi:10.1093/nar/gkac1022

940 Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J, Crowl AA,
941 Deng M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-

942 Rodríguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C,
943 Rodríguez-Correa H, Schulze ED, Simeone MC, Sork VL, Valencia-Avalos S.
944 2020. Genomic landscape of the global oak phylogeny. *New Phytologist*
945 **226**:1198–1212. doi:10.1111/nph.16162

946 Ikezaki Y, Kudo SN, Nakata T, Koita S, Munakata R, Yazaki K, Torimaru T, Tomaru
947 N, Isobe S, Hirakawa H, Kusumi J, Satake A. 2025. Molecular evolution of
948 terpene synthase underlying the diversification of isoprene emission in
949 Fagaceae. doi:10.1101/2025.07.31.667835

950 Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate
951 phylotypic period during organogenesis. *Nature Communications* **2**.
952 doi:10.1038/ncomms1248

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

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

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

964 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome
965 alignment and genotyping with HISAT2 and HISAT-genotype. *Nature*
966 *Biotechnology* **37**:907–915. doi:10.1038/s41587-019-0201-4

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

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

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

976 Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M,
977 Rosen BD, Cheng CY, Moreira W, Mock SA, Stubbs J, Sullivan JM, Krampis
978 K, Miller JR, Micklem G, Vaughn M, Town CD. 2015. Araport: The Arabidopsis

979 Information Portal. *Nucleic Acids Research* **43**:D1003–D1009.
980 doi:10.1093/nar/gku1200

981 Kubo T, Iwasa Y. 1996. Phenological pattern of tree regeneration in a model for
982 forest species diversity. *Theoretical Population Biology* **49**:90–117.
983 doi:10.1006/tpbi.1996.0004

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

987 Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins
988 and gene expression levels are coupled in *Drosophila* and are independently
989 associated with mRNA abundance, protein length, and number of protein-
990 protein interactions. *Molecular Biology and Evolution* **22**:1345–1354.
991 doi:10.1093/molbev/msi122

992 Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev
993 E, Silver DH, Feder M, Fernandez-Valverde SL, Nakanishi N, Simmons D,
994 Simakov O, Larsson T, Liu SY, Jerafi-Vider A, Yaniv K, Ryan JF, Martindale
995 MQ, Rink JC, Arendt D, Degnan SM, Degnan BM, Hashimshony T, Yanai I.
996 2016. The mid-developmental transition and the evolution of animal body
997 plans. *Nature* **531**:637–641. doi:10.1038/nature16994

998 Li B, Dewey CN. 2011. RSEM: Accurate transcript quantification from RNA-Seq
999 data with or without a reference genome. *BMC Bioinformatics* **12**.
1000 doi:10.1186/1471-2105-12-323

1001 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-
1002 Wheeler transform. *Bioinformatics* **25**:1754–1760.
1003 doi:10.1093/bioinformatics/btp324

1004 Liao BY, Zhang J. 2006. Low rates of expression profile divergence in highly
1005 expressed genes and tissue-specific genes during mammalian evolution.
1006 *Molecular Biology and Evolution* **23**:1119–1128. doi:10.1093/molbev/msj119

1007 Liu H, Zhang R, Zhou BF, Shen Z, Chen XY, Gao J, Wang B. 2023. Chromosome-
1008 scale genome assembly of sweet tea (Lithocarpus polystachyus Rehder).
1009 *Scientific Data* **10**. doi:10.1038/s41597-023-02791-y

1010 Lotharukpong JS, Zheng M, Luthringer R, Liesner D, Drost H-G, Coelho SM. 2024.
1011 A transcriptomic hourglass in brown algae. *Nature* **635**:129–135.
1012 doi:10.1038/s41586-024-08059-8

1013 Manos PS, Zhou Z, Cannon CH. 2001. Systematics of Fagaceae: phylogenetic
1014 tests of reproductive trait evolution. *International Journal of Plant Sciences*
1015 **162**:1361–1379. doi:10.1086/322949

1016 Mishra B, Ulaszewski B, Meger J, Aury JM, Bodénès C, Lesur-Kupin I, Pfenninger
1017 M, Da Silva C, Gupta DK, Guichoux E, Heer K, Lalanne C, Labadie K,
1018 Opgenoorth L, Ploch S, Le Provost G, Salse J, Scotti I, Wötzel S, Plomion C,
1019 Burczyk J, Thines M. 2022. A chromosome-level genome assembly of the
1020 European Beech (Fagus sylvatica) reveals anomalies for organelle DNA
1021 integration, repeat content and distribution of SNPs. *Frontiers in Genetics* **12**.
1022 doi:10.3389/fgene.2021.691058

1023 Miyawaki‐Kuwakado A, Han Q, Kitamura K, Satake A. 2024. Impacts of climate
1024 change on the transcriptional dynamics and timing of bud dormancy release
1025 in Yoshino‐cherry tree. *PLANTS, PEOPLE, PLANET* **6**:1505–1521.
1026 doi:10.1002/ppp3.10548

1027 Miyazaki Y, Maruyama Y, Chiba Y, Kobayashi MJ, Joseph B, Shimizu KK, Mochida
1028 K, Hiura T, Kon H, Satake A. 2014. Nitrogen as a key regulator of flowering in
1029 Fagus crenata: Understanding the physiological mechanism of masting by
1030 gene expression analysis. *Ecology Letters* **17**:1299–1309.
1031 doi:10.1111/ele.12338

1032 Morellato LPC, Camargo MGG, Gressler E. 2013. "A review of plant phenology in
1033 South and Central America." in Phenology: An Integrative Environmental
1034 Science. Springer.

1035 Nagano AJ, Kawagoe T, Sugisaka J, Honjo MN, Iwayama K, Kudoh H. 2019.
1036 Annual transcriptome dynamics in natural environments reveals plant
1037 seasonal adaptation. *Nature Plants* **5**:74–83. doi:10.1038/s41477-018-0338-
1038 z

1039 Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, Page DC. 2019.
1040 Conservation, acquisition, and functional impact of sex-biased gene
1041 expression in mammals. *Science (1979)* **365**. doi:10.1126/science.aaw7317

1042 Nishio H, Nagano AJ, Ito T, Suzuki Y, Kudoh H. 2020. Seasonal plasticity and diel
1043 stability of H3K27me3 in natural fluctuating environments. *Nature Plants*
1044 **6**:1091–1097. doi:10.1038/s41477-020-00757-1

1045 Nuzhdin S V., Wayne ML, Harmon KL, McIntyre LM. 2004. Common pattern of
1046 evolution of gene expression level and protein sequence in Drosophila.
1047 *Molecular Biology and Evolution* **21**:1308–1317. doi:10.1093/molbev/msh128

1048 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast
1049 and bias-aware quantification of transcript expression. *Nature Methods*
1050 **14**:417–419. doi:10.1038/nmeth.4197

1051 Pertea M, Pertea G. 2020. GFF Utilities: GffRead and GffCompare. *F1000Res* **9**.
1052 doi:10.12688/f1000research.23297.1

1053 Piao S, Liu Q, Chen A, Janssens IA, Fu Y, Dai J, Liu L, Lian X, Shen M, Zhu X.
1054 2019. Plant phenology and global climate change: Current progresses and
1055 challenges. *Global Change Biology* **25**:1922–1940. doi:10.1111/gcb.14619

1056 Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory
1057 evolution. *PNAS* **104**:8605–8612. doi:10.1073/pnas.0700488104

1058 Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ,
1059 Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. 2016.
1060 Chromosome-scale shotgun assembly using an in vitro method for long-range
1061 linkage. *Genome Research* **26**:342–350. doi:10.1101/gr.193474.115

1062 Quint M, Drost HG, Gabel A, Ullrich KK, Bönn M, Grosse I. 2012. A transcriptomic
1063 hourglass in plant embryogenesis. *Nature* **490**:98–101.
1064 doi:10.1038/nature11394

1065 Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and
1066 Smudgeplot for reference-free profiling of polyploid genomes. *Nature*
1067 *Communications* **11**. doi:10.1038/s41467-020-14998-3

1068 Satake A, Hagiwara T, Nagano AJ, Yamaguchi N, Sekimoto K, Shiojiri K, Sudo K.
1069 2024a. Plant molecular phenology and climate feedbacks mediated by
1070 BVOCs. *Annual Review of Plant Biology* **75**:605–627. doi:10.1146/annurev-
1071 arplant-060223-032108

1072 Satake A, Imai R, Fujino T, Tomimoto S, Ohta K, Na'iem M, Indrioko S, Widiyatno
1073 W, Purnomo S, Morales AM, Nizhynska V, Tani N, Suyama Y, Sasaki E,
1074 Kasahara M. 2024b. Somatic mutation rates scale with time not growth rate in
1075 long-lived tropical trees. *Elife* **12**. doi:10.7554/eLife.88456

1076 Satake A, Nagahama A, Sasaki E. 2022. A cross‐scale approach to unravel the
1077 molecular basis of plant phenology in temperate and tropical climates. *New*
1078 *Phytologist* **233**:2340–2353. doi:10.1111/nph.17897

1079 Satake A, Ohta K, Takeda-Kamiya N, Toyooka K, Kusumi J. 2023. Seasonal gene
1080 expression signatures of delayed fertilization in Fagaceae. *Molecular Ecology*
1081 **32**:4801–4813. doi:10.1111/mec.17079

1082 Schwartz MD. 2013. Phenology: An Integrative Environmental Science, 2nd ed.
1083 Springer Dordrecht.

1084 Smaczniak C, Immink RGH, Muiño JM, Blanvillain R, Busscher M, Busscher-
1085 Lange J, Dinh QD, Liu S, Westphal AH, Boeren S, Parcy F, Xu L, Carles CC,
1086 Angenent GC, Kaufmann K. 2012. Characterization of MADS-domain
1087 transcription factor complexes in *Arabidopsis* flower development. *PNAS*
1088 **109**:1560–1565. doi:10.1073/pnas.1112871109

1089 Smid M, Coebergh van den Braak RRJ, van de Werken HJG, van Riet J, van Galen
1090 A, de Weerd V, van der Vlugt-Daane M, Bril SI, Lalmahomed ZS, Kloosterman
1091 WP, Wilting SM, Foekens JA, IJzermans JNM, Martens JWM, Sieuwerts AM.
1092 2018. Gene length corrected trimmed mean of M-values (GeTMM) processing
1093 of RNA-seq data performs similarly in intersample analyses while improving
1094 intrasample comparisons. *BMC Bioinformatics* **19**. doi:10.1186/s12859-018-
1095 2246-7

1096 Thaben PF, Westermark PO. 2014. Detecting rhythms in time series with rain.
1097 *Journal of Biological Rhythms* **29**:391–400. doi:10.1177/0748730414553029

1098 True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary
1099 trajectories. *Evolution & Development* **3**:109–119. doi:10.1046/j.1525-
1100 142x.2001.003002109.x

1101 Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo
1102 H, Kissinger JC, Paterson AH. 2012. MCScanX: A toolkit for detection and
1103 evolutionary analysis of gene synteny and collinearity. *Nucleic Acids*
1104 *Research* **40**. doi:10.1093/nar/gkr1293

1105 Yang L, Gaut BS. 2011. Factors that contribute to variation in evolutionary rate
1106 among arabidopsis genes. *Molecular Biology and Evolution* **28**:2359–2369.
1107 doi:10.1093/molbev/msr058

1108 Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular*
1109 *Biology and Evolution* **24**:1586–1591. doi:10.1093/molbev/msm088

1110 Yoo SJ, Chung KS, Jung SH, Yoo SY, Lee JS, Ahn JH. 2010. BROTHER of FT
1111 and TFL1 (BFT) has TFL1-like activity and functions redundantly with TFL1 in
1112 inflorescence meristem development in *Arabidopsis*. *Plant Journal* **63**:241–
1113 253. doi:10.1111/j.1365-313X.2010.04234.x

1114 Zhang X, Yu G, Dai Y, Zhang H, Wang K, Han J. 2023. High-resolution Hi-C maps
1115 highlight multiscale chromatin architecture reorganization during cold stress in
1116 Brachypodium distachyon. *BMC Plant Biology* **23**. doi:10.1186/s12870-023-
1117 04269-w

1118 Zhou BF, Yuan S, Crowl AA, Liang YY, Shi Y, Chen XY, An QQ, Kang M, Manos
1119 PS, Wang B. 2022. Phylogenomic analyses highlight innovation and
1120 introgression in the continental radiations of Fagaceae across the Northern
1121 Hemisphere. *Nature Communications* **13**. doi:10.1038/s41467-022-28917-1

1122 Zhu GQ, Qu L, Xue HW. 2024. Casein kinase 1 AELs promote senescence by
1123 enhancing ethylene biosynthesis through phosphorylating WRKY22
1124 transcription factor. *New Phytologist* **244**:116–130. doi:10.1111/nph.19785

1125 Zoldos V, Papes D, Cerbah M, Panaud O, Besendorfer V, Siljak-Yakovlev S. 1999.
1126 Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal

conserved genome organization among 11 Quercus species. *Theoretical and Applied Genetics* **99**:969–977. doi:10.1007/s001220051404

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

Zrimec J, Buric F, Kokina M, Garcia V, Zelezniak A. 2021. Learning the Regulatory Code of Gene Expression. *Frontiers in Molecular Biosciences* **8**. doi:10.3389/fmolb.2021.673363

1138 **Figures**

1139 **Figure 1.** Synteny and distribution of genomic features of *Q. glauca* and *L. edulis*
1140 genomes. (A) A phylogenetic tree of five Fagaceae species from three genera,
1141 *Quercus*, *Lithocarpus*, and *Fagus*. Bootstrap values for all branches are 100. (B)
1142 Synteny relationship between the *Q. glauca* and *L. edulis* genomes. The collinear
1143 blocks within the genomes were displayed by lines. Black lines show reverse
1144 alignment and gray lines represent regular. Numbers correspond to the 12
1145 chromosomes. (C–D) Circos plots of the *Q. glauca* (C) and *L. edulis* genomes (D).
1146 Lanes depict circular representation of chromosome length (Mbp) (a), gene density
1147 (b), GC density (c), and repeat (d). Lines in the inner circle represent links between
1148 synteny-selected paralogs. (E) Orthogroup classification based on the orthogroup
1149 size difference between the *Q. glauca* and *L. edulis* genomes. The bar plot
1150 illustrates the number of genes in gene families categorized into 1:1 orthologs, N:N
1151 (many-to-many), 1:N (one-to-many), N:1 (many-to-one), and species-specific
1152 unique genes. (F–G) Mean gene 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).

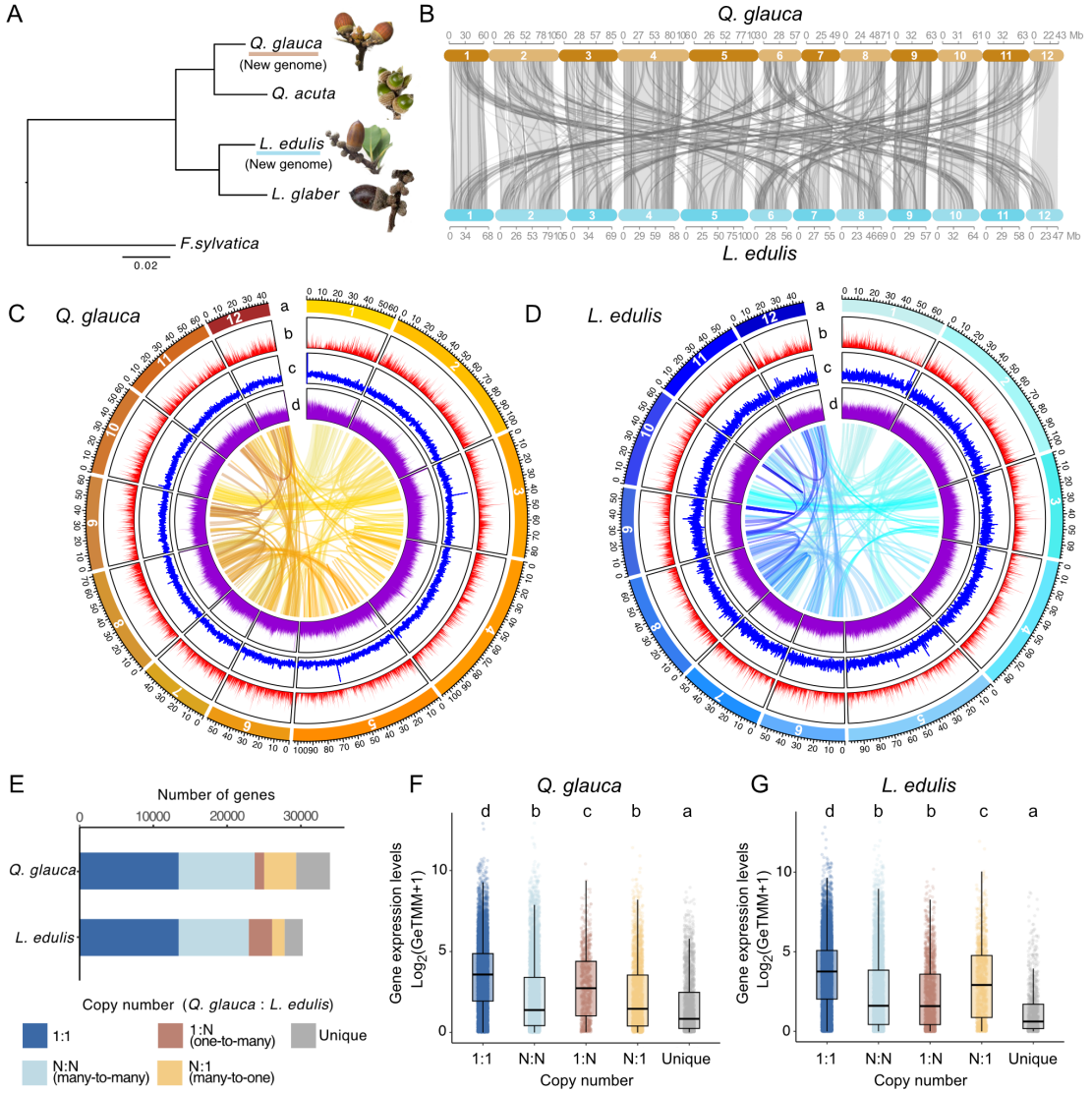


Figure 2. Seasonal gene expression dynamics. (A) Hierarchical clustering of 213 time points encompassing two tissues (leaf and bud) and four species (*Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber*), collected over two-year. Newly flushed leaf samples are indicated as asterisks. (B) Distribution 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.

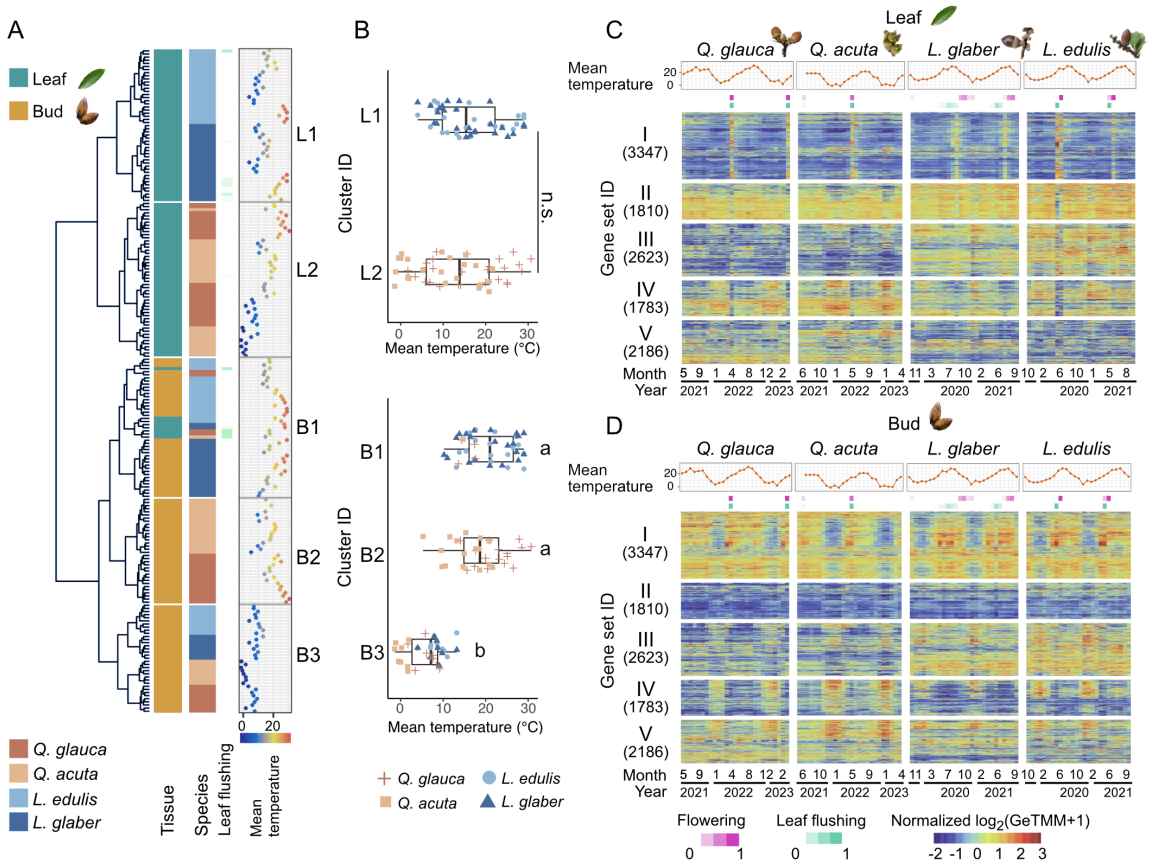


Figure 3. Distribution of period and peak month in rhythmic gene expression. (A–B) Distribution of period length (month) among the genes with significant rhythmicity calculated by RAIN for leaf (A) and bud (B) samples of each species. (C–D) Distribution of seasonal peaks among the genes with annual periodicity.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

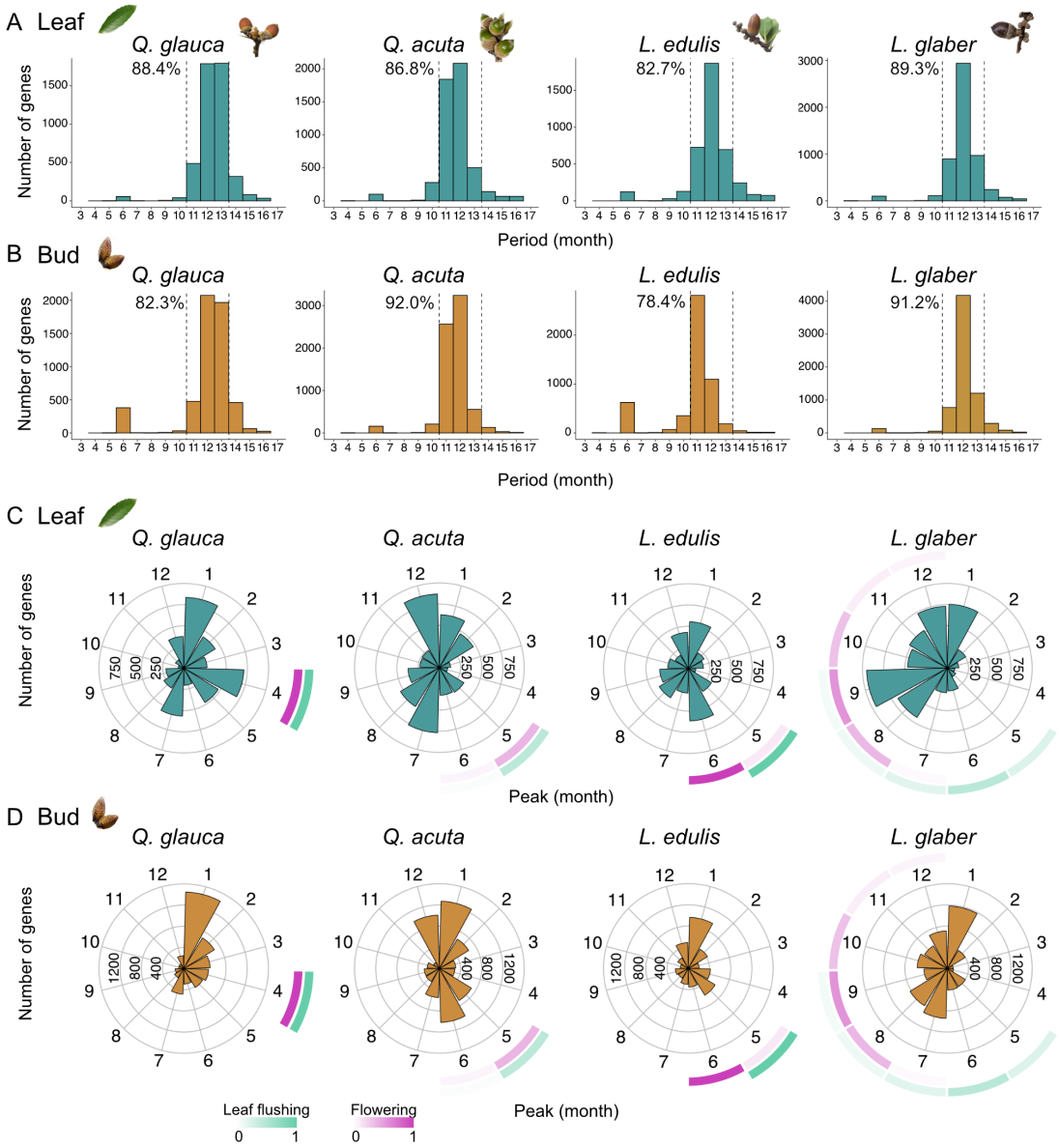

A Leaf
Q. glauca
88.4%
Q. acuta
86.8%
L. edulis
82.7%
L. glaber
89.3%
Number of genes
Period (month)
B Bud
Q. glauca
82.3%
Q. acuta
92.0%
L. edulis
78.4%
L. glaber
91.2%
Number of genes
Period (month)
C Leaf
Q. glauca
Q. acuta
L. edulis
L. glaber
Number of genes
Peak (month)
D Bud
Q. glauca
Q. acuta
L. edulis
L. glaber
Number of genes
Peak (month)
Leaf flushing
0
1
Flowering
0
1

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, $p < 0.05$).

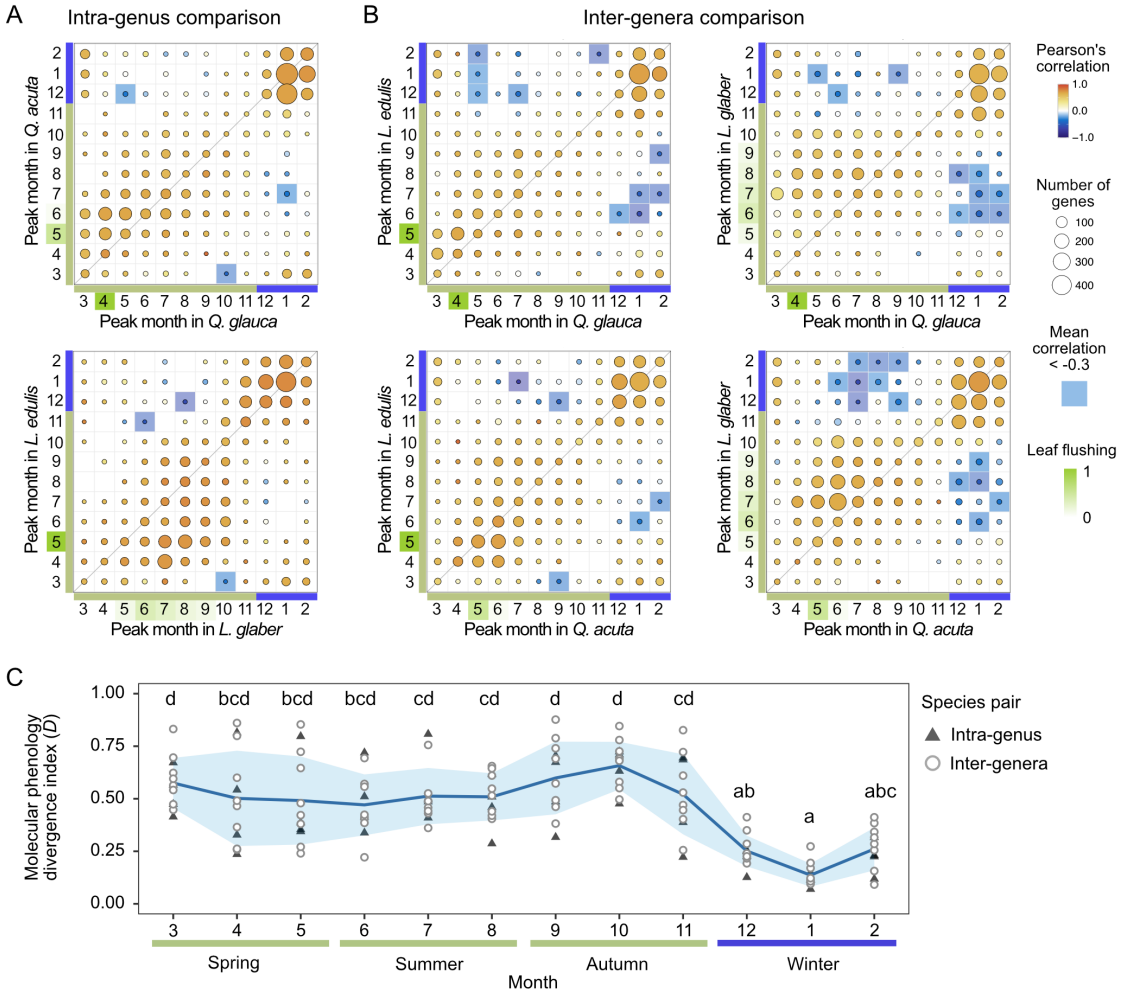


Figure 5. Phylogenetic constraints in the evolution of seasonal gene expression and the relationship between seasonal gene expression divergence and sequence evolution. (A) Distribution of pairwise Pearson's correlation coefficients for gene expression across species (Le: *L. edulis*, Lg: *L. glaber*, 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. The regression coefficient (β) and corresponding p-value are shown in each panel.

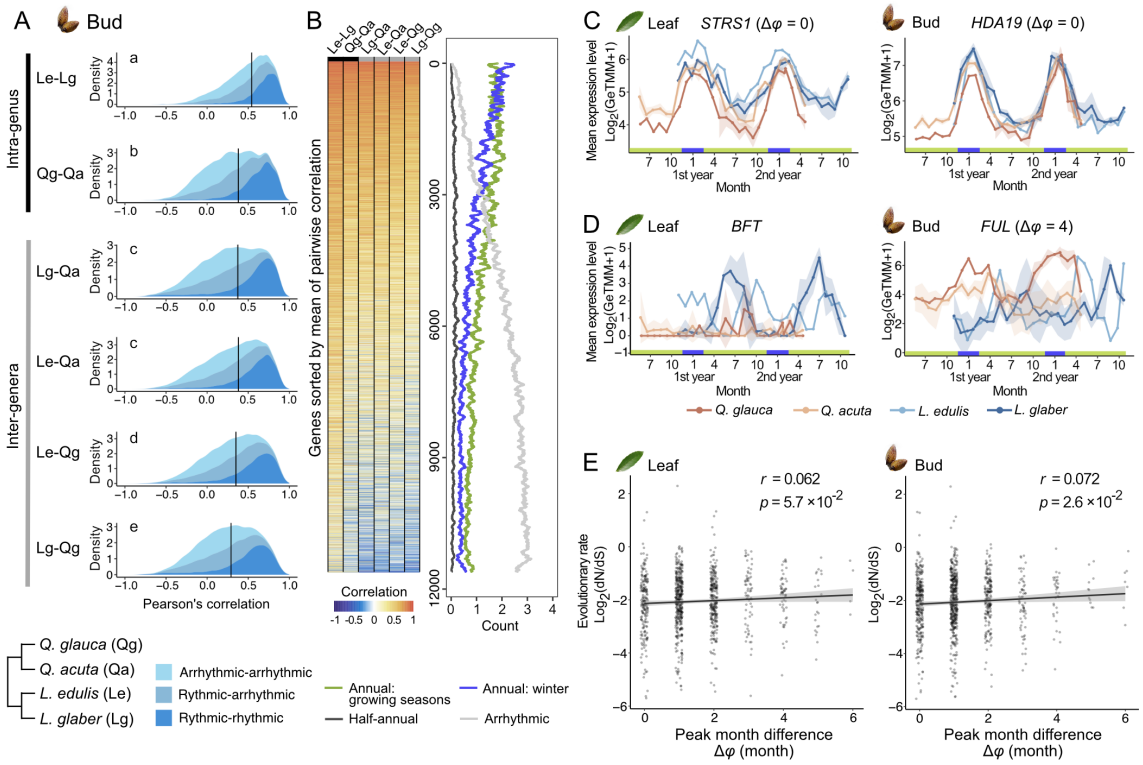
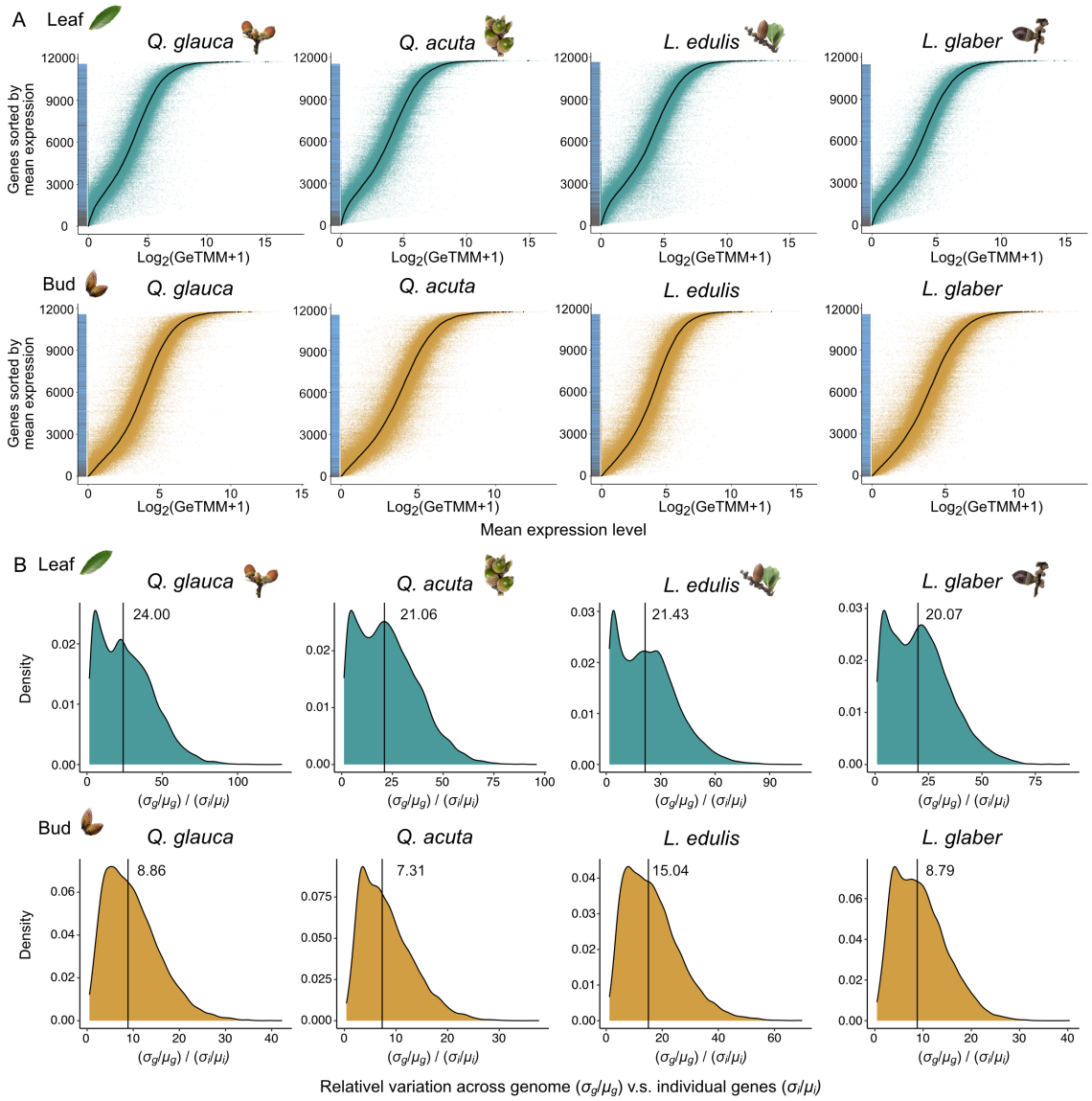
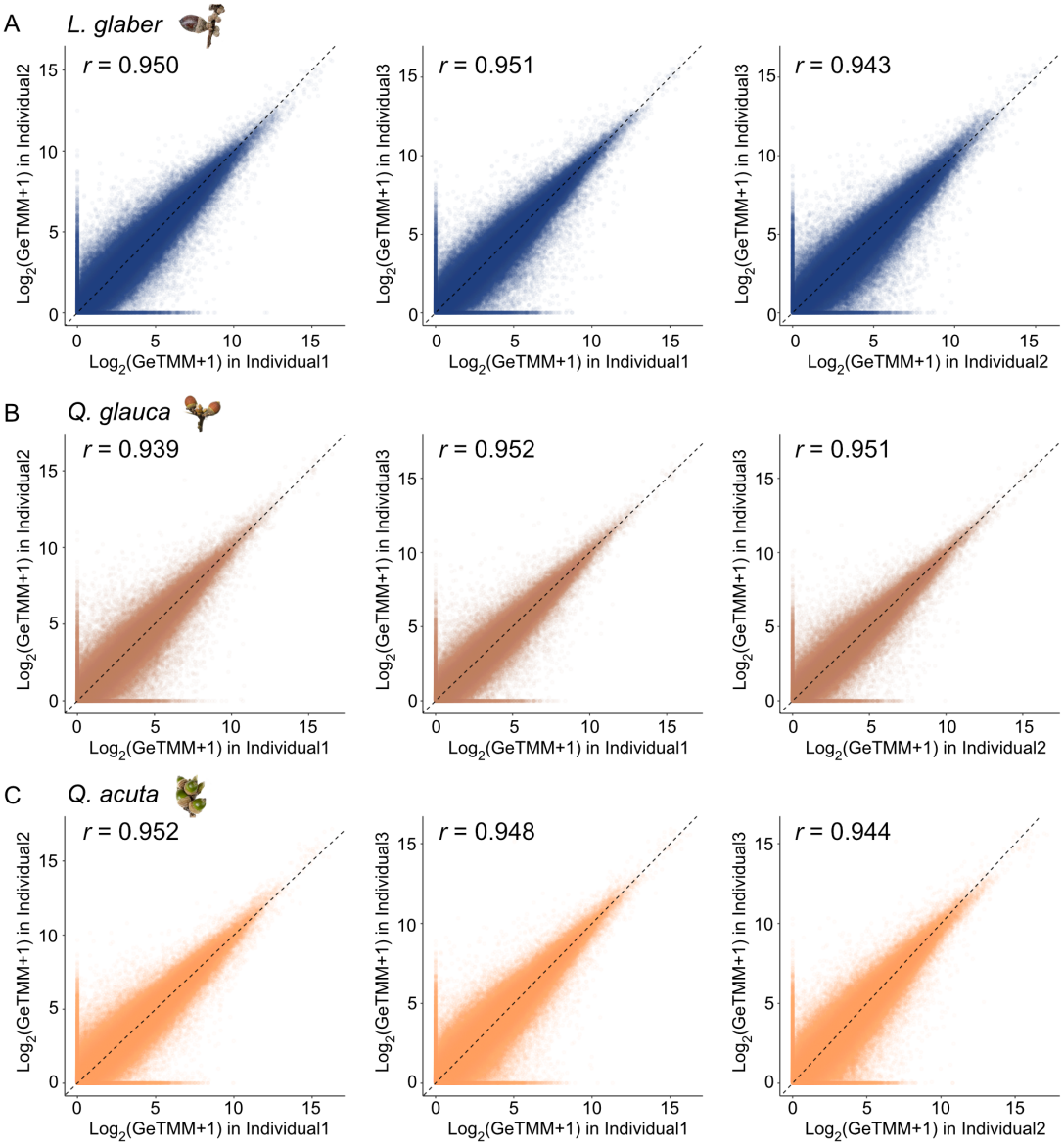


Figure supplements

Figure 2–figure supplement 1. Dynamic range of gene expression in seasonal environments. (A) Dynamic range of seasonal gene expression in leaves and buds for *Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber*. Each dot represents an expression level ($\log_2$(GeTMM+1)) on each sampling date, while black lines indicate the mean expression level of each gene over the entire sampling period. The bars on the left side of the panels distinguish between rhythmic (blue) and arhythmic (gray) genes. (B) Density plots showing the distribution of the ratio of standard deviation divided by the mean for each gene (σ_i / μ_i) and the overall variation across all genes in the genome (σ_g/μ_g) in leaves and buds for *Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber*.



1229 **Figure 2–figure supplement 2.** Comparison of gene expression across
1230 individuals. Scatter plots showing the expression levels of each gene measured
1231 from different individuals for *L. glaber* (A), *Q. glauca* (B), and *Q. acuta* (C). The
1232 Pearson's correlation coefficient for each pair of individuals is shown in the
1233 corresponding panel.



1234
1235

Figure 2–figure supplement 3. Seasonal fluctuations in environmental factors at the three study sites. Daily mean temperature and photoperiods at the three study sites (Ito campus, Kyushu university; Imajuku Field Activity Center; Mt. Sefuri in Kyushu, Japan) during the monitoring periods. The solid lines indicate the daily mean temperature, and the shaded area indicates the photoperiod. The study sites are located within a relatively small region, with distances between them ranging from approximately 10 to 20 km. The elevations vary, with Ito Campus (20–57 m a.s.l.) and Imajuku Field Activity Center (84–111 m a.s.l.) at lower altitudes, while Mt. Sefuri is at a higher elevation (970–974 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 since July 15, 2021.

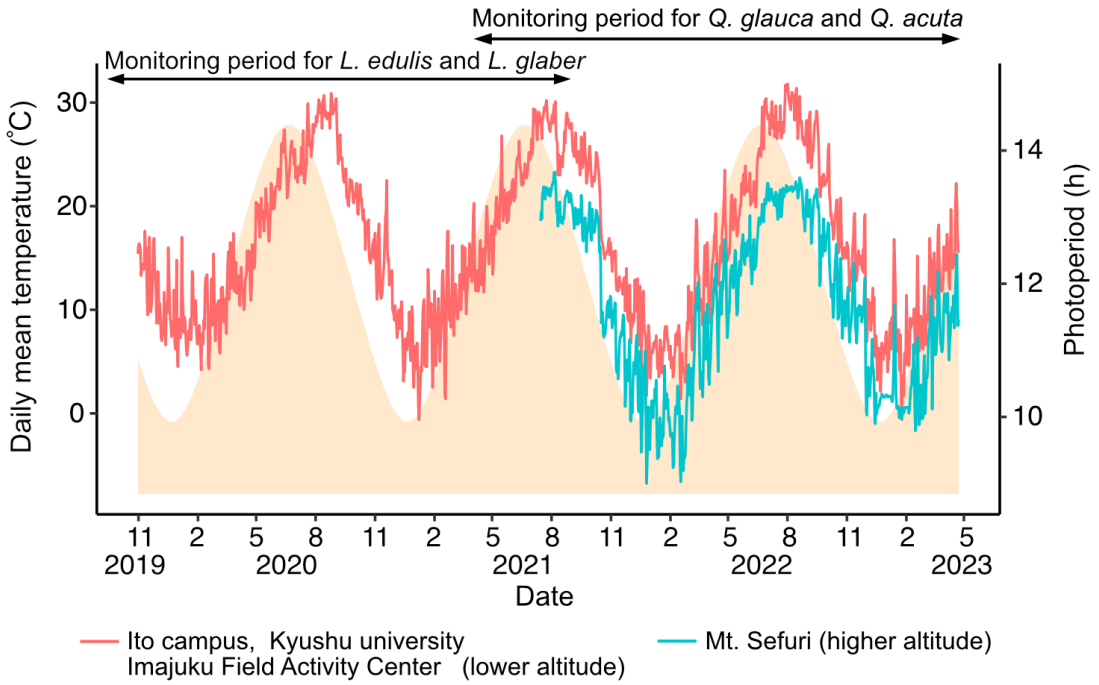


Figure 2–figure supplement 4. Determining the optimal number of clusters in hierarchical clustering. (A) Relationship between the number of clusters *k* and total Within-Cluster-Sum of Squared Error (WSS) on the clustering of 213 samples. We adopted the optimal *k* = 5 because it was the elbow point of the curve. (B) Relationship between the number of clusters *k* and total WSS on the clustering of 11749 genes. The elbow point was *k* = 4 or 5 and we selected the number of cluster *k* = 5. If *k* = 4 is selected, gene set IV and V (Figure 2D, E in main text) are merged.

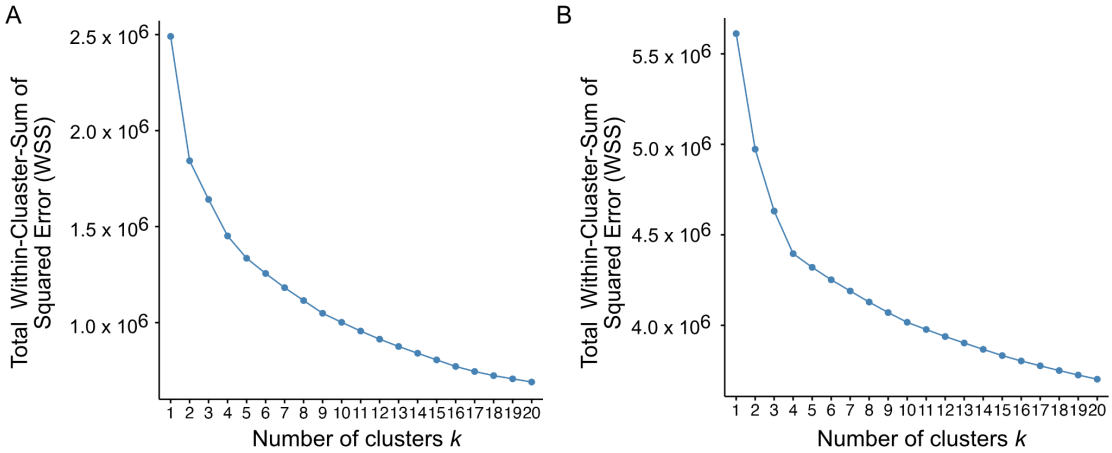


Figure 2–figure supplement 5. Phenological observations of leaf flushing, flowering (male and female flower) and winter buds of the four target species.

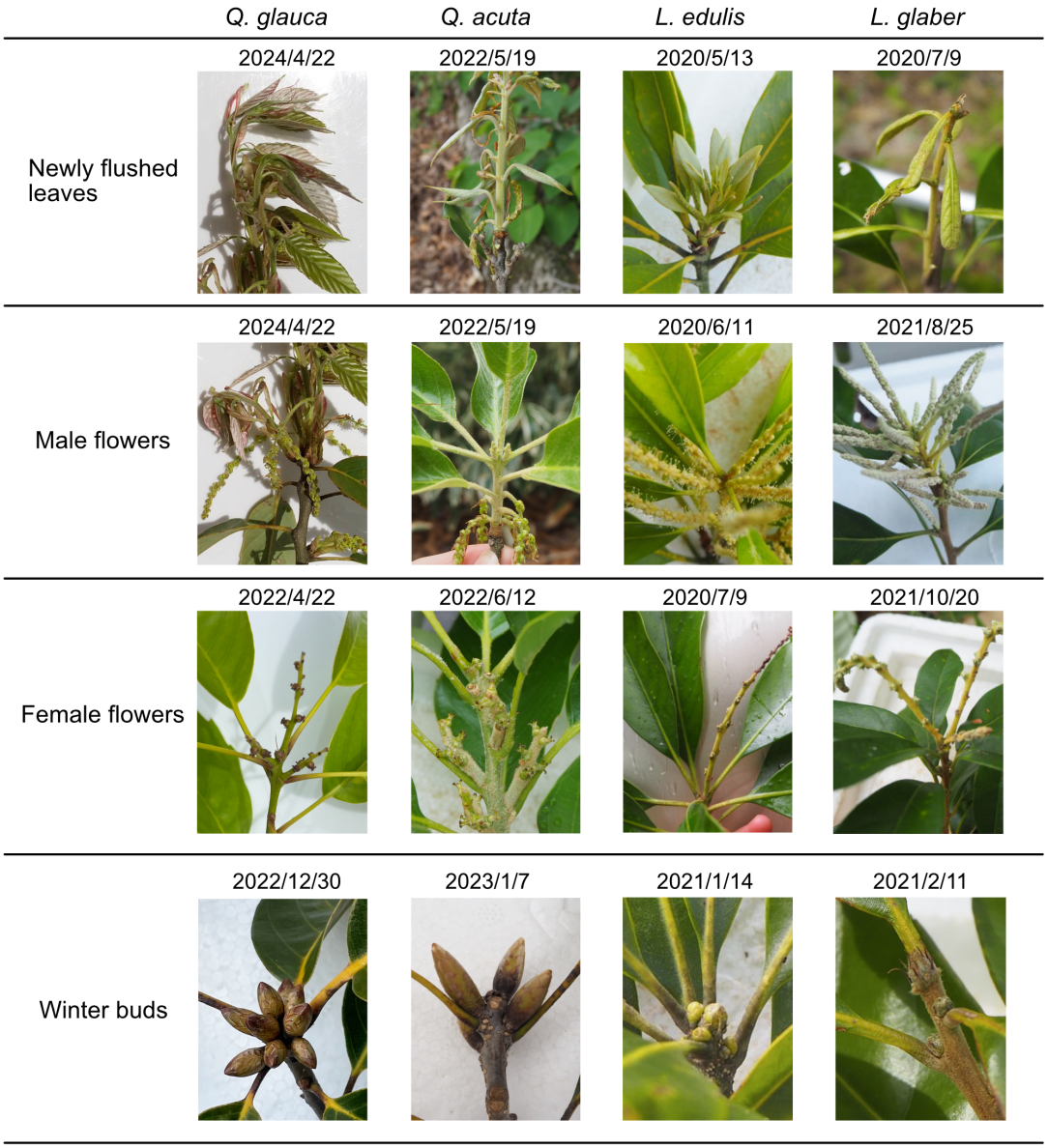


Figure 2–figure supplement 6. Seasonal gene expression dynamics observed within the same sampling period (from May 2021 to October 2022). (A) Hierarchical clustering of 160 time points encompassing two tissues (leaf and bud) and four species (*Q. glauca*, *Q. acuta*, *L. edulis*, and *L. glaber*). Newly flushed leaf samples are indicated as asterisks. (C–D) Hierarchical clustering of 11749 genes based on the similarity of gene expression dynamics. Daily mean temperature is shown in the top of the heatmaps.

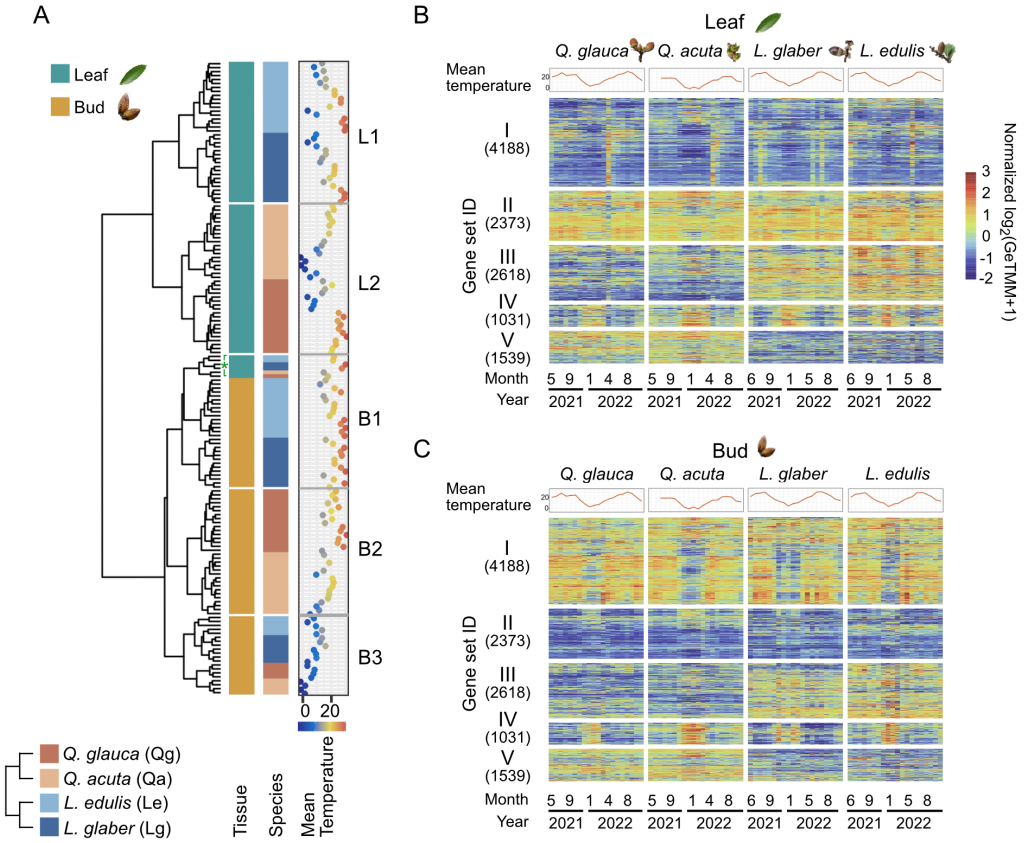


Figure 2–figure supplement 7. Seasonal gene expression dynamics under the identical environmental condition. (A) Hierarchical clustering of 161 time points from two tissues (leaf and bud) across three species (*Q. glauca*, *L. edulis*, and *L. glaber*) excluding *Q. acuta* sampled at higher elevation. Newly flushed leaf samples are indicated as asterisks. (C–D) Hierarchical clustering of 11749 genes based on the similarity of their gene expression dynamics. Daily mean temperature is shown in in the top of the heatmaps.

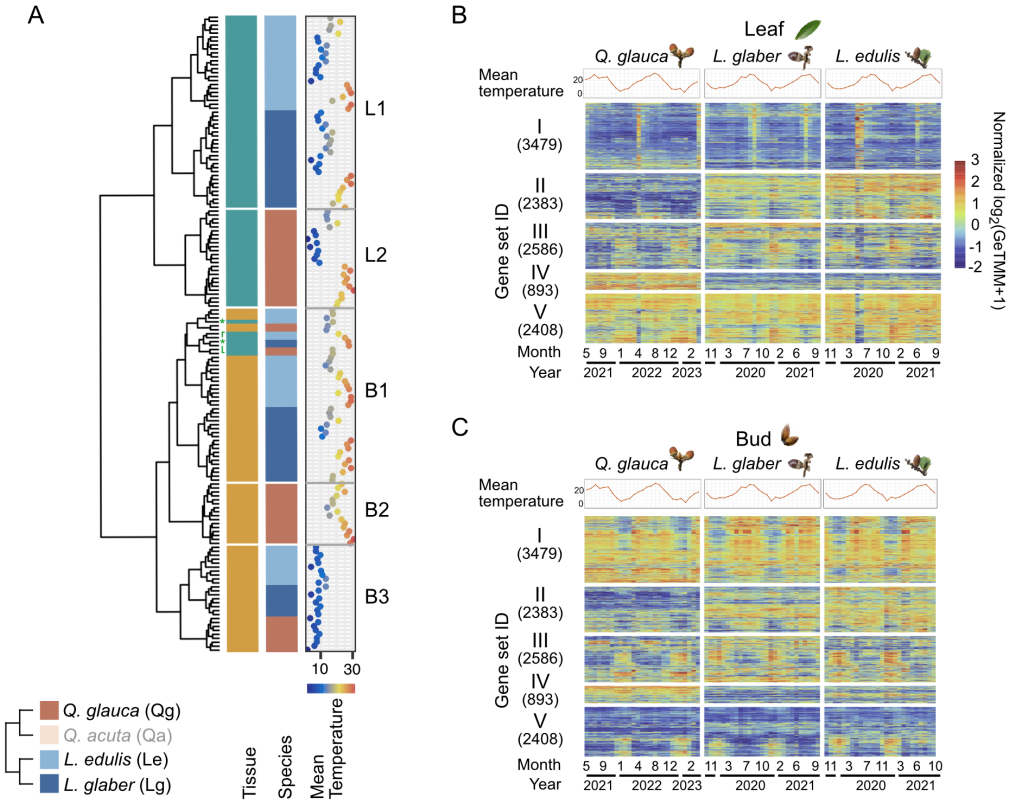


Figure 2–figure supplement 8. Distribution of environmental factors across clusters. Distribution of mean temperature for the 7 days prior to the sampling (A) and photoperiod for each sampling date (B) across clusters identified through hierarchical clustering in Figure 2A. The Mann-Whitney U test was used to compare climatic factors between leaf clusters (L1 and L2), while the Steel-Dwass test was applied for multiple comparisons among bud clusters (B1, B2, and B3). Newly flushed leaves were included within cluster B1. Different letters denote statistically significant differences ($p < 0.05$).

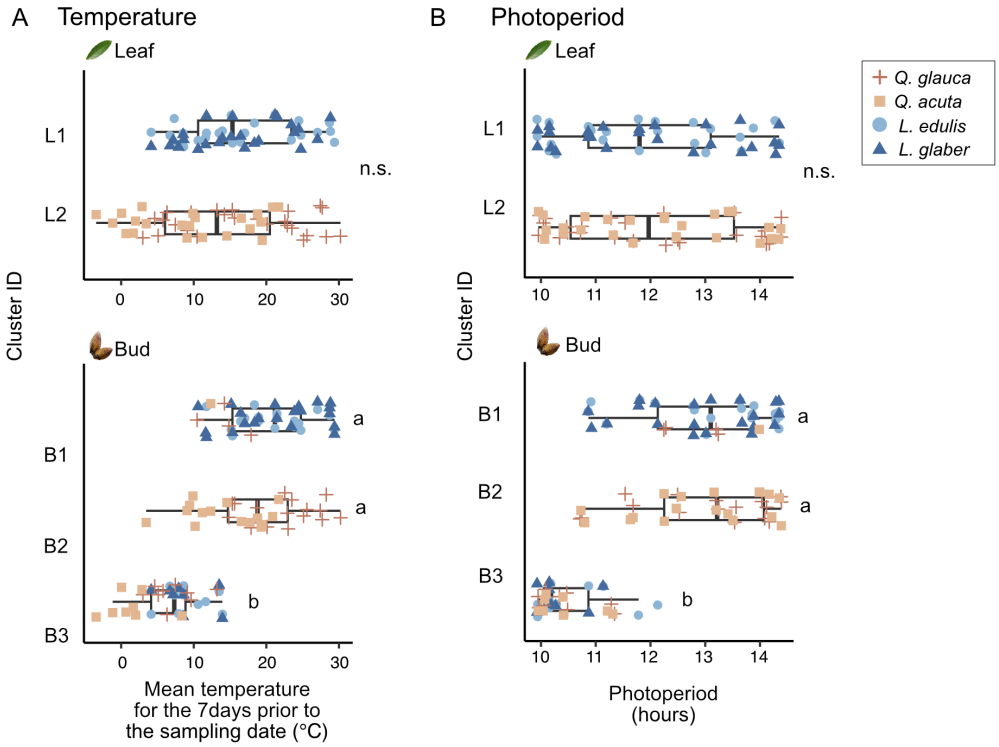


Figure 2–figure supplement 9. 3D plot of PC1–PC3 resulting from the PCA of 11749 orthologous genes in four Fagaceae species. Leaf samples are represented in a color gradient from green to blue, while bud samples range from orange to yellow. The color gradients indicate the mean temperature on the respective sampling dates for bud and leaf samples. The percentages in parentheses on each axis denote the variance explained by each principal component. The numbers within the plot indicate sampling months.

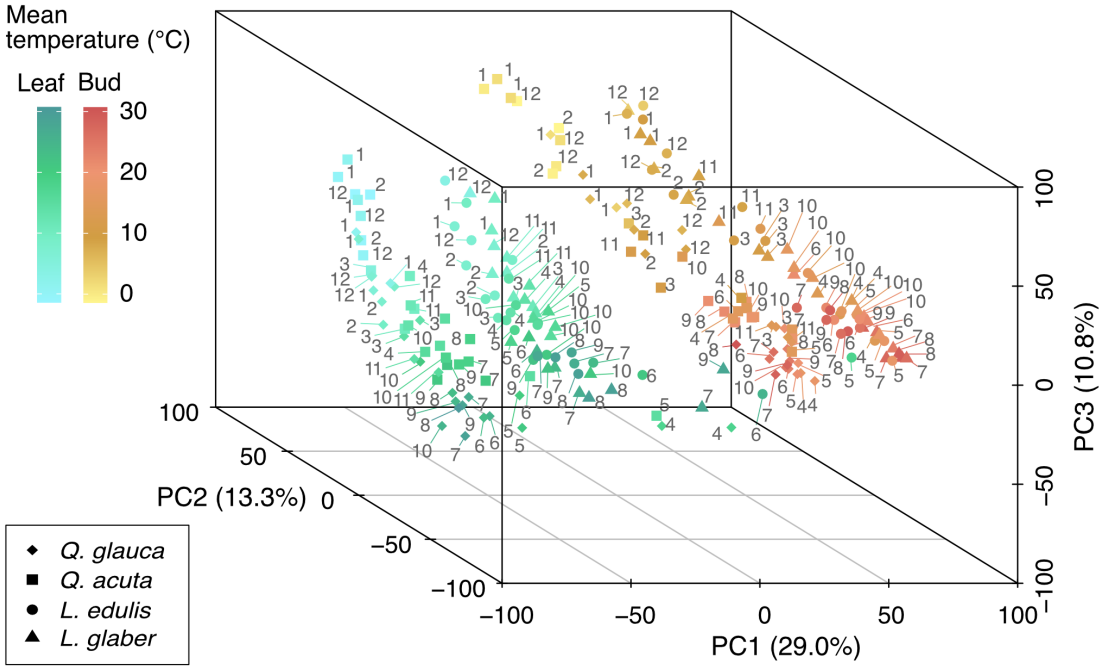


Figure 3–figure supplement 1. Number of rhythmic genes across species and distribution of peak months across half-annual genes. (A) Number of genes classified into annual (period range: 8-16 months), half-annual (< 8 month), and arrhythmic periodicity. (B–C) Number of common rhythmic genes shared among species within each species combination, categorized by annual (B) and half-annual (C) periodicity. Black dots beneath each bar indicate the species included in each combination. (D) Distribution of seasonal peaks across genes with half-annual periodicity. (E) Seasonal expression patterns of *WRKY22* in leaves and *KINγ1* in buds. For each gene, the species exhibiting a half-annual periodicity is highlighted with a thick, dark line, while the expression patterns in the other species are shown with thin, light lines.

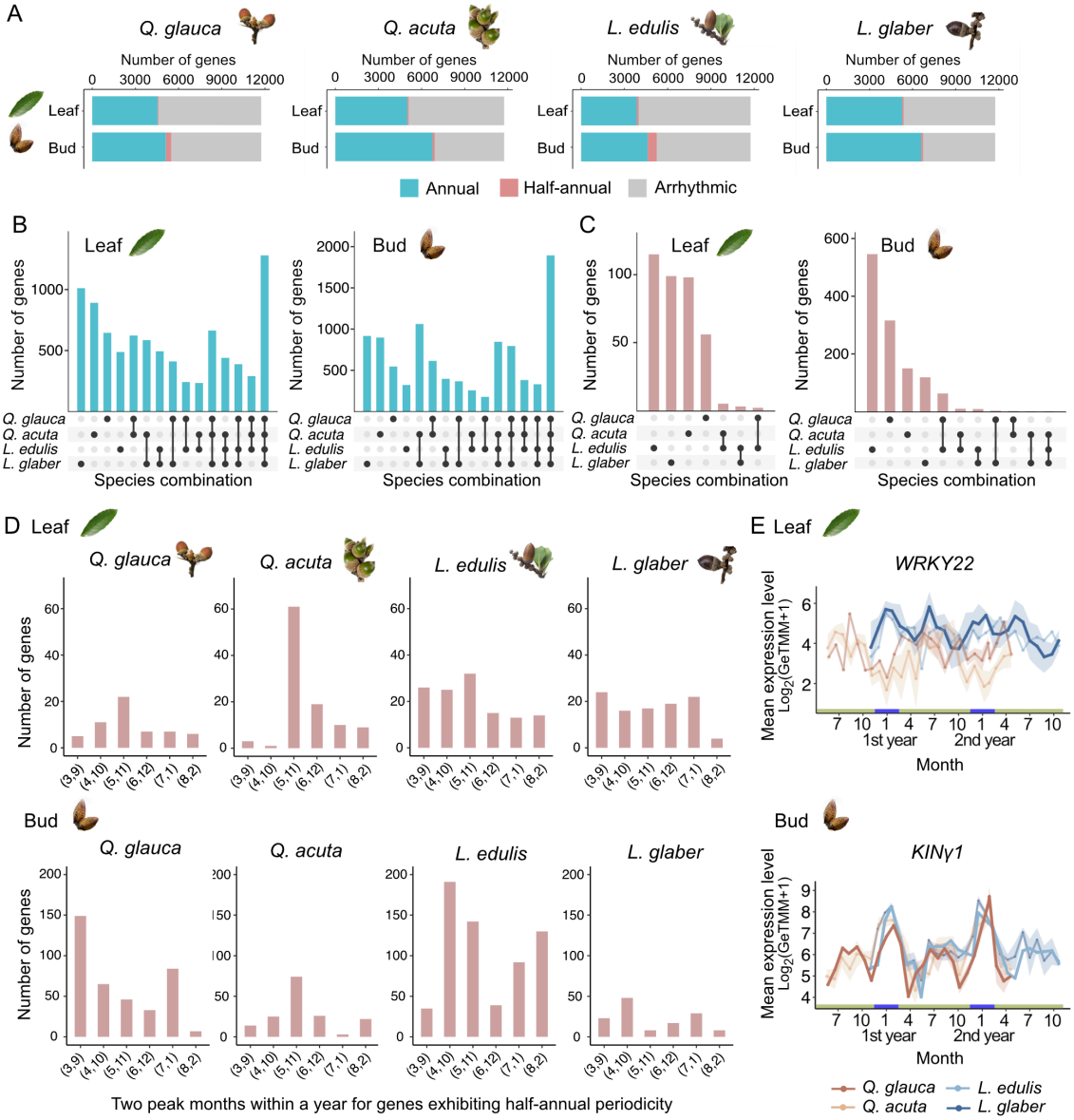
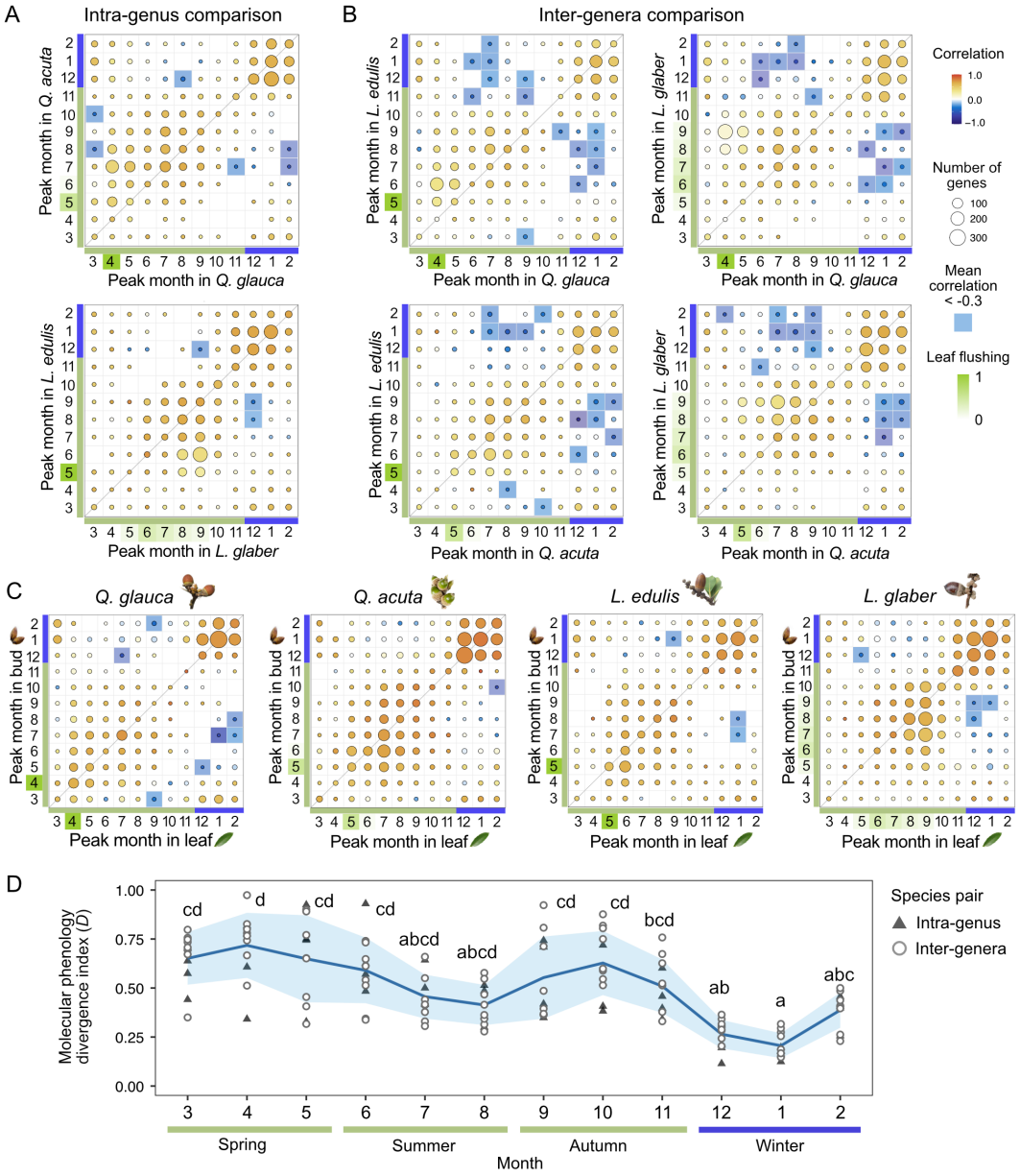
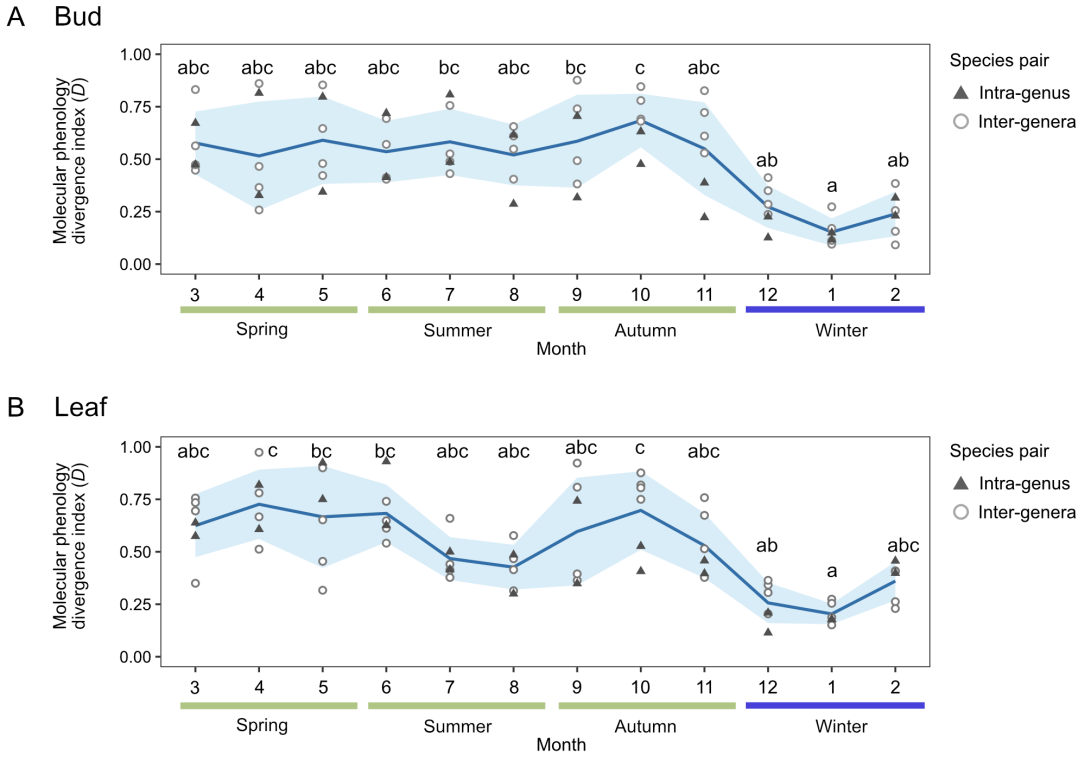


Figure 4–figure supplement 1. Comparison of peak months in seasonal gene expression across species in leaves. (A) Intra-genus and (B) inter-genera comparison of peak months in seasonal gene expression. (C) Comparison of peak months between leaves and buds. The size and color of the circles represent 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. (D) Plot of molecular phenology divergence index *D* across months. Triangles and circles indicate the *D* values calculated from intra- 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, $p < 0.05$).

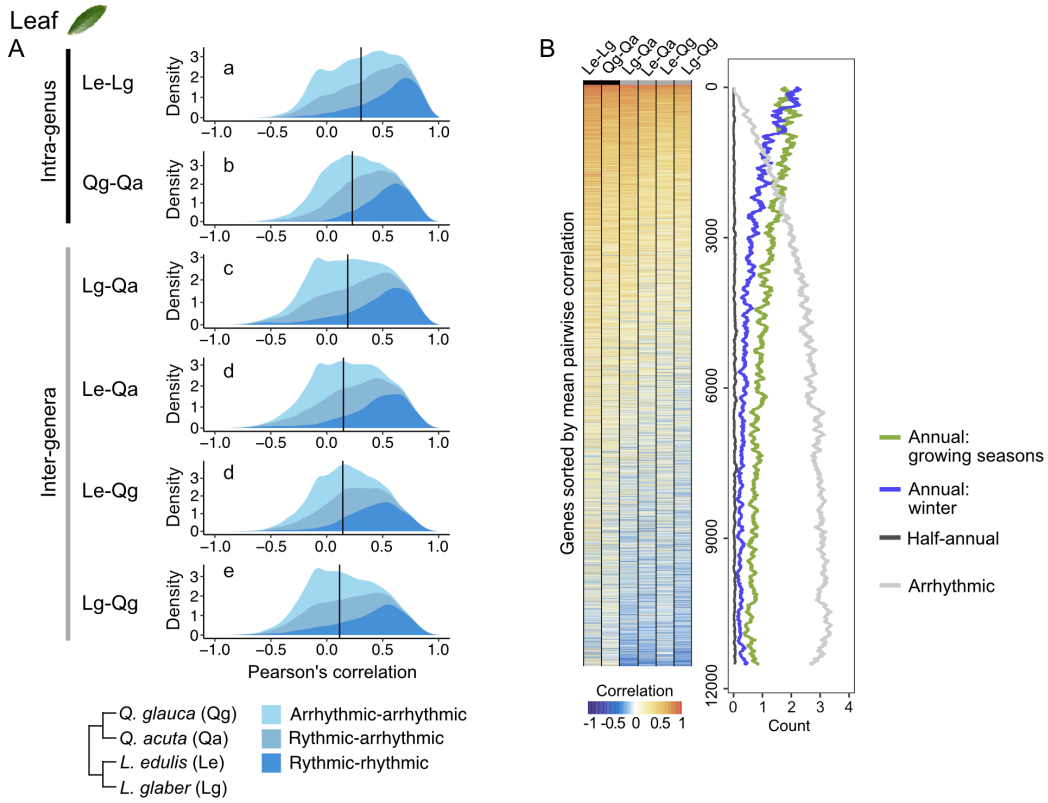


1325 **Figure 4–figure supplement 2.** Seasonal difference of molecular phenology
1326 divergence index *D* calculated for species under the identical environmental
1327 condition. Plot of molecular phenology divergence index *D* across months in buds
1328 (A) and leaves (B) calculated from the gene expression in *Q. glauca*, *L. edulis*, and
1329 *L. glaber*. Triangles and circles indicate the values of *D* calculated from intra-genus
1330 and inter-genera comparison, respectively. The line and shaded envelope indicate
1331 the mean and standard deviation (s.d.). Different letters denote statistically
1332 significant differences (Nemenyi test, $p < 0.05$).

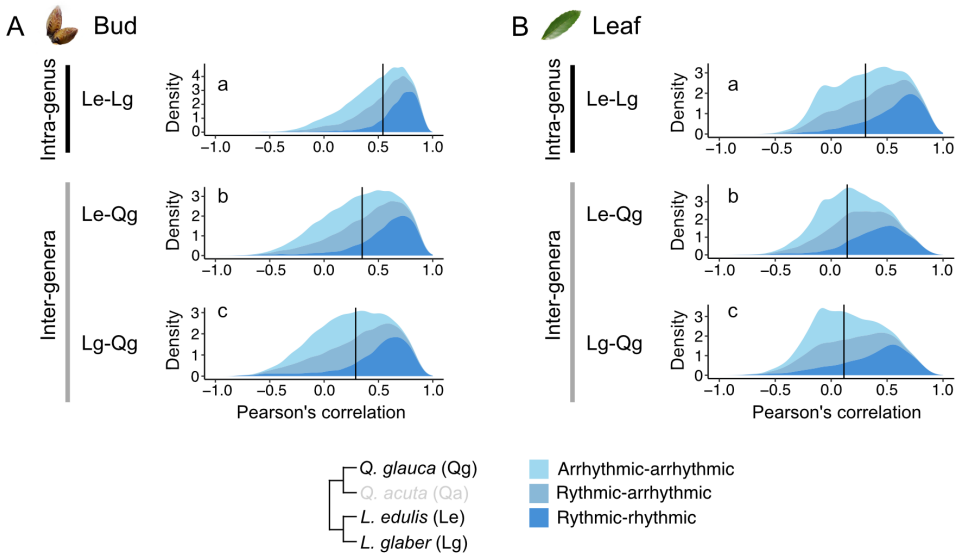


1333
1334

Figure 5–figure supplement 1. Phylogenetic constraints in the evolution of seasonal gene expression in leaves. (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 three categories based on the presence or absence of rhythmicity in gene expression. The median values of Pearson's correlation across all the genes were depicted as black bars. Different letters denote significantly different values (Nemenyi test, $p < 0.05$). (B) Heatmap of pairwise Pearson's correlation coefficients for each gene, along with the mean count of genes in four categories of seasonal gene expression: (1) rhythmic genes with annual periodicity peaking in the growing season (March–November), (2) rhythmic genes with annual periodicity peaking in winter (December–February), (3) rhythmic genes with half-annual periodicity, and (4) arrhythmic genes. The mean count was calculated for each block of 100 genes. Genes were ordered based on the mean Pearson's correlation of their expression across all species pairs.

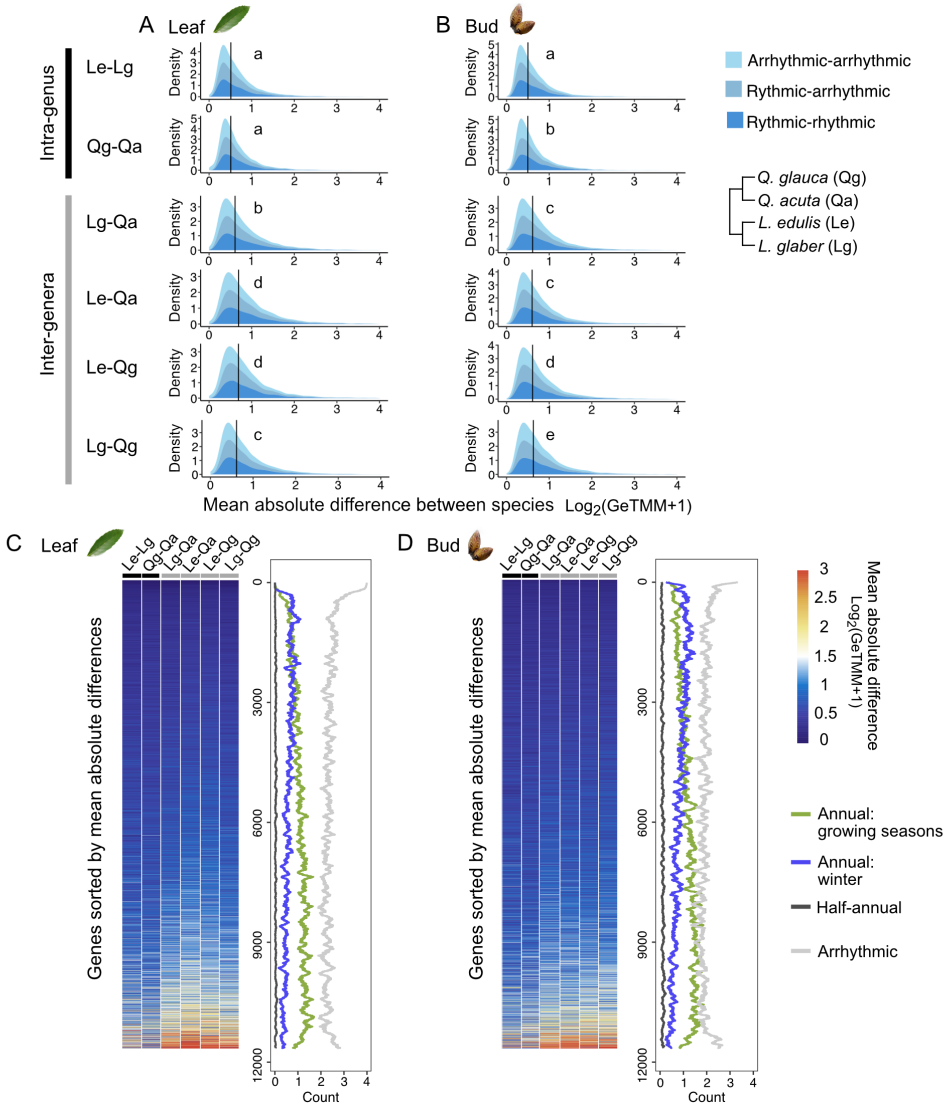


1352 **Figure 5–figure supplement 2.** Distribution of pairwise Pearson's correlation
1353 coefficients for gene expression across species under the identical environmental
1354 condition. Density plot of pairwise Pearson's correlation coefficients in buds (A)
1355 and leaves (B), calculated from the gene expression in *Q. glauca*, *L. edulis*, and *L.*
1356 *glaber* inhabiting nearly identical environments. Colors indicate rhythmicity
1357 categories. Black bars show median correlation values, and different letters
1358 indicate significant differences (Nemenyi test, $p < 0.05$).

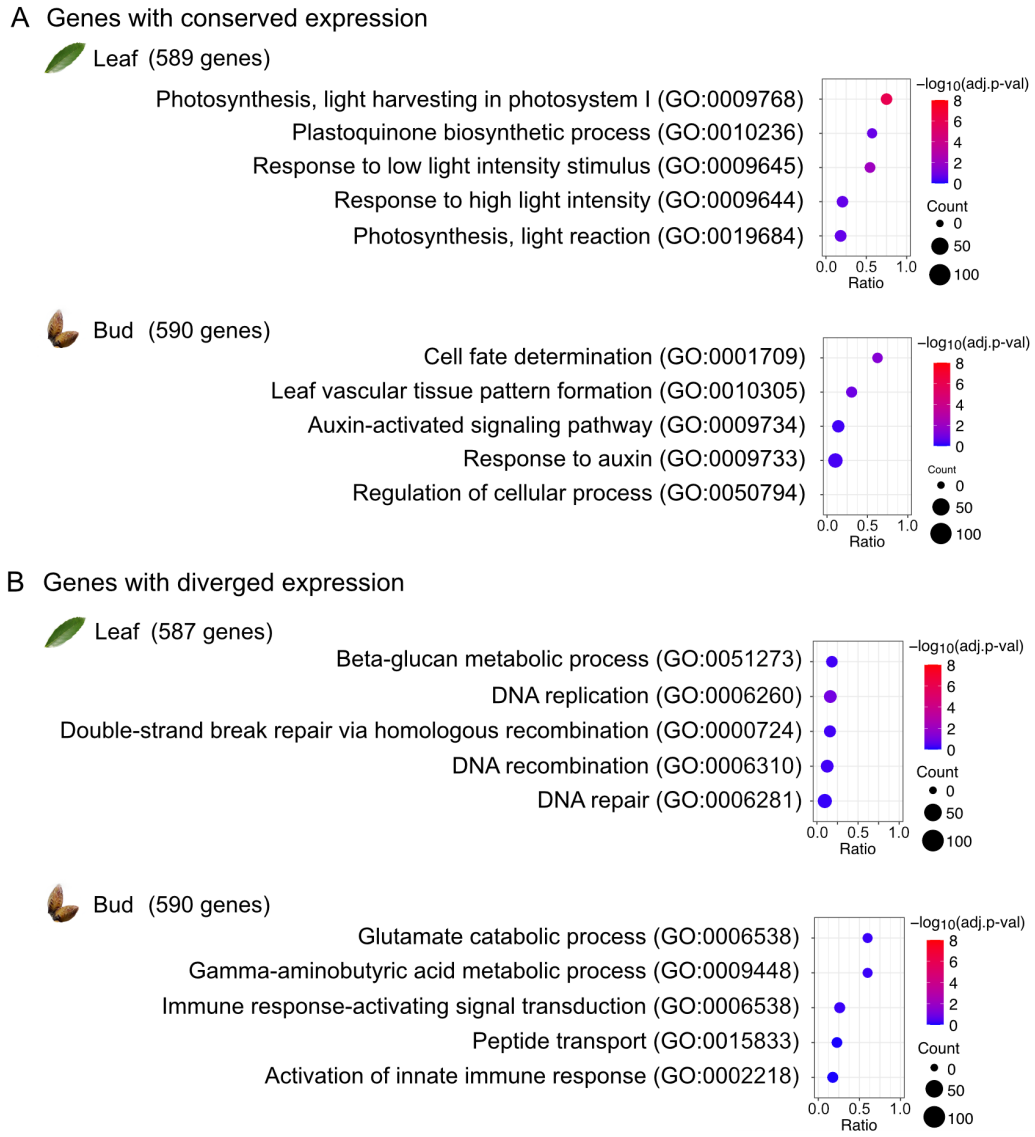


1359
1360

Figure 5–figure supplement 3. Pairwise mean absolute difference of gene expression dynamics. (A–B) Distribution of mean absolute difference of gene expression dynamics (Log_2(GeTMM+1)) in leaf (A) and bud samples (B) across species (Le: *L. edulis*, Lg: *L. glaber*, Qg: *Q. glauca*, Qa: *Q. acuta*; n = 11749). Colors indicate three categories based on the presence or absence of rhythmicity in gene expression. The median values across all the genes were depicted as black bars. Different letters denote significantly different values (Nemenyi test, $p < 0.05$). (C–D) Heatmap of the mean absolute difference for each gene, along with the mean count of genes in four categories of seasonal gene expression: (1) rhythmic genes with annual periodicity peaking in the growing season (March–November), (2) rhythmic genes with annual periodicity peaking in winter (December–February), (3) rhythmic genes with half-annual periodicity, and (4) arrhythmic genes. The mean count was calculated for each block of 100 genes. Genes were ordered based on the mean absolute difference of their expression across all the species pairs.

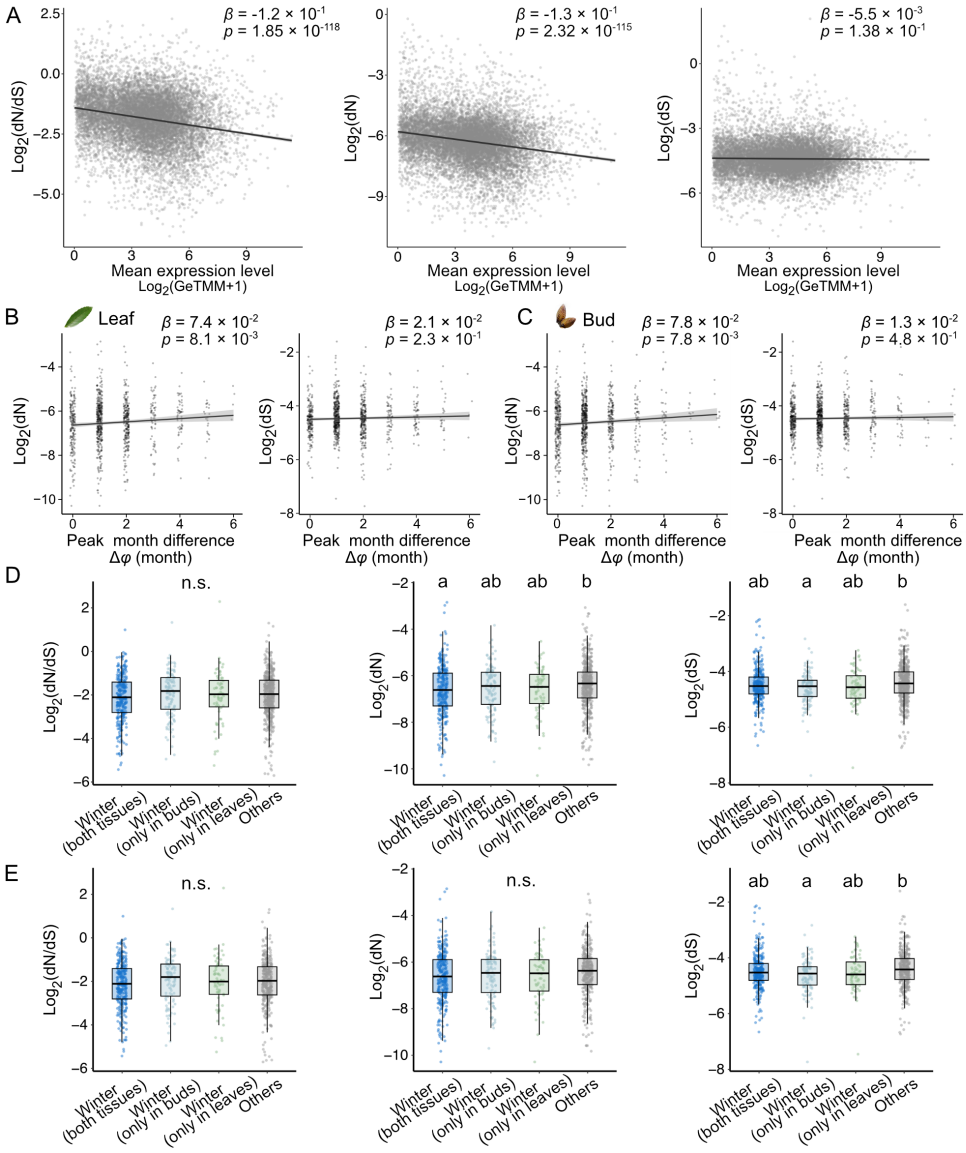


1377 **Figure 5–figure supplement 4.** Gene Ontology (GO) enrichment analysis. (A)
1378 Top 5 GO terms (ranked by the raw *p*-values) for the genes with conserved
1379 expression patterns across species (a mean correlation in gene expression
1380 ranking in the top 5%) in leaves and buds. (B) Top 5 GO terms for the genes with
1381 diverged expression patterns across species (a mean correlation in gene
1382 expression ranking in the bottom 5%) in leaves and buds. Genes used for the
1383 analysis are listed in Supplementary file 6.

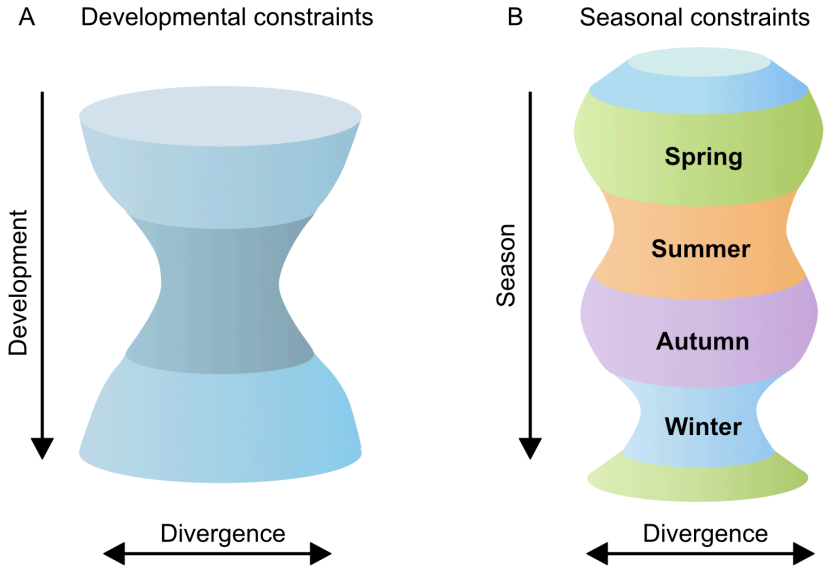


1384
1385

Figure 5–figure supplement 5. Relationship between sequence divergence and gene expression divergence. (A) Relationships between evolutionary measures (dN/dS, dN, dS), and mean expression levels of genes except for the outliers (*n* = 11223). (B–C) Relationships between the evolutionary measures (dN, dS) and peak differences (Δφ) between *Q. glauca* and *L. edulis* in leaves (B) and buds (C) using the same gene set as in Figure 5E (n = 951). The regression slope and corresponding p-value are shown in each panel. Black lines and transparent bands represent regression lines and 95% confidence intervals. (D–E) Comparison of the evolutionary measures (dN/dS, dN, dS) among genes with conserved annual periodicity across four categories (*n* = 951). Genes were classified into four categories: those expressed in winter in both tissues, those expressed in winter only in buds, those expressed in winter only in leaves, and others. (E) shows a subset of (D), including only genes with conserved seasonal peaks (peak month difference < 2).



1401 **Figure 5–figure supplement 6.** Developmental constraints and seasonal
1402 constraints in the evolution of gene expression. (A)The hourglass model of
1403 developmental constraints, illustrating similar gene expression during the mid-
1404 embryonic stage (40–45). (B) Seasonal constraints identified in this study. Similar
1405 gene expression patterns were observed during the winter season in both leaves
1406 and buds, as well as during the summer in leaves. The vertical axis represents
1407 time (developmental stages or seasons), while the horizontal axis represents the
1408 divergence of gene expression patterns.



1409
1410

Supplementary files

Supplementary file 1

Statistics for genome assemblies of two new sequenced species

Supplementary file 2

Summary of clustering 213 gene expression profiles

Supplementary file 3

Hierarchical clustering of 11749 gene expression dynamics

Supplementary file 4

Identification of genes with rhythmic expression

Supplementary file 5

Comparison of correlation and absolute difference among the species pairs

Supplementary file 6

List of genes with conserved and diverged gene expression across species.

Supplementary file 7

Statistical test of evolutionary rates