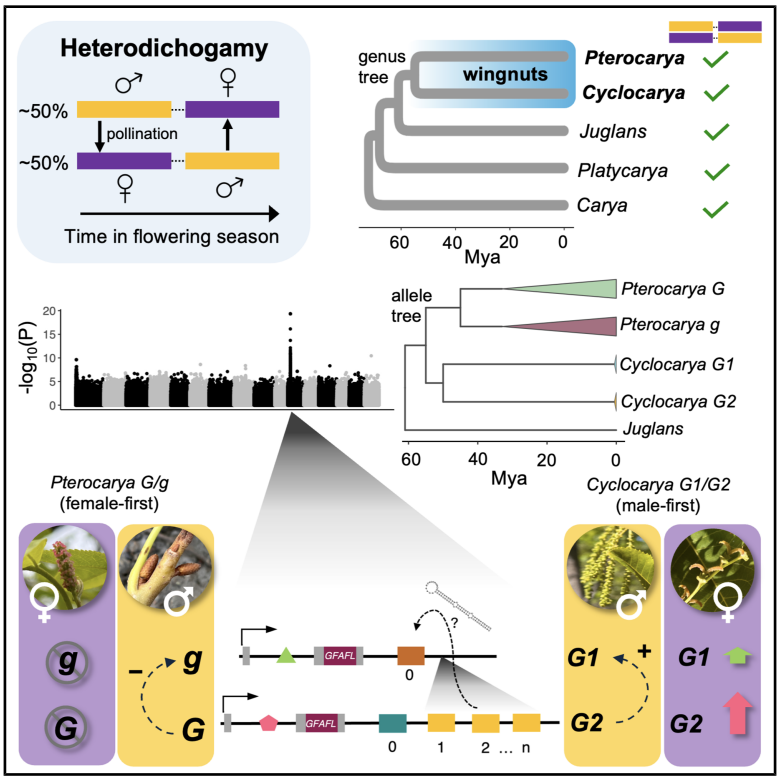
Article

Current Biology

Distinct haplotypes and reversed dominance at a single-gene balanced polymorphism controlling heterodichogamy in two genera of wingnuts

Graphical abstract



Highlights

- Heterodichogamy in wingnuts has a distinct inherited basis from other Juglandaceae
- *Pterocarya* and *Cyclocarya* reuse the same gene, *GFAFL*, but show reversed dominance
- Haplotypes across genera are distinct but show parallel structural variation
- Dominant haplotypes regulate recessive *GFAFL* expression in *trans* in male flowers

Authors

Jeffrey S. Groh, Gracie Ackerman, Michael Wenzel, Graham Coop

Correspondence

jgroh@berkeley.edu

In brief

In most species within the walnut family, two genetically determined morphs alternate between male and female flowering phases in time. Groh et al. identify a distinct locus for this dimorphism in two genera of wingnuts with reversed dominance and find *trans*-regulation of one allele by the other in both cases.

Groh et al., 2025, Current Biology *35*, 5192–5209
November 3, 2025 © 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
https://doi.org/10.1016/j.cub.2025.09.061

CellPress

CellPress

Current Biology

Article

Distinct haplotypes and reversed dominance at a single-gene balanced polymorphism controlling heterodichogamy in two genera of wingnuts

Jeffrey S. Groh,[1,2,4,5,*] Gracie Ackerman,[1] Michael Wenzel,[3] and Graham Coop[1,2]
[1]Department of Evolution and Ecology, University of California, Davis, Davis, CA 95616, USA
[2]Center for Population Biology, University of California, Davis, Davis, CA 95616, USA
[3]Sonoma Botanical Garden, Glen Ellen, CA 95442, USA
[4]Present address: Miller Institute for Basic Research in Science, University of California, Berkeley, Berkeley, CA 94720, USA
[5]Lead contact
*Correspondence: jgroh@berkeley.edu
https://doi.org/10.1016/j.cub.2025.09.061

SUMMARY

In the angiosperm mating system of heterodichogamy, two hermaphroditic morphs temporally alternate between male and female flowering phases, promoting disassortative mating between morphs. Heterodichogamy is widespread within Juglandaceae and is controlled by distinct ancient structural variants in *Juglans* and *Carya*. Here, we identify a novel inherited basis in two additional Juglandaceae genera, *Pterocarya* and *Cyclocarya*. In *Pterocarya*, we map dichogamy types to an ~70 kb trans-species structural variant overlapping a single candidate gene, *GFAFL*, with a dominant haplotype for female-first flowering. Heterodichogamy in *Cyclocarya* is associated with a distinct pair of ancient haplotypes at the same locus but with a dominant allele for male-first flowering. Both sets of alleles are ancient (ca. 45 and 50 Ma, respectively), but their divergence postdates the divergence between the genera. *GFAFL* expression in staminate flowers in both systems reveals *trans*-regulation of the recessive allele by the dominant haplotype, which in both systems contains tandem duplicates of sequence in the 3′ region of *GFAFL*. We suggest that reciprocal sex matching in heterodichogamy can result from a combination of *cis*-regulatory divergence and *trans*-regulation involving "fast" and "slow" alleles at a single flowering-time gene, yielding "early" vs. "late" sex-specific flowering. Our findings shed light on the molecular basis of heterodichogamy and contribute to an emerging view that diverse genetic pathways can be co-opted during its evolution and facilitate turnover in its genetic control.

INTRODUCTION

The timing of reproduction is a critical aspect of fitness and is shaped by adaptive evolution.[1–3] In hermaphroditic organisms that can produce both types of dimorphic gametes or gametophytes, the optimal timing of their development, i.e., reproduction through "male" vs. "female" pathways, may differ as a result of resource trade-offs, sex-specific changes in fecundity through time, and costs associated with self-fertilization and gamete interference. These factors are thought to underlie the evolution of diversity in sex-specific temporal mating strategies, such as sequential hermaphroditism in animals and dichogamy (temporal separation of male and female function) in flowering plants.[4–7]

Simultaneously with individual-level trade-offs, sex-specific temporal strategies are also shaped by sex-ratio selection acting within a population.[8,9] Heterodichogamy, found in at least 13 angiosperm families, refers to a stable dimorphism for the temporal order of staminate vs. pistillate flowering (i.e., male and female).[10,11] Protandrous (PA; male-first) individuals release pollen prior to their stigmas being receptive, whereas protogynous (PG; female-first) individuals show the complementary pattern. Pollen dispersal of one morph thus occurs synchronously with stigma receptivity of the other. This promotes disassortative outcrossing,[12] such that the rare mating type is favored and a 1:1 ratio of two morphs is expected at equilibrium.[13] This mating system is thus analogous to dioecy and heterostyly, in which discrete floral morphs are maintained by negative-frequency-dependent selection resulting from disassortative mating between morphs.[14,15] Heterodichogamy is suggested to facilitate the evolutionary transition from hermaphroditism to dioecy,[16–18] yet it can also persist stably over long evolutionary timescales.[19] Although this mating system has likely evolved at least 13 times,[11] to date its genomic basis has been characterized in very few systems.[19–21] As with dioecy and heterostyly, heterodichogamy appears to commonly be controlled by single-locus mechanisms, where the dominant haplotype contains sequence that is not shared by its counterpart (i.e., typically hemizygous) and with limited recombination between haplotypes,[19,20] suggesting generalities regarding the regulation and genomic consequences of different mating systems. However, the mechanisms of gene regulation controlling heterodichogamy remain

 © 2025 Elsevier Inc.
All rights are reserved, including those for text and data mining, AI training, and similar technologies.

poorly understood and may show distinct features from other dimorphic mating systems.

A classic example of heterodichogamy is its widespread occurrence in the walnut family (Juglandaceae).[22–25] This system is present in the many species of walnuts (*Juglans*) and hickories (*Carya*, e.g., pecan) examined to date, and in both genera it is controlled by Mendelian segregation of a dominant allele for protogyny (*G* dominant, *g* recessive).[13,26] In our previous work, we investigated the molecular basis and evolutionary history of heterodichogamy within Juglandaceae and found that two distinct genomic regions, with no shared homologous genes, control the mating system in *Juglans* and *Carya*.[19] The *Carya* locus resembles a supergene, with a region of reduced recombination spanning ~20 protein-coding genes, several of which have predicted roles in flowering processes. The *Juglans* system involves regulatory divergence of one clear candidate gene (*TPPD1*) linked to trehalose-6-phosphate signaling, a known flowering-time pathway.[27] Both systems are ancient (ca. 40 and 59 mya) (Figure 1D), yet the loci span relatively narrow genomic regions (~20–400 kb), unlike some heteromorphic sex chromosomes. Within Juglandaceae, heterodichogamy has been reported in two additional genera (*Cyclocarya* and *Platycarya*)[24,25]; however, we previously showed that the genetic basis of the mating system in these genera was not shared with either *Juglans* or *Carya*,[19] hypothesizing that additional genetic mechanisms control heterodichogamy in these other genera.

Here, we focus on the flowering behavior of two Juglandaceae genera known as wingnuts—*Pterocarya*, in which heterodichogamy had not been previously documented to our knowledge, and the sister genus *Cyclocarya* (wheel wingnut). Wingnuts are so named for their fruits (Figure 1D), which are suited for wind dispersal, unlike the animal-dispersed husked seeds of *Juglans* and *Carya*. As in *Juglans* and *Carya*, unisexual flowers of *Pterocarya* and *Cyclocarya* are born in separate staminate and pistillate inflorescences (Figures 1A and 1B) or, rarely, in the same inflorescence[28] (Figures S1A and S1B), and pollen is wind dispersed. Macrofossils of *Cyclocarya* appear in the early Paleocene,[29] whereas fossil *Pterocarya* fruits are not known prior to the Oligocene.[30,31] The fossil diversity and distribution of these genera throughout Neogene sediments of North America, Europe, and Asia indicate that wingnuts were once considerably more diverse and geographically widespread than today.[31,32] *Pterocarya* remains present in the fossil record of North America and Europe through the Pliocene, and appears to have gone extinct in these regions during the Pleistocene glaciations.[33,34] The current distribution is accordingly thought to reflect locations of glacial refugia; of the six recognized species, five are restricted to East Asia and a single species (Caucasian wingnut, *P. fraxinifolia*) is endemic to the Caucasus region. *P. stenoptera* is the most widespread species, being found across China, and it is used in urban landscaping in both Europe and North America. *Cyclocarya* contains a single modern representative native to China (*C. paliurus*), which consists of both diploid and autotetraploid lineages that diverged ~570,000 years ago.[35,36]

We combined *de novo* chromosome-level assemblies, resequencing of phenotyped flowering individuals, and gene expression data in order to characterize the evolutionary history and molecular basis of heterodichogamy in *Pterocarya* and *Cyclocarya*. Our study identifies a novel inherited basis underpinning this mating system in *Pterocarya* and parallel use of the same gene in *Cyclocarya*, with a different set of alleles and a reversed-dominance pattern. Together with a concurrent study from Liu et al.[37] on heterodichogamy in *Pterocarya*, our results inform models for the genetic regulation of heterodichogamy and further contribute to a picture of dynamic genetic control of heterodichogamy in Juglandaceae, despite the evolutionary stability of the mating system and conserved inheritance mechanisms within genera.

RESULTS

Pterocarya exhibits heterodichogamous flowering

We studied flowering of 89 individuals of *Pterocarya stenoptera* from germplasm collections, urban plantings, and "volunteer" trees to assess the relative timing of development of staminate and pistillate flowering structures. Individuals generally showed clear dichogamy of two types. Protandrous (male-first) individuals (Figure 1A) released pollen prior to stigma maturity, whereas protogynous (female-first) individuals showed mature stigmas prior to pollen release (Figure 1B). Among plants at the same locality, these timing differences resulted in the reciprocal matching of pollen release of one morph with stigma receptivity of the other.

The offset timing of staminate and pistillate flowering of the two morphs (approximately 1 week) was reflected in measurements of inflorescence size (Figure 1C). Female inflorescences, which progressively lengthen throughout flowering and post fertilization, are longer in protogynous individuals mid-season. On the other hand, male inflorescences, which begin development in the season prior to flowering, enter dormancy at a larger size in protandrous individuals (Figure 1C), a pattern that has also been observed in heterodichogamous *Cyclocarya*,[38] *Juglans*,[19,39] and *Carya*.[40]

Protandrous and protogynous morphs were common across sites and segregate among maternal siblings from mother trees from different collection sources (Figure S1C). Among all flowering individuals examined across sampling localities, we identified 36 protandrous individuals, 36 protogynous individuals, and 17 individuals whose dichogamy type could not be assessed due to limited observation.

We observed flowering in a single individual from each of two sister species, *P. macroptera* and *P. rhoifolia*. The *P. macroptera* individual showed protogynous flowering, whereas the *P. rhoifolia* individual exhibited protandrous flowering. Segregating dichogamy types in two clades spanning the deepest divergence within *Pterocarya* (Figure 1D), together with the prevalence of the mating system throughout Juglandaceae, strongly suggests that heterodichogamy occurs in all species of *Pterocarya*.

Genome assembly and annotation

To facilitate investigation into the molecular basis and evolutionary history of heterodichogamy in wingnuts, we generated haplotype-resolved chromosome-scale genome assemblies from three species of *Pterocarya* and from their sister group, *Cyclocarya paliurus*, using PacBio HiFi sequencing and scaffolding to Hi-C aided long-read assemblies from the same genera.[41,42] *Pterocarya* assembly samples originated from a single flowering

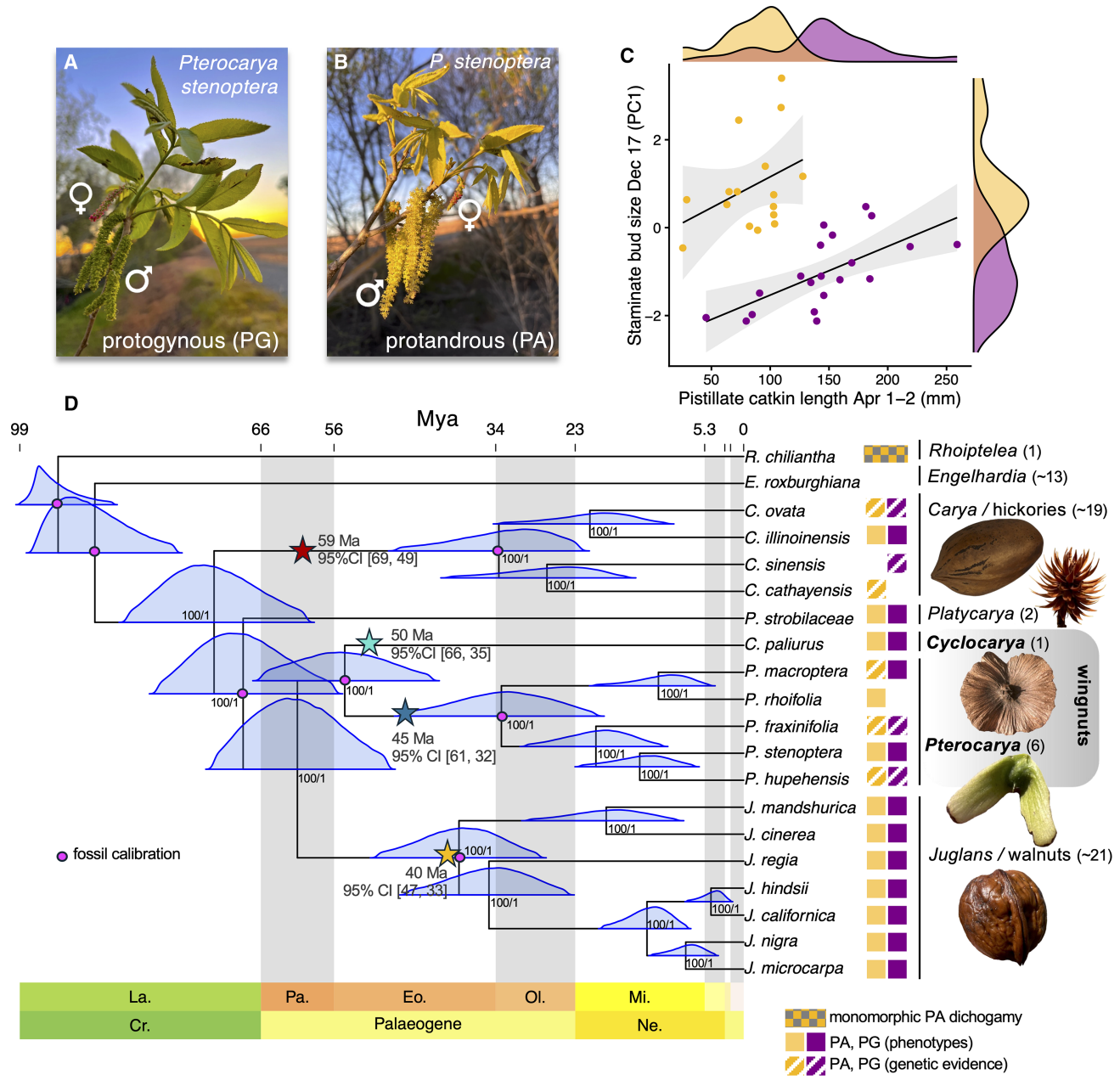


Figure 1. Heterodichogamy in *Pterocarya* and the phylogeny of Juglandaceae

(A and B) Individuals of *Pterocarya stenoptera* show either protogynous (A) or protandrous (B) dichogamy. In (A), mature stigmas (pink) are visible on the pistillate inflorescence, while anthers on the staminate inflorescence have not yet released pollen. In (B), anthers have released pollen, and pistillate flowers are beginning to reach maturity.

(C) Protandrous (gold) and protogynous (purple) individuals show reversed dimorphism for developmental stage of staminate (male) and pistillate (female) inflorescences. Density plots for each inflorescence type are shown at top and right. Developmental size of staminate buds (PC1 from length and width) and pistillate inflorescences are overall negatively correlated due to the dimorphism but are positively correlated within each type.

(D) Fossil-calibrated phylogeny of Juglandaceae. Time (mya) shown at top with abbreviations of geological periods and epochs at bottom. Full posterior distributions on node ages are shown in blue. Node labels indicate bootstrap support from maximum-likelihood analysis and local posterior probability values from ASTRAL. Pink circles indicate nodes with priors specified on the basis of fossil information (STAR Methods; Figures S1C and S1D). Colored rectangles at right indicate the type of dichogamy observed in each taxon. Absence of colored rectangles means dichogamy types have not been reported. Full genus names and common names (where relevant) are written on the right with numbers of accepted species in parentheses and a representative seed structure. Colored stars indicate the origins of separate genetic systems controlling heterodichogamy in Juglandaceae, with estimates and 95% credible intervals.

Photos in (A), (B), and (D) by J.S.G.

See also Figure S1 and Data S1.

individual each of *P. stenoptera*, *P. macroptera* var. *insignis* (both protogynous flowering types), and a non-reproductive *P. fraxinifolia* individual of unknown flowering type. The diploid *C. paliurus* individual from Sonoma Botanical Garden (formerly Quarryhill Botanical Garden) used for assembly showed protogynous flowering and is one of two known trees of this species established in North America.[43]

Our assemblies are highly contiguous; for example, haplotype 1 of *P. stenoptera* has a contig N50 of 32.14 Mb, with 97.9% of the 551.2 Mb assembled genome contained within 16 chromosome-scale scaffolds, the expected number from cytological counts,[44] and an average of 2.38 contigs per chromosome. After masking repetitive sequences (54% of the genome), we annotated our *P. stenoptera* assemblies, leveraging extrinsic evidence from protein homology and RNA sequencing (RNA-seq) datasets from 7 tissues (data from this study and Zhang et al.[41]), resulting in 30,681 gene models for haplotype 1 and a BUSCO completeness score of 98.4% (OrthoDBv10, eudicots). Other assemblies showed similarly favorable metrics (Data S1A).

A time-calibrated phylogeny of Juglandaceae

Although *Pterocarya* is well-resolved as monophyletic, recent studies using different molecular data types have reported conflicting intrageneric relationships.[31,45] Furthermore, although the *Pterocarya-Cyclocarya* divergence has been assumed to have occurred by the early Paleocene, based on *Cyclocarya* fossils (e.g., Lyson et al.[29] and Zhang et al.[46]), the lack of definitive *Pterocarya* fossils prior to the Oligocene[30] (30 Ma later) leaves considerable uncertainty over the crown age of extant *Pterocarya*. The extensive fossil record of Juglandaceae[47] presents an opportunity to place our findings in a robust time-calibrated phylogenetic framework together with other Juglandaceae heterodichogamy systems.[19] We therefore used genome-wide data (5,089 orthologs) for 7 of the 9 extant genera *sensu* APG IV[48] to infer a species phylogeny (using both ASTRAL and maximum likelihood) and performed Bayesian time calibration[49] using fossil-informed prior distributions on internal nodes (STAR Methods).

The time-calibrated phylogeny (Figure 1D; Data S1B) supports an origin of crown Juglandaceae in the Late Cretaceous (94 mya, 95% credible interval [CI] 86–99 mya), with a subsequent divergence of the various tribes occurring near the K-Pg boundary. This topology shows intergeneric relationships largely consistent with previous phylogenies of Juglandaceae[50–52] and resolves *Carya* rather than *Platycarya* as the outgroup within subfamily Juglandoideae. The *Pterocarya-Cyclocarya* divergence is estimated at 55 mya (95% CI 43–66 mya), despite specifying a prior favoring a minimum divergence of 65 mya (with soft bounds) on the basis of the earliest known Paleocene *Cyclocarya* fossil[29] (Figure S1E). This signal in the molecular data favoring a younger *Cyclocarya-Pterocarya* divergence, together with morphological divergence between Paleocene and extant *Cyclocarya*,[28] suggests the possibility that the oldest *Cyclocarya* fossils may represent stem lineages to both modern genera and that *Pterocarya* could have evolved from an ancestor with *Cyclocarya*-like fruit morphology. Within *Pterocarya*, we infer an early Oligocene or late Eocene crown age (31 mya, 95% CI 20–47 mya) and subsequent radiation during the Miocene epoch. Our results support reciprocal monophyly of the morphologically recognized intrageneric sections *Pterocarya* (*P. stenoptera*, *P. hupehensis*, and *P. fraxinifolia*) and *Platyptera* (*P. macroptera* and *P. rhoifolia*),[53] in agreement with Song et al.[45]

Ancient haplotypes at a single locus control heterodichogamy throughout *Pterocarya*

To identify the genetic basis of dichogamy type within *P. stenoptera*, we resequenced 20 phenotyped individuals (Illumina, mean depth 27×), combined with existing data from 10 additional individuals[19] with newly assigned dichogamy phenotypes. Our mapping population consists of a mixture of wild-derived seed collections and recently wild-derived material from several sources (Data S1C) and contains multiple sets of maternal siblings segregating for dichogamy type (Figure S1C).

Using a genome-wide association study (GWAS) with 11,663,910 called SNPs, we identified a single clear association peak on chromosome 11 (Figure 2A), consistent with a single-locus Mendelian inheritance mechanism. This region is not syntenic with *Juglans* chromosome 11 (Figure S2A) and does not contain homologs of any of the genes present in the *Juglans* or *Carya* heterodichogamy loci. The association peak is contained within single assembly contigs in each *P. stenoptera* haplotype assembly and does not appear close to the putative centromere (Figure S2B). The peak contains a set of SNPs perfectly predictive of dichogamy type spanning a 65.5 kb region in haplotype 2 of the assembly (Figures 2B and S3A) (vs. 77.2 kb in haplotype 1; Figures S3B–S3E). We refer to this as the G-locus. Gene annotation identified a single reliably predicted protein-coding gene within this region of haplotype 2 (STAR Methods), a member of the *FANTASTIC-FOUR* (*FAF*) superfamily (red; Figure 2B), with a reciprocal ortholog at the syntenic location in haplotype 1. We return to a possible functional role of this gene below.

Genotypes at these SNPs are consistent with a dominant allele (*G*) determining protogyny and indicate that our haplotype assemblies 1 and 2 represent the recessive (*g*) and dominant (*G*) alleles, respectively. In our GWAS sample, all 15 protandrous individuals are *gg* homozygotes, and, of the 15 protogynous individuals, 14 are *Gg* heterozygotes while one is a *GG* homozygote. This deficit of homozygotes for the dominant allele is expected to result under disassortative mating. We tested this hypothesis by combining our data (15 PA, 15 PG, 1 unknown) with published resequencing data from nine additional *P. stenoptera* individuals sampled from its current natural range.[54] Consistent with disassortative mating between morphs, we identified only two dominant allele homozygotes (out of 40 total), a significant deviation away from the Hardy-Weinberg expectation under random mating ($p = 1.27 \times 10^{-8}$, $\chi^2 = 32.38$, d.f. = 1). This test is conservative with respect to population structure, which would increase the proportion of homozygotes in the sample.

Patterns of read depth across this region showed clear stratification by phenotype (Figure 2C). Protogynous heterozygotes appear hemizygous for a set of indels with respect to both assembled haplotypes, whereas protandrous individuals almost completely lack read depth across the entirety of the *G* haplotype but show typical read depth across the *g* haplotype (Figure S3F). These patterns point to structural differences and/or high divergence between the two haplotypes. Indeed, compared with the genome-wide average level of polymorphism per base pair ($\pi = 0.008$), the divergence between haplotypes is

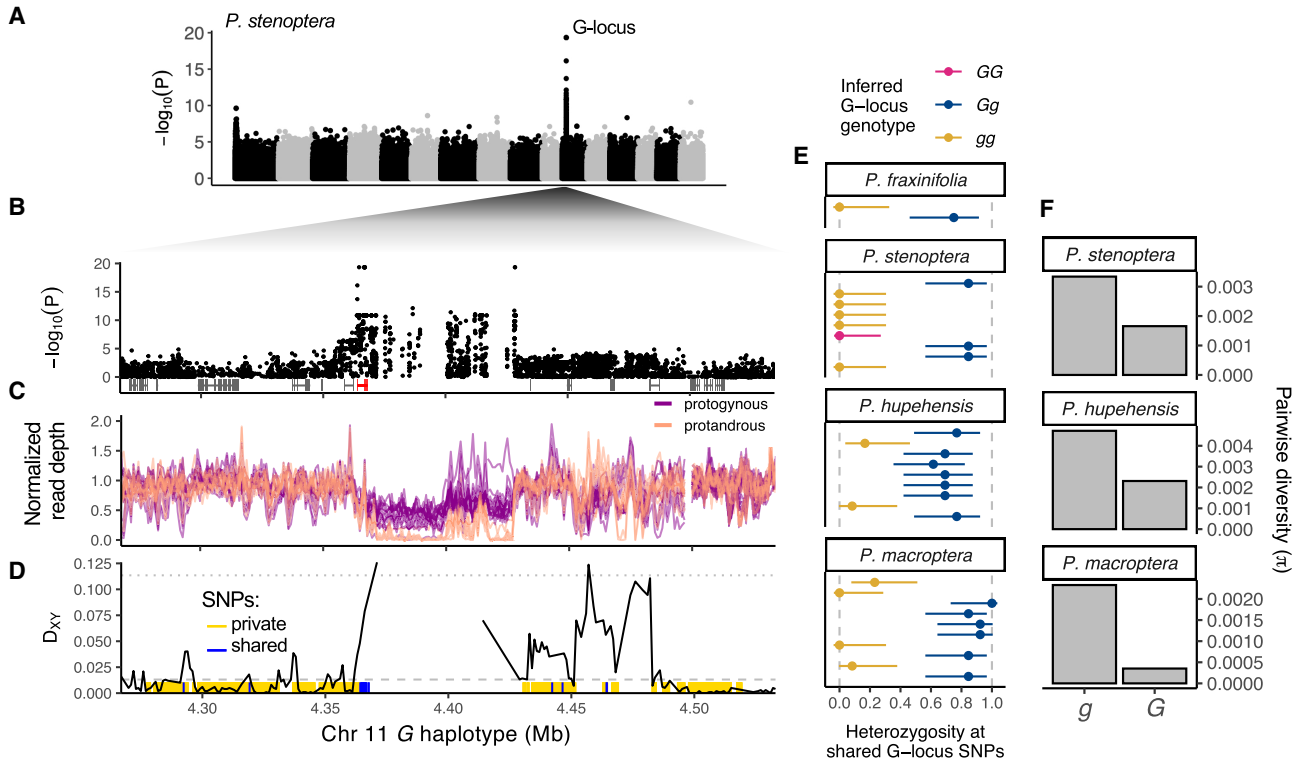


Figure 2. A single genetic locus controls heterodichogamous mating types across *Pterocarya*

(A) GWAS for dichogamy type in 30 individuals of *P. stenoptera* reveals a single clear peak on chromosome 11 (shown for reads aligned to haplotype 2, which contains the dominant allele). Chromosomes are arranged in increasing number, alternating black and gray.

(B) Zoomed-in view of the GWAS peak (G-locus). Predicted gene models are shown at bottom with a *FAF*-like gene (*GFAFL*) in red.

(C) Normalized average read depth in 1 kb windows across the G-locus. Lines are colored by each individual's dichogamy type.

(D) Nucleotide divergence in 1 kb windows across the region. Dashed and dotted lines show the genome-wide mean and 99% quantile, respectively. Colored tick marks at bottom show positions of SNPs from a comparison of *P. stenoptera* and *P. macroptera* haplotypes where alleles group by species (gold) or are shared across species and group by haplotype (blue).

(E) Individuals in four *Pterocarya* species from published datasets[19,54] segregate for the shared SNP haplotypes identified in (D) between *P. stenoptera* and *P. macroptera*. Bars show 95% confidence intervals of the proportion (Agresti-Coull method).

(F) Nucleotide diversity is reduced among *G* haplotypes compared with *g* haplotypes in three species.

See also Data S1 and Figures S1–S3 and S6.

exceptionally high in this region (D_{xy} = 0.126), falling in the top 1% of windowed 1 kb values (for windows with >500 aligned bp and <50% repeat content) genome wide (Figure 2D). High divergence between the two assembled haplotypes extends beyond the GWAS peak in a region of high repeat content and could in part reflect alignment error.

The high divergence across the G-locus suggests that it could be an ancient balanced polymorphism. We leveraged our haplotype assemblies from the protogynous *P. macroptera* individual and identified a high concentration of *trans*-species polymorphic SNPs at the GWAS peak (blue; Figure 2D), whereas SNPs immediately adjacent to the GWAS peak group by species (yellow; Figure 2D). We next examined heterozygosity at this set of *trans*-species polymorphic SNPs in published whole-genome resequencing data from multiple individuals of each of three species (*P. stenoptera*, *P. macroptera*, and *P. hupehensis*).[54] We find multiple individuals within each species that are highly heterozygous at this set of SNPs, as expected if these individuals were heterozygous (*Gg*) for the same pair of haplotypes, while others showed low heterozygosity across these SNPs, as expected if they were homozygous for one haplotype (Figure 2E). Furthermore, the latter were homozygous for the recessive allele SNP (*gg*) seen in *P. stenoptera* and *P. macroptera*, with only one homozygote for the dominant allele (*GG*) seen in this dataset (in *P. stenoptera*). These genotypes suggest that the phenotypic morphs may be close to their 1:1 predicted ratio under disassortative mating in these species. We also detected shared polymorphism at these SNPs in one of two previously sequenced individuals of *P. fraxinifolia*[19] (Figure 2E) and confirmed that a previously sequenced protandrous *P. rhoifolia* individual lacked shared polymorphism at these SNPs. We used estimates of species divergence times from the genome-wide phylogeny— together with divergence at 4-fold-degenerate coding sites or intronic sequences—to calibrate a molecular clock for this region (described in STAR Methods), inferring that the *Pterocarya* haplotypes diverged 45 mya (95% CI: 32–61 mya) in the common ancestor of all extant *Pterocarya*.

The region tightly linked to the *G* allele is predicted to share evolutionary features with Y-linked regions in XY sex-determination systems (or W-linked regions in ZW systems). As

disassortative mating maintains the *G* haplotype at lower frequency compared with the *g* haplotype, under a neutral model we expect a 3-fold reduction in diversity among *G* vs. *g* haplotypes with perfect disassortative mating. In addition, the physical extent of conserved *trans*-species polymorphism indicates a lack of historical recombination between *G* and *g* haplotypes, and *G* may also rarely recombine with itself because *GG* homozygotes are rare. Thus, effects of Muller's ratchet and/or hitchhiking should be stronger for *G* haplotypes. Using phased resequencing data across the coding sequence of the *FAF*-like gene, we find support for these predictions, with *G* haplotypes showing a 2- to 6.6-fold reduction in several species (Figure 2F). Y-linked regions and other permanently heterozygous supergenes often accumulate repetitive sequences at higher rates than autosomally inherited regions.[55–58] Despite the fact that the *G* haplotype is nearly always heterozygous, the *G* and *g* haplotypes are of similar lengths in both *Pterocarya* species. G-locus haplotypes are only modestly enriched in the percentage of repeat content compared with their local genomic context, and this enrichment is not stronger for the *G* haplotype (Figure S2C). The lack of considerable expansion of the dominant haplotype suggests that the degenerative impacts of heterodichogamous mating on molecular variation are limited in *Pterocarya*.

Allele-specific regulation of a single flowering gene at the *Pterocarya* G-locus

The region of *trans*-species G-locus SNPs coincides closely with the transcribed region of a FAF-domain-containing gene, which we refer to here as *G-LOCUS-FAF-LIKE* (*GFAFL*) (Figure 3A). The FAF domain (Figure S4A) is a plant-specific domain of unknown molecular function named for a group of four Arabidopsis genes identified as regulators of shoot meristem size.[59] Other FAF-domain-containing (*FAF*-like) genes have since been found to regulate diverse developmental processes, including abscisic acid signaling, abiotic stress tolerance, endosperm starch metabolism,[60–62] and flowering time.[63] We identified *FAF*-like paralogs in several embryophytes and found that all previously functionally characterized members of this group are distant paralogs of *GFAFL* (Figure 3C). Nonetheless, orthologs of *GFAFL* in *Cyclocarya*, *Juglans*, and Arabidopsis show clear expression in flowers or pollen and little or no expression in other tissues (Figures S4C–S4H), supporting the idea that *GFAFL* orthologs are plausibly involved in flowering.

In the coding region of *GFAFL*, we see only one nonsynonymous SNP difference between haplotypes that is polymorphic in both *P. stenoptera* and *P. macroptera*—a nonconservative K-to-N replacement in the recessive haplotype lineage at an otherwise conserved amino acid site (Figure S4A). Within species, the haplotypes show greater divergence across several conserved regions; the *P. macroptera* haplotypes are ~2.5% diverged at the level of amino acid sequence ($\frac{dN}{dS}$ = 0.58), while the *P. stenoptera* haplotypes are (~8%) diverged in amino acid sequence ($\frac{dN}{dS}$ = 0.65). Using phased resequencing data, we did not detect any signal of adaptive molecular evolution, as there was no excess of nonsynonymous vs. synonymous substitutions between the haplotypes using nonsynonymous and synonymous polymorphism among *g* haplotypes as a control (MacDonald-Kreitman test).

To investigate the role of gene-expression regulation in the developmental basis of heterodichogamy, we performed RNA-seq on both staminate (male) and pistillate (female) flower tissues of both morphs collected on the same day during flowering (March 20), as well as on pre-dormant staminate catkin buds (on August 2) in the same year and pistillate buds just prior to bud break in the following season (February 18). These tissues show distinct genome-wide expression patterns across biological replicates (Figure S5A), and we see a signal of transcriptome-wide differentiation between morphs within mature flowering tissues (Figure S5B–S5D). Examining RNA-seq read alignments from both reproductive and vegetative tissues,[19,41] we found that *GFAFL* is the only gene in the G-locus region with (1) unequivocal evidence of expression in floral tissue and (2) conserved presence across multiple G-locus haplotypes of *Pterocarya* (STAR Methods).

In *P. stenoptera*, we found expression of *GFAFL* in staminate catkins and no other tissues (Figure S5E), with a large increase in expression between the bud stage and floral maturity (Figure 3B). Neither overall expression levels nor splicing patterns of the gene appear to differ between morphs. However, we identified patterns of allele-specific expression using G-locus SNPs within RNA-seq reads. In early development of staminate flowers, heterozygotes for the G-locus show a strong bias toward expression of the *G* allele of *GFAFL* ($p < 0.01$, $\chi^2 = 7.11$, d.f. = 1). Indeed, expression of *g* in heterozygotes is entirely lacking, which is significantly less than what would be expected compared with half the expression level seen in *gg* homozygotes ($p < 0.001$, $\chi^2 = 15.11$, d.f. = 1). The highest expression of *G* was seen in the only *GG* homozygote, which could suggest a dosage-sensitive effect, but this effect is weak and not statistically significant. In mature staminate flowers, the pattern of allele-specific expression in heterozygotes appears to weaken or even reverse, and the level of expression of *g* is not reduced compared with its expectation from the *gg* homozygote.

In the early stages of staminate flower development, despite the lack of *g* expression of *GFAFL* in heterozygotes, we see clear expression of the *G* copy (Figure 3B). These facts suggest a mechanism of allele-specific *trans*-silencing of the *g* copy of *GFAFL* by the *G* haplotype. This result is further supported by independent data by data from a concurrent study by Liu et al.,[37] which identified the same candidate gene for heterodichogamy in *Pterocarya* using different methods. Furthermore, the apparent lack of expression of *GFAFL* in both early stage (i.e., before bud break) and mature pistillate flowers suggests that G-locus haplotypes might regulate the female timing difference through differential regulation of additional genes not found in the GWAS peak.

Heterodichogamy in *Cyclocarya* is also associated with *GFAFL* but with reversed dominance

The protogynous *C. paliurus* individual we used for assembly is not heterozygous for the two *Pterocarya* dichogamy haplotypes described above, nor is it homozygous for the dominant haplotype seen in *Pterocarya*. This implied that there must be a distinct inheritance mechanism for heterodichogamy in *Cyclocarya*. Remarkably, we find evidence that a distinct pair of haplotypes at the same locus (*G1* and *G2*) with reversed dominance control heterodichogamy in both the diploid and tetraploid lineages of *Cyclocarya*.

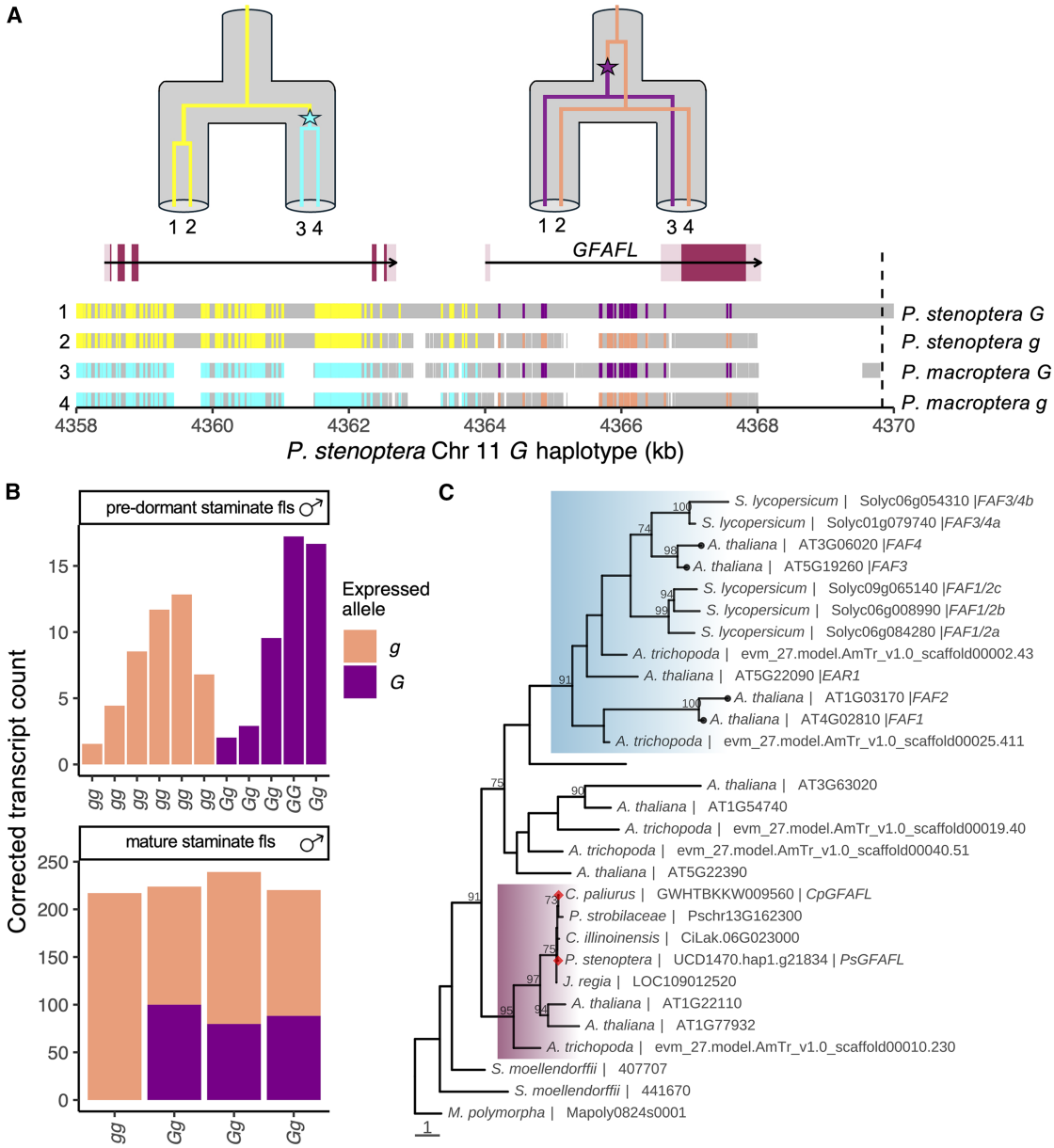


Figure 3. *Trans*-species polymorphism and allele-specific regulation at *GFAFL* underlie heterodichogamy in *Pterocarya*

(A) Outside the region of copy number variation, *trans*-species polymorphic SNPs at the *Pterocarya* G-locus localize to the transcribed region of *GFAFL*. Each horizontal bar represents a haplotype alignment to the *P. stenoptera G* reference, with gray portions showing aligned regions. Positions with colored tick marks are biallelic SNPs (singletons not shown) with the alleles in two different colors. Above, we show the implied tree sequence, noting that across sites either allele may be the derived mutation. The dotted vertical line on the right indicates the start position of the region homologous to the repeats contained within the dominant haplotype (*G-0*).

(B) *GFAFL* shows allele-specific expression in developing staminate flowers. Each vertical bar represents a unique individual whose genotype is shown at the bottom, as determined from genomic sequence data. The total height of each bar represents the combined transcript counts of both *GFAFL* alleles, and the contribution of each allele is represented by the height of the portion shaded purple (*G*) or orange (*g*).

(C) Maximum-likelihood phylogeny of *FAF*-like genes in a selection of embryophytes from a codon-based alignment of the *FAF* domain, rooted using a single-copy gene in *Marchantia*. The subclade highlighted in blue contains the canonical Arabidopsis *FANTASTIC FOUR* genes (black points at tips). *GFAFL* is a member of a different subclade (maroon), which we infer is related to the previous via gene duplication in the ancestor of angiosperms. Copies in *Pterocarya* and *Cyclocarya* are indicated with red diamonds. Alignment across a larger coding region in this subclade reveals that *GFAFL* shows closer homology to the Arabidopsis gene *AT1G22110* than to *AT1G77932* (Figure S4A). Node labels give bootstrap support (only shown for values >70%). Branch lengths represent substitutions per codon.

See also Figures S4 and S5 and Data S1.

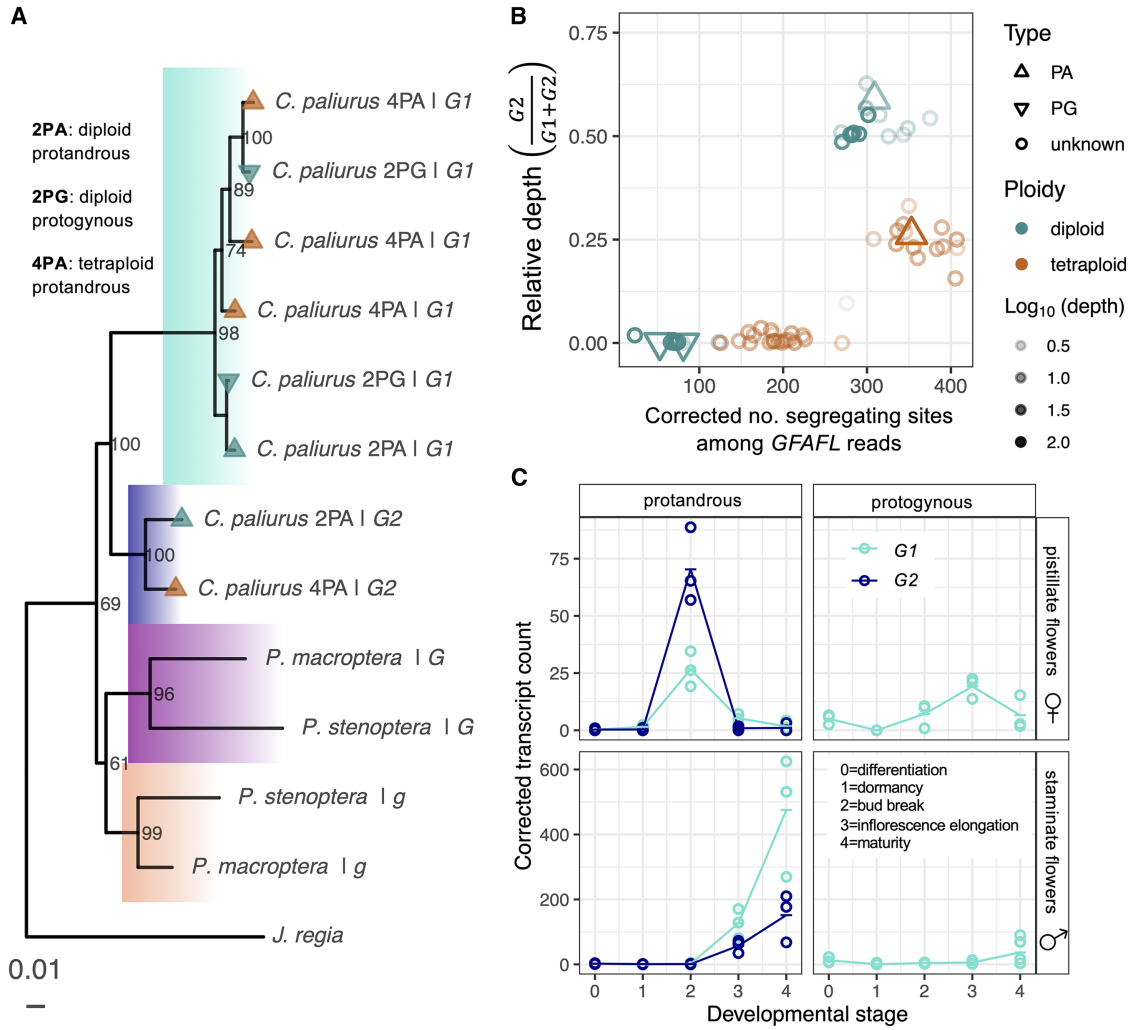


Figure 4. A distinct pair of ancient *GFAFL* alleles in *Cyclocarya* are associated with dichogamy type and show allele-specific regulation in flowers

(A) Phylogeny of the *GFAFL* intron across *Pterocarya* and *Cyclocarya* assemblies, rooted with *Carya illinoinensis* as an outgroup (not shown). Tip labels from left to right indicate species name, followed by ploidy and phenotype of the individual used for assembly (*Cyclocarya* only), followed by G-locus allele (*Pterocarya* and *Cyclocarya* only). Tip shapes and colors further indicate the flowering type and ploidy of the individual from which each *Cyclocarya GFAFL* copy derives; see legend for (B).

(B) Segregation of the *G2* allele in both diploid and tetraploid lineages of *C. paliurus*. Each point represents a unique individual. The *y* axis shows relative read depth across alternative alleles of the *GFAFL* transcribed region from a competitive mapping to both alleles. Diploids show either zero or $\sim\frac{1}{2}$ relative depth over *G2*, while tetraploids show either zero or $\sim\frac{1}{4}$. The *x* axis shows ascertained segregating sites from short reads aligned to an assembly containing a single copy of *GFAFL*, with a correction applied for variable sequencing depth. Levels of variation among the *GFAFL* sequences within individuals are much higher for individuals carrying a copy of the *G2* allele. The higher numbers of segregating sites in tetraploids vs. diploids can be explained by the sampling of four vs. two total copies of *GFAFL*.

(C) Expression of the two *GFAFL* alleles across both pistillate and staminate flower development in protandrous and protogynous morphs of *C. paliurus*. Points show biological replicates and lines connect means across replicates.

Data reanalyzed from Qu et al.,[35] with three points in right two panels added in this study.

See also Figures S6 and S7.

In published RNA-seq data from *C. paliurus*,[64] we detected flower-specific expression of two deeply diverged copies of *GFAFL* orthologs (termed *G1* and *G2*), both of which map to the syntenic location of *GFAFL* in our assembly and previous assemblies[42] of this species. Divergence at 4-fold-degenerate sites and intronic sequence (STAR Methods) gives a divergence-time estimate for these alleles of 50 mya (95% CI 35–66 mya), younger than the inferred divergence time between *Cyclocarya* and *Pterocarya* (55 Ma, 95% CI 43–66 mya). These alleles do not differ at either of the two conserved *GFAFL* coding-sequence polymorphisms ancestral to *Pterocarya* (Figure 3A), and they cluster together as a separate clade, sister to the *Pterocarya* alleles, in a phylogeny of the *GFAFL* intron (Figure 4A). These *Cyclocarya GFAFL* alleles are also associated with

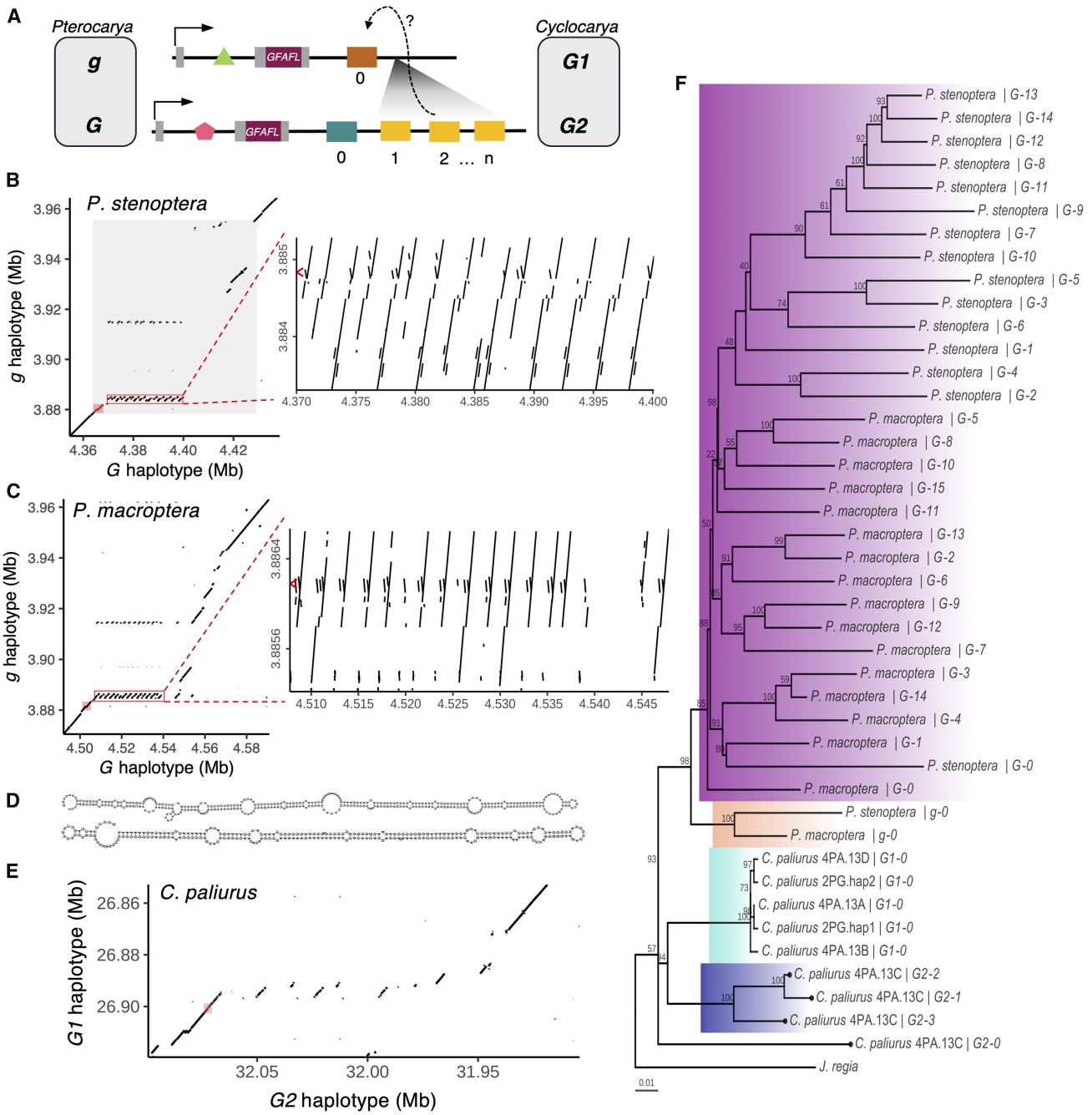


Figure 5. Parallel structural variation at *Pterocarya* and *Cyclocarya GFAFL* haplotypes

(A) Simplified schematic of the G-locus repeat structure in both genera. Solid arrows indicate the transcription start site of *GFAFL*. Green triangle and pink pentagon represent hypothetical intronic *cis*-regulatory elements that differ between haplotypes. Dotted arrow indicates a hypothesized *trans*-interaction between the alleles whereby the dominant allele regulates expression of the recessive allele.

(B and C) Alignment dot plots between the dominant and recessive G-locus haplotypes of two *Pterocarya* species. The gray region in (B) indicates the region associated with dichogamy type in *P. stenoptera*. Pink rectangles at bottom left in (B) and (C) show the position of *GFAFL* in each assembly. Insets on the right show a zoomed-in view of the tandem repeats. Red arrows indicate the location of homology in the recessive haplotypes to an inverted-repeat structure found in the dominant haplotypes of both species.

(D) Predicted secondary RNA-folding structures for two selected inverted repeats *P. stenoptera G-7* (top) and *P. macroptera G-0* (bottom).

(E) Alignment dot plot between dominant and recessive heterodichogamy haplotypes of *C. paliurus* across the region of segregating copy number variation downstream of *GFAFL*.

(legend continued on next page)

CellPress

structural variation downstream of the *GFAFL* coding region (Figure 5E). The presence of this ancient structural variant at the same locus that controls heterodichogamy in the sister genus *Pterocarya* suggested that these *Cyclocarya* alleles could be associated with the regulation of heterodichogamy and maintained by long-term balancing selection.

We found that these alleles also show an association with known dichogamy types. One allele (*G1*) appears in both assembled haplotypes of our diploid protogynous *C. paliurus* individual, as well as in a published assembly from another known diploid protogynous individual.[42] The *G2* sequence was not found in the raw assembly reads of either sample; thus, both protogynous individuals appear to be homozygous for the *G1* allele. The other allele (*G2*) appears at the syntenic location in a published assembly from a diploid protandrous individual (we also confirmed this individual possessed *G1*) and at the syntenic location in one of four published haplotype-resolved assemblies from a protandrous tetraploid (the other three haplotypes having *G1*). Thus, both protandrous assemblies derive from individuals heterozygous for the *G1* and *G2* alleles. In a separate flower RNA-seq dataset,[38] we detected expression of *G2* only in labeled protandrous samples and not in protogynous samples, whereas *G1* was expressed in both types (representing 9 individuals of each type). Although we lack the samples for a GWAS, the available data are consistent with an association between *GFAFL* alleles and dichogamy type, where protogynous individuals are *G1* homozygotes and protandrous individuals are heterozygous for *G1* and *G2*. Strikingly, the association of these alleles with dichogamy types shows a reversed-dominance pattern compared with that of *Pterocarya*, where here a dominant allele (*G2*) appears associated with protandry rather than protogyny.

We next examined segregation patterns of these alleles in published genomic resequencing data from a larger sample of both diploids and tetraploids[42] of unknown flowering type. Among all diploids examined (n = 13), genotypes were either *G1/G1* (n = 4) or *G1/G2* (n = 9), with zero *G2/G2* homozygotes seen despite the high frequency of the *G2* allele. This deficit of the *G2/G2* homozygote deviates from the expectation under random mating ($p = 0.056$, $\chi^2 = 3.64$, d.f. = 1). Although this result shows borderline statistical significance, note that the complete absence of *G2/G2* is the most extreme observation possible for a given frequency of *G2*. This is consistent with the prediction under disassortative mating between *G1* homozygotes and *G1/G2* heterozygotes. Among tetraploids (n = 36), individuals appear to possess either zero or at most one copy of *G2* for every three copies of *G1* (Figure 4B), consistent with the segregation of *G1* and *G2* alleles at a single locus during meiosis.

Cyclocarya GFAFL alleles show specialized expression in male and female flowers

To explore the possible functional role of *GFAFL* in heterodichogamy in *Cyclocarya*, we next examined expression of the two alleles in a published RNA-seq dataset representing a developmental time series of staminate and pistillate flowers of both protandrous and protogynous morphs.[35] The two alleles show striking patterns of unisexual flower-specific expression (Figure 4C). In protandrous heterozygotes, *G2* is more highly expressed during development of pistillate flowers, whereas *G1* is more highly expressed during development of staminate flowers. The expression of *G2* in pistillate flowers is consistent with neo-functionalization, as expression of the *GFAFL* orthologs in *Pterocarya*, *Juglans*, and Arabidopsis appears restricted to staminate flowers or the androecium (Figures S4C–S4H and S5E).

The expression of both alleles in *C. paliurus* heterozygotes is strongly correlated, suggesting they could have similar regulatory roles in flowering. Yet the transcripts show ~8% divergence in amino acid sequence ($^{dN}/_{dS} = 0.53$). Several substitutions occur in conserved regions, in addition to a change in the location of the start codon, indicating that they could encode proteins with partially divergent functions (Figure S4A). These patterns suggest a model wherein both alleles regulate the developmental rate of flowering structures, but one acts as a "fast" allele (*G1*), while the other acts as a "slow" allele (*G2*). For this model to be a sufficient basis for heterodichogamy, it would also predict upregulation of the fast allele in staminate flowers of protandrous morphs. Consistent with this, we see that the *G1* allele is highly upregulated in staminate flowers of protandrous morphs heterozygous for *G1* compared with protogynous morphs homozygous for *G1*, despite its lower relative copy number in heterozygotes.

Together, these expression patterns illustrate the potential for heterodichogamy to be controlled by regulatory variation at a single locus with multiple effects on the expression of a single gene. If two alleles differ quantitatively in the rate of promoting the transition to flowering, the combination of (1) sex-specific expression in heterozygotes and (2) morph-specific expression of the recessive allele may be sufficient to generate reciprocal matching of sexes between morphs.

In summary, several lines of evidence support the idea that ancient haplotypes at the *GFAFL* locus control heterodichogamy in *Cyclocarya*. (1) The *Cyclocarya GFAFL* locus shows ancient balanced polymorphism for two alleles. (2) A similar structural variant at the orthologous locus controls heterodichogamy in the closest living relative (*Pterocarya*). (3) Alleles at this locus show an association with flowering types (albeit with low sample size). (4) Genotype frequencies at this locus are biased in the direction predicted by disassortative mating. Finally, (5) *GFAFL* alleles show morph-specific expression patterns in flowers. We cannot formally exclude the alternative hypothesis that a different locus controls heterodichogamy in *Cyclocarya*. However, the above facts are best explained by the hypothesis that the *GFAFL* polymorphism is the causal determinant.

Convergent genomic structure of the dominant haplotype in *Pterocarya* and *Cyclocarya*

Patterns of allele-specific *GFAFL* expression in G-locus heterozygotes of both *Pterocarya* and *Cyclocarya* suggest the

(F) Maximum-likelihood phylogeny of repeat motifs in each assembly rooted with *Juglans* as an outgroup. *Pterocarya* G-locus haplotype sequences are highlighted in purple (*G*) and salmon (*g*). Also highlighted are *Cyclocarya G2* repeats 1–3 (dark blue) and *G1-0* sequences (light blue). All *Cyclocarya G2* sequences are indicated with points at tips.
Scale bar indicates substitutions per site. Node labels indicate bootstrap support.
See also Figure S6 and Data S1.

existence of both *cis*-regulatory divergence between G-locus haplotypes as well as *trans*-acting allele-specific regulation. We noted a striking parallel in haplotype structure between the dominant G-locus haplotypes in both genera that could reflect similar mechanisms of gene regulation.

In *Pterocarya*, pairwise alignments between the dominant and recessive haplotypes in both *P. stenoptera* and *P. macroptera* revealed the presence of an array of 14–15 ~2 kb tandem repeats in both dominant haplotypes (Figures 5B and 5C). These tandem repeats do not overlap any predicted genes but are homologous to sequence present ~1.5 kb downstream of *GFAFL* in both the dominant and recessive haplotypes of these *Pterocarya* species. Similarly, in *Cyclocarya*, we detected tandem repeats within the dominant haplotype showing homology to the same location downstream of the transcribed region of *GFAFL* (Figure 5E). The repeated motifs in *Pterocarya* and *Cyclocarya* show ~80% sequence homology to one another over ~1.5 kb. In *Cyclocarya*, the repeated sequence occurs in at least three copies in the putative dominant haplotype of a protandrous tetraploid (Figure 5E) and at least two copies in the collapsed assembly from a protandrous diploid (Figure S6B).

To understand the evolutionary relationships between these sequence motifs, we next constructed a phylogeny of the dominant haplotype repeats and the homologous sequences in the recessive haplotypes downstream of *GFAFL* (Figure 5F). All *Pterocarya* sequences form a well-supported clade (98% bootstrap support), with *G* haplotype sequences and *g* haplotype sequences forming two reciprocally monophyletic sister groups (85% and 100% bootstrap support). We found no signal of phylogenetic conservation of specific repeats within the tandem array across the two *Pterocarya* species, indicating considerable turnover in repeat sequence content within species. One or possibly two protogynous individuals of *P. stenoptera* appear to lack at least part of this repeat array (Figures S6D and S6E), potentially reflecting ongoing contraction and expansion. In *Cyclocarya*, we see a well-supported clade of recessive *G1* haplotype sequences and a well-supported clade of the repeats in the dominant *G2* haplotype (100% bootstrap support); both of these clades are reciprocally monophyletic with respect to *Pterocarya* sequences. However, we note that the *Cyclocarya G2* repeats (*G2-1*, *G2-2*, and *G2-3*) appear phylogenetically closer to the *Cyclocarya G1-0* sequence than to the syntenic sequence present on the same haplotype as the repeats themselves (*G2-0*). Despite turnover of these repeats within and across genera, the parallelism of this haplotype structure (and the lack of other notable shared haplotype features) highlights these repeats as potential candidates for regulating *GFAFL*.

Considering that RNA-seq data in both genera suggest a mechanism of *trans*-regulation of the recessive copy of *GFAFL* by the dominant haplotype, we considered the possibility that the repeats on the dominant haplotype encode a *trans*-acting RNA intermediary that either post-transcriptionally regulates the recessive copy of *GFAFL* or that transcriptionally regulates *GFAFL* by targeting a *cis*-regulatory element of *GFAFL* present on the recessive haplotype. Here, it is interesting to note a striking parallel to the molecular structure of the *Juglans* G-locus. In that case, the dominant haplotype contains tandem duplicates of the 3′ region of *TPPD1*, and small RNAs (sRNAs) are transcribed from these repeats in developing staminate flowers, corresponding with allele-specific regulation of *TPPD1* in the same tissue.[19]

Intriguingly, we noted the presence of a shared ~300 bp inverted-repeat structure present in multiple copies within the repeat region of both *P. stenoptera* and *P. macroptera* dominant haplotypes (Figures 5B and 5C). These inverted repeats are homologous to the same region in the respective recessive haplotype from each species that is not within the transcribed region of *GFAFL*. Furthermore, they are predicted to form thermodynamically stable hairpin structures (Figure 5D),[65] suggesting that they could act as precursors for the production of regulatory microRNAs (miRNAs). Although miRNAs often act as post-transcriptional regulators, they can also act as transcriptional regulators by directing methylation to *cis*-regulatory elements.[66] This sequence motif and its possible role in directing methylation are promising avenues for further research into the molecular mechanism underlying heterodichogamy in these genera. Notably, a concurrent study by Liu et al.[37] identifies the presence of sRNAs closely matching the sequence within the hairpin predicted here that could plausibly target the recessive haplotype to cause allele-specific silencing in *trans*.

Alternative mechanisms not involving an RNA intermediary may be possible.[67] For instance, an alternative hypothesis for the expression data in *Pterocarya* is that an enhancer region within the *G* haplotype outcompetes the *g* haplotype for recruitment of a limiting shared transcription factor, which transiently promotes *GFAFL* transcription during a specific developmental stage of flowering. Whatever the specific mechanism, these repeated homologous sequences found downstream of *GFAFL* within dominant haplotypes of both *Pterocarya* and *Cyclocarya* are a promising link for further investigation into the molecular-genetic mechanisms of gene regulation that underlie heterodichogamy.

DISCUSSION

In this work, we documented the occurrence of heterodichogamy in *Pterocarya*. Thus, all five genera within subfamily Juglandoideae have been confirmed to possess this mating system. Remarkably, our results now demonstrate that, among these, at least three distinct genetic loci control heterodichogamy.[19] In three genera (*Juglans*, *Pterocarya*, and *Carya*), a dominant allele controls female-first flowering (protogynous heterogamety), while only *Cyclocarya* appears to show protandrous heterogamety. The genetic basis of heterodichogamy in *Platycarya* has not yet been identified and may involve a separate mechanism. To our knowledge, there are no reports of the temporal flowering behavior of subfamily Engelhardioideae, but considering the similarities of inflorescence structure and pollination mode with Juglandoideae, it seems plausible that heterodichogamy may be found there also.

Neither the *Pterocarya* nor *Cyclocarya GFAFL* alleles, nor the other genetic systems we have previously mapped,[19] are estimated to be old enough to have originated in the common ancestor of this group. Thus, we cannot definitely rule out convergent evolutionary origins of heterodichogamy in these different genera. However, considering the occurrence of heterodichogamy in all of subfamily Juglandoideae, together with its apparent scarcity in related families, it is reasonable to suggest

that phenotypic heterodichogamy evolved in the common ancestor of Juglandoideae[24] and has been maintained for >70 Ma. If that is the case, the ancestral heterodichogamy mechanism was controlled by a mechanism that was susceptible to subsequent invasion by novel dichogamy-determining alleles at unlinked loci. Why might this occur? One possible adaptive explanation could be that heterodichogamy morphs controlled by an ancestral genetic system showed imperfect temporal matching between opposing sexual phases. Such temporal mismatches might arise due to environmental changes that differentially affected male and female flowering time. A novel dichogamy-determining allele conferring closer sex matching to the other morph under these conditions could then gain a fitness advantage and take over control of dichogamy type, analogous to a popular model of sex chromosome turnover.[68] Such alleles may be predisposed to arise at genes that already participate in regulatory networks underlying morph-specific flower development. Consistent with this idea, we see morph-specific temporal expression patterns among previously identified candidate genes in *Juglans* and *Carya* in the flower expression data of *Cyclocarya* (Figure S7).

In the wingnut genera *Pterocarya* and *Cyclocarya*, the same gene (*GFAFL*) appears to regulate heterodichogamy. Reciprocal monophyly of the *GFAFL* alleles in each genus and a lack of shared polymorphisms indicate that both sets of alleles diverged after the divergence of these genera (Figure 4). Although this could be consistent with convergent evolution of heterodichogamy via the same gene, it seems likely that polymorphism at this gene controlled heterodichogamy in the common ancestor of these sister genera. In this scenario, i.e., allelic turnover, one or both pairs of alleles replaced an ancestral pair of alleles at this locus, while phenotypic heterodichogamy was maintained (an analogous model has been offered for mimicry polymorphism in *Papilio* butterflies[69]). If the ancestor of these two genera was indeed heterodichogamous, as seems likely, the apparent male-first dominance of *Cylocarya* would require that dominance reversal must have occurred. Given that *Cyclocarya* may be the only case of protandrous heterogamety in Juglandoideae, this could represent a derived state. The expression of *GFAFL* in pistillate flowers of *Cyclocarya* also appears to be derived. Thus, if *GFAFL* controlled heterodichogamy in the ancestor of these genera, the molecular mechanism might have more closely resembled that of the *Pterocarya* system.

Dominance reversals are also known in various sex-determination systems.[70–72] This can even occur while retaining the same gene as the key sex regulator. For example, a transition from XY to ZW sex determination in poplar was mediated by different means of silencing the feminizing factor *ARR17* in males.[73] In contrast, the alleles in both the *Pterocarya* and putative *Cyclocarya* heterodichogamy systems are functional copies of the same protein-coding gene, both of which are expressed in flowers and likely differ quantitatively in their regulatory effects on flowering time rather than on sex allocation.

Although genetic control of heterodichogamy naturally parallels that of dioecy, it may differ in other key ways. For example, although many plant sex-determining regions lack heteromorphism,[74] perhaps as a result of purifying selection in the haploid phase,[75,76] the current data suggest that heteromorphism might be even less pronounced in heterodichogamy. Two factors could contribute to such a pattern. First, the apparent lack of phenotypic specialization of dichogamy morphs, other than in their timing, suggests that there may be limited genetic variation with morph-antagonist effects (e.g., protogynous beneficial/protandrous deleterious alleles or vice versa), precluding an adaptive advantage of recombination suppression that would subsequently trigger differential degeneration and heteromorphism.[19] In this regard, heterodichogamy may be more similar to heterostyly, where both morphs are hermaphroditic but differ in the relative heights of male and female floral organs. Interestingly, in all known cases, convergently evolved distyly loci systems involve hemizygosity for a small number of protein-coding genes in the dominant haplotypes of short-styled morphs, without chromosome-scale heteromorphism between haplotypes.[77–85] Second, although disassortative mating renders the dominant homozygote rare in heterodichogamy, some level of intra-morph mating evidently occurs, as viable dominant homozygotes are now known in several genera with mapped heterodichogamy systems (*Juglans*, *Carya*, and *Pterocarya*). Moreover, heterodichogamy morphs are not associated with self-incompatibility, unlike in heterostyly, where this is often the case.[86] This may enable occasional recombination between dominant haplotypes and a higher rate of purging of deleterious insertions or deletions compared with permanently heterozygous sex chromosomes.

As mentioned above, rather than wholesale silencing of one sex in dioecy, the molecular basis of heterodichogamy requires quantitative regulatory control of both male and female flowering in each morph in order to achieve reciprocal temporal sex matching. How might this be achieved at the genetic level? One possibility (pleiotropy hypothesis) is that pleiotropy of a single regulatory change could simultaneously delay flowering of one sex while advancing the other through a resource trade-off.[19] However, if such a trade-off were to exist, then the developmental rates of male and female flowering structures would likely be negatively correlated within dichogamy types. Our data on the relative developmental rates of male and female inflorescences in *Pterocarya* show the opposite pattern, where male and female developmental rates within each dichogamy type are positively correlated (Figure 1C). Although the sex-specific phenology variation we see within types is only a subset of the total variation, it is difficult to imagine how the direction of pleiotropy would reverse over some threshold shift in flowering time of either sex. A second possibility is that variants across multiple genes with sex-specific effects on either male or female flowering time become coupled together on the same haplotype, enabling co-inheritance of variation that delays flowering of one sex while advancing that of the other. The clear evidence of recombination suppression across multiple genes putatively involved in flowering at the *Carya* G-locus haplotypes lends support to this model.[19]

Our findings in the wingnut group suggest a third possibility. If alleles at a single gene regulating the developmental transition to flowering vary quantitatively in their effect, heterodichogamy can result from a combination of sex-specific *cis*-regulatory divergence and *trans*-regulation of one allele by the other. The *Cyclocarya* data are consistent with this model; the *G2* allele found in protandrous heterozygote individuals shows higher relative expression in *cis* in pistillate (female) flowers, while the *G2*

haplotype upregulates the *G1* allele in *trans* in staminate (male) flowers, such that *G1* shows higher relative expression. These patterns are consistent with the *G1* copy of GFAFL acting as a fast protein allele, while the *G2* copy acts as a slow allele, corresponding to early vs. late flowering. Such a model would allow the *G2* allele found only in the protandrous heterozygote to slow the development of pistillate flowers, while also driving higher expression of the *G1* allele in *trans* to accelerate the development of staminate flowers relative to the protogynous *G1* homozygotes, thereby reversing the order of sex-specific flowering to generate intermorph synchronization. If *trans*-regulation of the *G1* allele by the *G2* haplotype is key to the acceleration of staminate flowering in the protandrous morph of *Cyclocarya*, this predicts that a *G2*/*G2* individual would show delayed staminate flowering relative to a protandrous *G1*/*G2* individual. In this way, the deficit of *G2*/*G2* homozygotes per se could be a cause as well as a consequence of disassortative mating in this system.

The data from staminate flowers *Pterocarya* are also consistent with this model, but the data for pistillate flowers are not. The *G* haplotype found in protogynous heterozygotes seems to cause allele-specific silencing in *trans* of the *g* copy of *GFAFL* in early staminate flower development, while the *G* allele of *GFAFL* is itself expressed. In this case, the *G* allele may thus be acting as a slow allele compared with a fast counterpart, *g*. However, this model does not seem to explain the pistillate flowering offset, as we did not detect *GFAFL* expression in pistillate flower tissue of *Pterocarya*—neither at the time of bud break nor at the time of flowering. Although we cannot exclude the possibility of expression in pistillate flowers during a narrow developmental window, the data suggest that *GFAFL* regulation is not likely the causal basis for the female timing difference at the G-locus in *Pterocarya*. One or both of the G-locus haplotypes may be controlling female flowering through a *trans*-acting mechanism (for example, small RNAs) differentially regulating flowering-time genes elsewhere in the genome. Results from a concurrent study by Liu et al.[37] suggest that small RNAs transcribed from the dominant haplotype of *Pterocarya* could plausibly regulate the expression of additional flowering-time genes that show morph-specific expression patterns. We additionally note that the current data for the *Juglans* locus could also be consistent with this model[19]; however, further study is needed.

It is natural to suggest that the *trans*-regulatory effects of the dominant haplotype on the expression of the recessive copy of *GFAFL* in both *Pterocarya* and *Cyclocarya* is directly related to the molecular mechanism of dominance. Furthermore, the convergent structure of the dominant haplotypes in both genera, together with our identification of tandem duplicates of a hairpin structure in the dominant haplotype of *Pterocarya* matching sequence on the recessive haplotype, hints that sRNAs could potentially mediate dominance in these systems. Together with data from *Juglans*, where tandem repeats of the 3′ region of a flowering-time candidate gene (*TPPD1*) occur in the dominant haplotype and encode putative regulatory sRNAs, these findings suggest that this type of regulatory architecture could be a more general feature of heterodichogamy; however, data from more systems are needed. If the tandem duplicate structure in the dominant G-locus haplotypes of wingnuts is responsible for *trans*-regulation of *GFAFL*, it is possible that dominance was established gradually, and perhaps secondarily, after the original appearance of balanced polymorphism at the locus. Indeed, balanced polymorphisms are classic models where dominance modifiers are suggested to evolve secondarily[87–94] Direct evidence for the evolution of dominance modifiers per se at mating type loci comes from the self-incompatibility systems of Brassicaceae, where *trans*-acting sRNAs encoded by dominant haplotypes mediate transcriptional silencing of recessive haplotype genes at the same locus.[95,96]

Our findings show that heterodichogamy can be achieved through diverse genetic pathways, which may facilitate its evolutionary turnover. In wingnuts, and perhaps walnuts, it appears that changes affecting a single protein-coding gene may be sufficient to control reciprocal temporal matching of opposing sexual phases. Given this, why then have the current-day systems been conserved over vast time periods? Perhaps the molecular architectures of these systems—that is, the putative regulatory repeats—enable sufficient flexibility for adaptation of finely tuned reciprocal matching, making these systems robust to invasion.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to, and will be fulfilled by, the lead contact, Jeffrey S. Groh (jgroh@berkeley.edu).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- Genome assemblies for *P. stenoptera*, *P. fraxinifolia*, *P. macroptera*, and *C. paliurus* have been deposited in NCBI GenBank under accession numbers PRJNA1234227–PRJNA1234228, PRJNA1238886–PRJNA1238887, PRJNA1238908–PRJNA1238909, and PRJNA1239020–PRJNA1239021. Raw assembly reads and resequencing data have been deposited under accession number PRJNA1109986, and RNA-seq data have been deposited under accession number PRJNA1123722. The above sequence data are publicly available as of the date of publication.
- Scripts used to perform the analyses in this paper are publicly available at https://doi.org/10.5281/zenodo.17118673.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We are grateful to Sonoma Botanical Garden, USDA Wolfskill Experimental Orchard, and Hunter Wilson of UC Davis for providing access to living collections sampled in this study. We thank Pat Brown, Blake Meyers, Chuck Langley, Steven Manchester, Stephen Wright, and members of the Coop lab for helpful comments and discussion. We thank Michael Turelli and three anonymous reviewers for feedback on an earlier version of the manuscript. Funding was provided by the National Institutes of Health (NIH) (R35 GM136290 to G.C.) and by a research award from the Center for Population Biology, UC Davis (to J.S.G.).

AUTHOR CONTRIBUTIONS

J.S.G. and G.C. conceived and designed the research. J.S.G. collected phenological data and tissue samples; performed data curation, bioinformatics, formal analysis, and visualization; and wrote the original draft of the manuscript. G.A. isolated nucleic acids from tissue samples. M.W. facilitated tissue collections and contributed phenological observations. G.C. assisted with tissue collections, contributed resources and expertise, and edited the manuscript. All authors reviewed the manuscript and approved the final submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Phenotyping
  - Genome assembly and annotation
  - Species tree inference
  - Short read sequence alignment, variant identification and filtering
  - Investigation of candidate genes
  - Genetic structure and GWAS
  - Analysis of polymorphism and divergence within and between haplotypes
  - Haplotype age estimation
  - Relatedness and investigation of function of *GFAFL* homologs
  - Gene expression analysis
  - Analysis of the *GFAFL* region in *Cyclocarya*
  - Analysis of haplotype structure
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2025.09.061.

Received: March 21, 2025
Revised: June 25, 2025
Accepted: September 23, 2025
Published: October 16, 2025

REFERENCES

1. Hall, M.C., and Willis, J.H. (2006). Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. Evolution *60*, 2466–2477. https://doi.org/10.1111/j.0014-3820.2006.tb01882.x.
2. Franks, S.J., Sim, S., and Weis, A.E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proc. Natl. Acad. Sci. USA *104*, 1278–1282. https://doi.org/10.1073/pnas.0608379104.
3. Gaudinier, A., and Blackman, B.K. (2020). Evolutionary processes from the perspective of flowering time diversity. New Phytol. *225*, 1883–1898. https://doi.org/10.1111/nph.16205.
4. Warner, R.R. (1975). The adaptive significance of sequential hermaphroditism in animals. Am. Nat. *109*, 61–82. https://doi.org/10.1086/282974.
5. Munday, P.L., Buston, P.M., and Warner, R.R. (2006). Diversity and flexibility of sex-change strategies in animals. Trends Ecol. Evol. *21*, 89–95. https://doi.org/10.1016/j.tree.2005.10.020.
6. Lloyd, D.G., and Webb, C.J. (1986). The avoidance of interference between the presentation of pollen and stigmas in angiosperms I. Dichogamy. N. Z. J. Bot. *24*, 135–162. https://doi.org/10.1080/0028825X.1986.10409725.
7. Routley, M.B., Bertin, R.I., and Husband, B.C. (2004). Correlated evolution of dichogamy and self-incompatibility: a phylogenetic perspective. Int. J. Plant Sci. *165*, 983–993. https://doi.org/10.1086/423881.
8. Leigh, E.G.J., Charnov, E.L., and Warner, R.R. (1976). Sex ratio, sex change, and natural selection. Proc. Natl. Acad. Sci. USA *73*, 3656–3660. https://doi.org/10.1073/pnas.73.10.3656.
9. Brunet, J., and Charlesworth, D. (1995). Floral sex allocation in sequentially blooming plants. Evolution *49*, 70–79. https://doi.org/10.1111/j.1558-5646.1995.tb05959.x.
10. Renner, S.S. (2001). How common is heterodichogamy? Trends Ecol. Evol. *16*, 595–597. https://doi.org/10.1016/S0169-5347(01)02280-7.
11. Endress, P.K. (2020). Structural and temporal modes of heterodichogamy and similar patterns across angiosperms. Bot. J. Linn. Soc. *193*, 5–18. https://doi.org/10.1093/botlinnean/boaa001.
12. Bai, W.N., Zeng, Y.F., and Zhang, D.Y. (2007). Mating patterns and pollen dispersal in a heterodichogamous tree, *Juglans mandshurica* (Juglandaceae). New Phytol. *176*, 699–707. https://doi.org/10.1111/j.1469-8137.2007.02202.x.
13. Gleeson, S.K. (1982). Heterodichogamy in walnuts: inheritance and stable ratios. Evolution *36*, 892–902. https://doi.org/10.1111/j.1558-5646.1982.tb05461.x.
14. Fisher, R.A. (1930). The Genetical Theory of Natural Selection (Clarendon Press). https://doi.org/10.5962/bhl.title.27468.
15. Ganders, F.R. (1979). The biology of heterostyly. N. Z. J. Bot. *17*, 607–635. https://doi.org/10.1080/0028825X.1979.10432574.
16. Dommée, B., Bompar, J.L., and Denelle, N. (1990). Sexual tetramorphism in *Thymelaea hirsuta* (Thymelaeaceae): evidence of the pathway from heterodichogamy to dioecy at the infraspecific level. Am. J. Bot. *77*, 1449–1462. https://doi.org/10.1002/j.1537-2197.1990.tb12555.x.
17. Pendleton, R.L., Freeman, D.C., McArthur, E.D., and Sanderson, S.C. (2000). Gender specialization in heterodichogamous *Grayia brandegei* (Chenopodiaceae): evidence for an alternative pathway to dioecy. Am. J. Bot. *87*, 508–516. https://doi.org/10.2307/2656594.
18. Pannell, J.R., and Verdú, M. (2006). The evolution of gender specialization from dimorphic hermaphroditism: paths from heterodichogamy to gynodioecy and androdioecy. Evolution *60*, 660–673. https://doi.org/10.1554/05-481.1.
19. Groh, J.S., Vik, D.C., Davis, M., Monroe, J.G., Stevens, K.A., Brown, P.J., Langley, C.H., and Coop, G. (2025). Ancient structural variants control sex-specific flowering time morphs in walnuts and hickories. Science *387*, eado5578. https://doi.org/10.1126/science.ado5578.
20. Zhao, J.-L., Dong, Y., Huang, A.-D., Duan, S.-C., Peng, X.-C., Liao, H., Chen, J.-H., Luo, Y.-L., Lan, Q.-Y., Wang, Y.-L., et al. (2025). Ginger genome reveals the *SMPED1* gene causing sex-phase synchrony and outcrossing in a flowering plant. Nat. Plants. Published online October 7, 2025. https://doi.org/10.1038/s41477-025-02125-3.
21. Chen, S., Peng, X., Xie, Z., Zhang, M., Huang, A., Wang, W., Xiang, G., Zhang, K., Gao, R., Duan, B., et al. (2025). Genetic and genomic insights into dichogamy in Zingiberaceae. Plant Commun. *6*, 101352. https://doi.org/10.1016/j.xplc.2025.101352.
22. Delpino, F. (1874). Ulteriori osservazioni e considerazioni sulla dicogamia nel regno vegetale. Atti della Societa Italiana Scientia Naturale *17*, 402–407.
23. Darwin, C. (1877). The Different Forms of Flowers on Plants of the Same Species (John Murray).
24. Fukuhara, T., and Tokumaru, S.i. (2014). Inflorescence dimorphism, heterodichogamy and thrips pollination in *Platycarya strobilacea* (Juglandaceae). Ann. Bot. *113*, 467–476. https://doi.org/10.1093/aob/mct278.
25. Mao, X., Liu, J., Li, X., Qin, J., and Fu, X. (2016). Flowering biological characteristics and mating system in immature plantations of heterodichogamous *Cyclocarya paliurus*. J. Nanjing For. Univ. *40*, 47–55.
26. Thompson, T.E., and Romberg, L.D. (1985). Inheritance of heterodichogamy in pecan. J. Hered. *76*, 456–458. https://doi.org/10.1093/oxfordjournals.jhered.a110144.
27. Wahl, V., Ponnu, J., Schlereth, A., Arrivault, S., Langenecker, T., Franke, A., Feil, R., Lunn, J.E., Stitt, M., and Schmid, M. (2013). Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. Science *339*, 704–707. https://doi.org/10.1126/science.1230406.

28. Pigg, K.B., DeVore, M.L., and Taylor, W. (2023). New features of *Cyclocarya brownii* Manchester & Dilcher from the Late Paleocene of North Dakota, USA. Int. J. Plant Sci. *184*, 282–303. https://doi.org/10.1086/724496.

29. Lyson, T.R., Miller, I.M., Bercovici, A.D., Weissenburger, K., Fuentes, A. J., Clyde, W.C., Hagadorn, J.W., Butrim, M.J., Johnson, K.R., Fleming, R. F., et al. (2019). Exceptional continental record of biotic recovery after the Cretaceous–Paleogene mass extinction. Science *366*, 977–983. https://doi.org/10.1126/science.aay2268.

30. Manchester, S.R. (1991). *Cruciptera*, a new juglandaceous winged fruit from the Eocene and Oligocene of western North America. Syst. Bot. *16*, 715–725. https://doi.org/10.2307/2418873.

31. Yan, H., Zhou, P., Wang, W., Ye, J.F., Tan, S.L., Guo, C.C., Zhang, W.G., Zhu, Z.W., Liu, Y.Z., and Xiang, X.G. (2024). Biogeographic history of *Pterocarya* (Juglandaceae) inferred from phylogenomic and fossil data. J. Syst. Evol. *62*, 1165–1176. https://doi.org/10.1111/jse.13055.

32. Manchester, S.R., and Dilcher, D.L. (1982). Pterocaryoid fruits (Juglandaceae) in the Paleogene of North America and their evolutionary and biogeographic significance. Am. J. Bot. *69*, 275–286. https://doi.org/10.1002/j.1537-2197.1982.tb13258.x.

33. Song, Y.G., Walas, Ł., Pietras, M., Sâm, H.V., Yousefzadeh, H., Ok, T., Farzaliyev, V., Worobiec, G., Worobiec, E., Stachowicz-Rybka, R., et al. (2021). Past, present and future suitable areas for the relict tree *Pterocarya fraxinifolia* (Juglandaceae): Integrating fossil records, niche modeling, and phylogeography for conservation. Eur. J. Forest Res. *140*, 1323–1339. https://doi.org/10.1007/s10342-021-01397-6.

34. Stults, D.Z., Tiffney, B.H., and Axsmith, B.J. (2022). New observations on the last *Pterocarya* (Juglandaceae) occurrences in eastern North America. Int. J. Plant Sci. *183*, 380–392. https://doi.org/10.1086/720182.

35. Qu, Y., Shang, X., Zeng, Z., Yu, Y., Bian, G., Wang, W., Liu, L., Tian, L., Zhang, S., Wang, Q., et al. (2023). Whole-genome duplication reshaped adaptive evolution in a relict plant species, *Cyclocarya paliurus*. Genomics Proteomics Bioinformatics *21*, 455–469. https://doi.org/10.1016/j.gpb.2023.02.001.

36. Yu, R.M., Zhang, N., Zhang, B.W., Liang, Y., Pang, X.X., Cao, L., Chen, Y. D., Zhang, W.P., Yang, Y., Zhang, D.Y., et al. (2023). Genomic insights into biased allele loss and increased gene numbers after genome duplication in autotetraploid *Cyclocarya paliurus*. BMC Biol. *21*, 168. https://doi.org/10.1186/s12915-023-01668-1.

37. Liu, H.-S., Wang, W.-H., Song, Y.-F., Yang, Y., Zhang, D.-Y., Bai, W.-N., and Zhang, B.-W. (2025). A pan-genome framework reveals structural variation and small RNA regulation underlying heterodichogamy in *Pterocarya*.. Published online October 16, 2025. Curr. Biol. https://doi.org/10.1016/j.cub.2025.09.017

38. Chen, X., Mao, X., Huang, P., and Fang, S. (2019). Morphological characterization of flower buds development and related gene expression profiling at bud break stage in heterodichogamous *Cyclocarya paliurus* (Batal.) Iljinskaja. Genes *10*, 818. https://doi.org/10.3390/genes10100818.

39. Luza, J.G., and Polito, V.S. (1988). Microsporogenesis and anther differentiation in *Juglans regia* L.: A developmental basis for heterodichogamy in walnut. Botanical Gazette *149*, 30–36. https://doi.org/10.1086/337688.

40. Rhein, H.S., Sreedasyam, A., Cooke, P., Velasco-Cruz, C., Grimwood, J., Schmutz, J., Jenkins, J., Kumar, S., Song, M., Heerema, R.J., et al. (2023). Comparative transcriptome analyses reveal insights into catkin bloom patterns in pecan protogynous and protandrous cultivars. PLoS One *18*, e0281805. https://doi.org/10.1371/journal.pone.0281805.

41. Zhang, Z.Y., Xia, H.X., Yuan, M.J., Gao, F., Bao, W.H., Jin, L., Li, M., and Li, Y. (2024). Multi-omics analyses provide insights into the evolutionary history and the synthesis of medicinal components of the Chinese wingnut. Plant Divers. *46*, 309–320. https://doi.org/10.1016/j.pld.2024.03.010.

42. Qu, Y., Shang, X., Fang, S., Zhang, X., and Fu, X. (2023). Genome assembly of two diploid and one auto-tetraploid *Cyclocarya paliurus* genomes. Sci. Data *10*, 507. https://doi.org/10.1038/s41597-023-02402-w.

43. Sutton, J. (2019). '*Cyclocarya paliurus*'. Trees and Shrubs Online. https://www.treesandshrubsonline.org/articles/cyclocarya/cyclocarya-paliurus/.

44. Oginuma, K. (1999). Karyomorphology and evolution in Juglandales: a Review. Acta Phytotax Geobot. *50*, 229–241.

45. Song, Y.G., Li, Y., Meng, H.H., Fragnière, Y., Ge, B.J., Sakio, H., Yousefzadeh, H., Bétrisey, S., and Kozlowski, G. (2020). Phylogeny, taxonomy, and biogeography of *Pterocarya* (Juglandaceae). Plants (Basel) *9*, 1524. https://doi.org/10.3390/plants9111524.

46. Zhang, J.B., Li, R.Q., Xiang, X.G., Manchester, S.R., Lin, L., Wang, W., Wen, J., and Chen, Z.D. (2013). Integrated fossil and molecular data reveal the biogeographic diversification of the eastern Asian-eastern North American disjunct hickory genus (*Carya* Nutt.). PLoS One *8*, e70449. https://doi.org/10.1371/journal.pone.0070449.

47. Manchester, S.R. (1987). The fossil history of the Juglandaceae. Monogr. Syst. Bot. Mo. Bot. Gard. *21*, 1–137.

48. APG IV (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. *181*, 1–20. https://doi.org/10.1111/boj.12385.

49. dos Reis, M.d., and Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. Mol. Biol. Evol. *28*, 2161–2172. https://doi.org/10.1093/molbev/msr045.

50. Mu, X.Y., Tong, L., Sun, M., Zhu, Y.X., Wen, J., Lin, Q.W., and Liu, B. (2020). Phylogeny and divergence time estimation of the walnut family (Juglandaceae) based on nuclear RAD-seq and chloroplast genome data. Mol. Phylogenet. Evol. *147*, 106802. https://doi.org/10.1016/j.ympev.2020.106802.

51. Zhou, H., Hu, Y., Ebrahimi, A., Liu, P., Woeste, K., Zhao, P., and Zhang, S. (2021). Whole genome based insights into the phylogeny and evolution of the Juglandaceae. BMC Ecol. Evol. *21*, 191. https://doi.org/10.1186/s12862-021-01917-3.

52. Ding, Y.M., Pang, X.X., Cao, Y., Zhang, W.P., Renner, S.S., Zhang, D.Y., and Bai, W.N. (2023). Genome structure-based Juglandaceae phylogenies contradict alignment-based phylogenies and substitution rates vary with DNA repair genes. Nat. Commun. *14*, 617. https://doi.org/10.1038/s41467-023-36247-z.

53. Manning, W.E. (1978). The classification within the Juglandaceae. Ann. Mo. Bot. Gard. *65*, 1058–1087. https://doi.org/10.2307/2398782.

54. Geng, F., Zhang, X., Ma, J., Liu, H., Ye, H., Hao, F., Liu, M., Dang, M., Zhou, H., Li, M., et al. (2024). Genome assembly and winged fruit gene regulation of Chinese wingnut: insights from genomic and transcriptomic analyses. Genomics Proteomics Bioinformatics *22*, qzae087. https://doi.org/10.1093/gpbjnl/qzae087.

55. Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. *14*, 113–124. https://doi.org/10.1038/nrg3366.

56. Stolle, E., Pracana, R., Howard, P., Paris, C.I., Brown, S.J., Castillo-Carrillo, C., Rossiter, S.J., and Wurm, Y. (2019). Degenerative expansion of a young supergene. Mol. Biol. Evol. *36*, 553–561. https://doi.org/10.1093/molbev/msy236.

57. Akagi, T., Fujita, N., Shirasawa, K., Tanaka, H., Nagaki, K., Masuda, K., Horiuchi, A., Kuwada, E., Kawai, K., Kunou, R., et al. (2025). Rapid and dynamic evolution of a giant Y chromosome in *Silene latifolia*. Science *387*, 637–643. https://doi.org/10.1126/science.adk9074.

58. Moraga, C., Branco, C., Rougemont, Q., Jedlička, P., Mendoza-Galindo, E., Veltsos, P., Hanique, M., Rodríguez de la Vega, R.C., Tannier, E., Liu, X., et al. (2025). The *Silene latifolia* genome and its giant Y chromosome. Science *387*, 630–636. https://doi.org/10.1126/science.adj7430.

59. Wahl, V., Brand, L.H., Guo, Y.L., and Schmid, M. (2010). The FANTASTIC FOUR proteins influence shoot meristem size in *Arabidopsis thaliana*. BMC Plant Biol. *10*, 285. https://doi.org/10.1186/1471-2229-10-285.

60. Wang, K., He, J., Zhao, Y., Wu, T., Zhou, X., Ding, Y., Kong, L., Wang, X., Wang, Y., Li, J., et al. (2018). EAR1 negatively regulates ABA signaling by enhancing 2C protein phosphatase activity. Plant Cell *30*, 815–834. https://doi.org/10.1105/tpc.17.00875.

61. Lim, C.W., Bae, Y., and Lee, S.C. (2022). Differential role of *Capsicum annuum* FANTASTIC FOUR-like gene CaFAF1 on drought and salt stress responses. Environ. Exp. Bot. *199*, 104887. https://doi.org/10.1016/j.envexpbot.2022.104887.

62. Zhang, X., Mogel, K.J.H.V., Lor, V.S., Hirsch, C.N., De Vries, B., Kaeppler, H.F., Tracy, W.F., and Kaeppler, S.M. (2019). Maize *sugary enhancer1* (*se1*) is a gene affecting endosperm starch metabolism. Proc. Natl. Acad. Sci. USA *116*, 20776–20785. https://doi.org/10.1073/pnas.1902747116.

63. Zhang, D., Ai, G., Ji, K., Huang, R., Chen, C., Yang, Z., Wang, J., Cui, L., Li, G., Tahira, M., et al. (2024). EARLY FLOWERING is a dominant gain-of-function allele of FANTASTIC FOUR 1/2c that promotes early flowering in tomato. Plant Biotechnol. J. *22*, 698–711. https://doi.org/10.1111/pbi.14217.

64. Zhang, S.Y., Peng, Y.Q., Xiang, G.S., Song, W.L., Feng, L., Jiang, X.Y., Li, X.J., He, S.M., Yang, S.C., Zhao, Y., et al. (2024). Functional characterization of genes related to triterpene and flavonoid biosynthesis in *Cyclocarya paliurus*. Planta *259*, 50. https://doi.org/10.1007/s00425-023-04282-1.

65. Axtell, M.J., and Meyers, B.C. (2018). Revisiting criteria for plant microRNA annotation in the era of big data. Plant Cell *30*, 272–284. https://doi.org/10.1105/tpc.17.00851.

66. Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., and Qi, Y. (2010). DNA methylation mediated by a microRNA pathway. Mol. Cell *38*, 465–475. https://doi.org/10.1016/j.molcel.2010.03.008.

67. Matzke, M., Mette, M.F., Jakowitsch, J., Kanno, T., Moscone, E.A., Van Der Winden, J., and Matzke, A.J. (2001). A test for transvection in plants: DNA pairing may lead to trans-activation or silencing of complex heteroalleles in tobacco. Genetics *158*, 451–461. https://doi.org/10.1093/genetics/158.1.451.

68. Van Doorn, G.S., and Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. Nature *449*, 909–912. https://doi.org/10.1038/nature06178.

69. Palmer, D.H., and Kronforst, M.R. (2020). A shared genetic basis of mimicry across swallowtail butterflies points to ancestral co-option of *doublesex*. Nat. Commun. *11*, 6. https://doi.org/10.1038/s41467-019-13859-y.

70. Augstenová, B., Johnson Pokorná, M., Altmanová, M., Frynta, D., Rovatsos, M., and Kratochvíl, L. (2018). ZW, XY, and yet ZW: Sex chromosome evolution in snakes even more complicated. Evolution *72*, 1701–1707. https://doi.org/10.1111/evo.13543.

71. Nielsen, S.V., Banks, J.L., Diaz, R.E., Jr., Trainor, P.A., and Gamble, T. (2018). Dynamic sex chromosomes in old world chameleons (Squamata: Chamaeleonidae). J. Evol. Biol. *31*, 484–490. https://doi.org/10.1111/jeb.13242.

72. Hu, N., Sanderson, B.J., Guo, M., Feng, G., Gambhir, D., Hale, H., Wang, D., Hyden, B., Liu, J., Smart, L.B., et al. (2023). Evolution of a ZW sex chromosome system in willows. Nat. Commun. *14*, 7144. https://doi.org/10.1038/s41467-023-42880-5.

73. Müller, N.A., Kersten, B., Leite Montalvão, A.P., Mähler, N., Bernhardsson, C., Bräutigam, K., Carracedo Lorenzo, Z., Hoenicka, H., Kumar, V., Mader, M., et al. (2020). A single gene underlies the dynamic evolution of poplar sex determination. Nat. Plants *6*, 630–637. https://doi.org/10.1038/s41477-020-0672-9.

74. Renner, S.S., and Müller, N.A. (2021). Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nat. Plants *7*, 392–402. https://doi.org/10.1038/s41477-021-00884-3.

75. Chibalina, M.V., and Filatov, D.A. (2011). Plant Y chromosome degeneration is retarded by haploid purifying selection. Curr. Biol. *21*, 1475–1479. https://doi.org/10.1016/j.cub.2011.07.045.

76. Sandler, G., Beaudry, F.E.G., Barrett, S.C.H., and Wright, S.I. (2018). The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants. Evol. Lett. *2*, 368–377. https://doi.org/10.1002/evl3.60.

77. Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., Oosterhout, C.V., and Gilmartin, P.M. (2016). Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. Nat. Plants *2*, 16188. https://doi.org/10.1038/nplants.2016.188.

78. Shore, J.S., Hamam, H.J., Chafe, P.D.J., Labonne, J.D.J., Henning, P.M., and McCubbin, A.G. (2019). The long and short of the S-locus in *Turnera* (Passifloraceae). New Phytol. *224*, 1316–1329. https://doi.org/10.1111/nph.15970.

79. Gutiérrez-Valencia, J., Hughes, P.W., Berdan, E.L., and Slotte, T. (2021). The genomic architecture and evolutionary fates of supergenes. Genome Biol. Evol. *13*, evab057. https://doi.org/10.1093/gbe/evab057.

80. Fawcett, J.A., Takeshima, R., Kikuchi, S., Yazaki, E., Katsube-Tanaka, T., Dong, Y., Li, M., Hunt, H.V., Jones, M.K., Lister, D.L., et al. (2023). Genome sequencing reveals the genetic architecture of heterostyly and domestication history of common buckwheat. Nat. Plants *9*, 1236–1251. https://doi.org/10.1038/s41477-023-01474-1.

81. Yang, J., Xue, H., Li, Z., Zhang, Y., Shi, T., He, X., Barrett, S.C.H., Wang, Q., and Chen, J. (2023). Haplotype-resolved genome assembly provides insights into the evolution of S-locus supergene in distylous *Nymphoides indica*. New Phytol. *240*, 2058–2071. https://doi.org/10.1111/nph.19264.

82. Zhao, Z., Zhang, Y., Shi, M., Liu, Z., Xu, Y., Luo, Z., Yuan, S., Tu, T., Sun, Z., Zhang, D., et al. (2023). Genomic evidence supports the genetic convergence of a supergene controlling the distylous floral syndrome. New Phytol. *237*, 601–614. https://doi.org/10.1111/nph.18540.

83. Raimondeau, P., Ksouda, S., Marande, W., Fuchs, A.L., Gryta, H., Theron, A., Puyoou, A., Dupin, J., Cheptou, P.O., Vautrin, S., et al. (2024). A hemizygous supergene controls homomorphic and heteromorphic self-incompatibility systems in Oleaceae. Curr. Biol. *34*, 1977–1986.e8. https://doi.org/10.1016/j.cub.2024.03.029.

84. Castric, V., Batista, R.A., Carré, A., Mousavi, S., Mazoyer, C., Godé, C., Gallina, S., Ponitzki, C., Theron, A., Bellec, A., et al. (2024). The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene. Curr. Biol. *34*, 1967–1976.e6. https://doi.org/10.1016/j.cub.2024.03.047.

85. Yuan, S., Barrett, S.C.H., Tang, C., Zhang, Y., Sun, Q., Zhao, Z., Zhang, Y., Zhang, D., and Luo, S. (2025). Genomic evidence unveils the genetic architecture and evolution of the S-locus controlling heterostyly in Rubiaceae. New Phytol. *247*, 1925–1941. https://doi.org/10.1111/nph.70311.

86. Barrett, S.C.H. (2019). 'A most complex marriage arrangement': recent advances on heterostyly and unresolved questions. New Phytol. *224*, 1051–1067. https://doi.org/10.1111/nph.16026.

87. Clarke, C.A., and Sheppard, P.M. (1960). The evolution of dominance under disruptive selection. Heredity *14*, 73–87. https://doi.org/10.1038/hdy.1960.6.

88. Charlesworth, D., and Charlesworth, B. (1975). Theoretical genetics of Batesian mimicry III. Evolution of dominance. J. Theor. Biol. *55*, 325–337. https://doi.org/10.1016/s0022-5193(75)80083-x.

89. Otto, S.P., and Bourguet, D. (1999). Balanced polymorphisms and the evolution of dominance. Am. Nat. *153*, 561–574. https://doi.org/10.1086/303204.

90. Billiard, S., and Castric, V. (2011). Evidence for fisher's dominance theory: how many 'special cases'? Trends Genet. *27*, 441–445. https://doi.org/10.1016/j.tig.2011.06.005.

91. Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V., and Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. Nat. Commun. *5*, 5644. https://doi.org/10.1038/ncomms6644.

92. Barson, N.J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G.H., Fiske, P., Jacq, C., Jensen, A.J., Johnston, S.E., Karlsson, S., et al. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature *528*, 405–408. https://doi.org/10.1038/nature16062.

93. Pearse, D.E., Barson, N.J., Nome, T., Gao, G., Campbell, M.A., Abadía-Cardoso, A., Anderson, E.C., Rundio, D.E., Williams, T.H., Naish, K.A., et al. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. Nat. Ecol. Evol. *3*, 1731–1742. https://doi.org/10.1038/s41559-019-1044-6.

94. Lesaffre, T., Pannell, J.R., and Mullon, C. (2024). An explanation for the prevalence of XY over ZW sex determination in species derived from hermaphroditism. Proc. Natl. Acad. Sci. USA *121*, e2406305121. https://doi.org/10.1073/pnas.2406305121.

95. Tarutani, Y., Shiba, H., Iwano, M., Kakizaki, T., Suzuki, G., Watanabe, M., Isogai, A., and Takayama, S. (2010). Trans-acting small RNA determines dominance relationships in *Brassica* self-incompatibility. Nature *466*, 983–986. https://doi.org/10.1038/nature09308.

96. Durand, E., Méheust, R., Soucaze, M., Goubet, P.M., Gallina, S., Poux, C., Fobis-Loisy, I., Guillon, E., Gaude, T., Sarazin, A., et al. (2014). Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. Science *346*, 1200–1205. https://doi.org/10.1126/science.1259442.

97. Cheng, H., Jarvis, E.D., Fedrigo, O., Koepfli, K.P., Urban, L., Gemmell, N. J., and Li, H. (2022). Haplotype-resolved assembly of diploid genomes without parental data. Nat. Biotechnol. *40*, 1332–1335. https://doi.org/10.1038/s41587-022-01261-x.

98. Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., and Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. *23*, 258. https://doi.org/10.1186/s13059-022-02823-7.

99. Lovell, J.T., Bentley, N.B., Bhattarai, G., Jenkins, J.W., Sreedasyam, A., Alarcon, Y., Bock, C., Boston, L.B., Carlson, J., Cervantes, K., et al. (2021). Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding. Nat. Commun. *12*, 4125. https://doi.org/10.1038/s41467-021-24328-w.

100. Gabriel, L., Brůna, T., Hoff, K.J., Ebel, M., Lomsadze, A., Borodovsky, M., and Stanke, M. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. *34*, 769–777. https://doi.org/10.1101/gr.278090.123.

101. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

102. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

103. Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E.V., and Zdobnov, E.M. (2023). OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. *51*, D445–D451. https://doi.org/10.1093/nar/gkac998.

104. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. *20*, 275. https://doi.org/10.1186/s13059-019-1905-y.

105. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

106. Marrano, A., Britton, M., Zaini, P.A., Zimin, A.V., Workman, R.E., Puiu, D., Bianco, L., Pierro, E.A.D., Allen, B.J., Chakraborty, S., et al. (2020). High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome. GigaScience *9*, giaa050. https://doi.org/10.1093/gigascience/giaa050.

107. Fitz-Gibbon, S., Mead, A., O'Donnell, S., Li, Z.Z., Escalona, M., Beraut, E., Sacco, S., Marimuthu, M.P.A., Nguyen, O., and Sork, V.L. (2023). Reference genome of California walnut, *Juglans californica*, and resemblance with other genomes in the order Fagales. J. Hered. *114*, 570–579. https://doi.org/10.1093/jhered/esad036.

108. Zhu, T., Wang, L., You, F.M., Rodriguez, J.C., Deal, K.R., Chen, L., Li, J., Chakraborty, S., Balan, B., Jiang, C.Z., et al. (2019). Sequencing a *Juglans regia* × *J. microcarpa* hybrid yields high-quality genome assemblies of parental species. Hortic. Res. *6*, 55. https://doi.org/10.1038/s41438-019-0139-1.

109. Zhang, W.P., Cao, L., Lin, X.R., Ding, Y.M., Liang, Y., Zhang, D.Y., Pang, E.L., Renner, S.S., and Bai, W.N. (2022). Dead-end hybridization in walnut trees revealed by large-scale genomic sequence data. Mol. Biol. Evol. *39*, msab308. https://doi.org/10.1093/molbev/msab308.

110. Li, X., Cai, K., Zhang, Q., Pei, X., Chen, S., Jiang, L., Han, Z., Zhao, M., Li, Y., Zhang, X., et al. (2022). The Manchurian walnut genome: insights into juglone and lipid biosynthesis. GigaScience *11*, giac057. https://doi.org/10.1093/gigascience/giac057.

111. Guzman-Torres, C.R., Trybulec, E., LeVasseur, H., Akella, H., Amee, M., Strickland, E., Pauloski, N., Williams, M., Romero-Severson, J., Hoban, S., et al. (2024). Conserving a threatened North American walnut: a chromosome-scale reference genome for butternut (*Juglans cinerea*). G3 (Bethesda) *14*, jkad189. https://doi.org/10.1093/g3journal/jkad189.

112. Zhang, W.P., Ding, Y.M., Cao, Y., Li, P., Yang, Y., Pang, X.X., Bai, W.N., and Zhang, D.Y. (2024). Uncovering ghost introgression through genomic analysis of a distinct eastern Asian hickory species. Plant J. *119*, 1386–1399. https://doi.org/10.1111/tpj.16859.

113. Zhou, H., Zhang, X., Liu, H., Ma, J., Hao, F., Ye, H., Wang, Y., Zhang, S., Yue, M., and Zhao, P. (2024). Chromosome-level genome assembly of *Platycarya strobilacea*. Sci. Data *11*, 269. https://doi.org/10.1038/s41597-024-03107-4.

114. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics *37*, 1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016.

115. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

116. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

117. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

118. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19*, 153. https://doi.org/10.1186/s12859-018-2129-y.

119. Zhang, C., and Mirarab, S. (2022). Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. Mol. Biol. Evol. *39*, msac215. https://doi.org/10.1093/molbev/msac215.

120. Emms, D., and Kelly, S. (2018). STAG: species tree inference from all genes. Preprint at bioRxiv. https://doi.org/10.1101/267914.

121. Yang, Z., and Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. *23*, 212–226. https://doi.org/10.1093/molbev/msj024.

122. Puttick, M.N. (2019). MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics *35*, 5321–5322. https://doi.org/10.1093/bioinformatics/btz554.

123. Friis, E.M., Pedersen, K.R., and Schönenberger, J. (2006). Normapolles plants: a prominent component of the Cretaceous rosid diversification. Plant Syst. Evol. *260*, 107–140. https://doi.org/10.1007/s00606-006-0440-y.

124. Wolfe, J.A. (1973). Fossil forms of Amentiferae. Brittonia *25*, 334–355. https://doi.org/10.2307/2805639.

125. Heřmanová, Z., Kvaček, J., and Friis, E.M. (2011). *Budvaricarpus serialis* Knobloch & Mai, an unusual new member of the Normapolles complex from the Late Cretaceous of the Czech Republic. Int. J. Plant Sci. *172*, 285–293. https://doi.org/10.1086/657278.

126. Manchester, S.R., and Dilcher, D.L. (1997). Reproductive and vegetative morphology of *Polyptera* (Juglandaceae) from the Paleocene of Wyoming and Montana. Am. J. Bot. *84*, 649. https://doi.org/10.2307/2445902.

127. Belt, E.S., Hartman, J.H., Diemer, J.A., Kroeger, T.J., Tibert, N.E., and Curran, H.A. (2004). Unconformities and age relationships, Tongue River and older members of the Fort Union Formation (Paleocene), western Williston Basin, USA. Rocky Mt. Geol. *39*, 113–140. https://doi.org/10.2113/39.2.113.

128. Manos, P.S., Soltis, P.S., Soltis, D.E., Manchester, S.R., Oh, S.H., Bell, C. D., Dilcher, D.L., and Stone, D.E. (2007). Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. Syst. Biol. *56*, 412–430. https://doi.org/10.1080/10635150701408523.

129. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

130. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

131. Song, B., Marco-Sola, S., Moreto, M., Johnson, L., Buckler, E.S., and Stitzer, M.C. (2022). AnchorWave: Sensitive alignment of genomes with high sequence diversity, extensive structural polymorphism, and whole-genome duplication. Proc. Natl. Acad. Sci. USA *119*, e2113075119. https://doi.org/10.1073/pnas.2113075119.

132. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

133. Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. *108*, 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

134. To, T.H., Jung, M., Lycett, S., and Gascuel, O. (2016). Fast dating using least-squares criteria and algorithms. Syst. Biol. *65*, 82–97. https://doi.org/10.1093/sysbio/syv068.

135. Zumajo-Cardona, C., Little, D.P., Stevenson, D., and Ambrose, B.A. (2021). Expression analyses in *Ginkgo biloba* provide new insights into the evolution and development of the seed. Sci. Rep. *11*, 21995. https://doi.org/10.1038/s41598-021-01483-0.

136. Chakraborty, S., Britton, M., Martínez-García, P.J., and Dandekar, A.M. (2016). Deep RNA-seq profile reveals biodiversity, plant–microbe interactions and a large family of NBS-LRR resistance genes in walnut (*Juglans regia*) tissues. AMB Express *6*, 12. https://doi.org/10.1186/s13568-016-0182-3.

137. Hardwood Genomics Team. (2015). Comparative genomics of environmental stress responses in North American hardwoods. NCBI. https://www.ncbi.nlm.nih.gov/bioproject/273267.

138. Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., and Penin, A.A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. Plant J. *88*, 1058–1070. https://doi.org/10.1111/tpj.13312.

139. Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nat. Genet. *37*, 501–506. https://doi.org/10.1038/ng1543.

140. Jiang, H., Lei, R., Ding, S.W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics *15*, 182. https://doi.org/10.1186/1471-2105-15-182.

141. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419. https://doi.org/10.1038/nmeth.4197.

142. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

143. Qu, Y., Kong, W., Wang, Q., and Fu, X. (2021). Genome-wide identification MIKC-Type MADS-Box gene family and their roles during development of floral buds in wheel wingnut (*Cyclocarya paliurus*). Int. J. Mol. Sci. *22*, 10128. https://doi.org/10.3390/ijms221810128.

144. Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. J. Comput. Biol. *7*, 203–214. https://doi.org/10.1089/10665270050081478.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Pterocarya & *Cyclocarya* plants	This study	See Data S1A and S1C
Deposited data		
Whole genome sequence data (Pacbio and Illumina)	This study	Genbank: PRJNA1109986
RNA sequence data	This study	GenBank: PRJNA1123722
Genome assemblies	This study	GenBank: PRJNA1234227, GenBank: PRJNA1234228, GenBank: PRJNA1238886, GenBank: PRJNA1238887, GenBank: PRJNA1238908, GenBank: PRJNA1238909, GenBank: PRJNA1239020, GenBank: PRJNA1239021
Software and algorithms		
Hifiasm 0.19.9	https://doi.org/10.1038/s41592-020-01056-5	https://github.com/chhylp123/hifiasm
Ragtag v2.1.0	https://doi.org/10.1186/s13059-022-02823-7	https://github.com/malonge/RagTag
BRAKER3	doi: 10.1101/gr.278090.123	https://github.com/Gaius-Augustus/BRAKER
STAR 2.7.11	https://doi.org/10.1093/bioinformatics/bts635	https://github.com/alexdobin/STAR
EDTA v2.2.2	https://doi.org/10.1186/s13059-019-1905-y	https://github.com/oushujun/EDTA
RepeatModeler 2.0.5	https://doi.org/10.1073/pnas.1921046117	https://github.com/Dfam-consortium/RepeatModeler
RepeatMasker 4.1.7	https://www.repeatmasker.org/RepeatMasker/	https://www.repeatmasker.org/RepeatMasker/
OrthoFinder 2.5.5	https://doi.org/10.1186/s13059-019-1832-y	https://github.com/davidemms/OrthoFinder
muscle 5.1	https://doi.org/10.1186/1471-2105-5-113	https://github.com/rcedgar/muscle
IQ-TREE 2.2.5	https://doi.org/10.1093/molbev/msu300	https://iqtree.github.io/
ASTRAL-IV	https://doi.org/10.1186/s12859-018-2129-y	https://github.com/chaoszhang/ASTER
MCMCTree (paml 4.10.7)	https://doi.org/10.1093/molbev/msr045	https://github.com/abacus-gene/paml
skewer 0.2.2	https://doi.org/10.1186/1471-2105-15-182	https://github.com/relipmoc/skewer
Salmon 0.13.1	https://doi.org/10.1038/nmeth.4197	https://github.com/COMBINE-lab/salmon
DESeq2 1.44.0	https://doi.org/10.1186/s13059-014-0550-8	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
bwa 0.7.17	https://doi.org/10.1093/bioinformatics/btp324	https://github.com/lh3/bwa
samtools 1.17	https://doi.org/10.1093/gigascience/giab008	https://www.htslib.org/
bcftools 1.17	https://doi.org/10.1093/gigascience/giab008	https://github.com/samtools/bcftools
gemma 0.98.3	https://doi.org/10.1038/ng.2310	https://github.com/genetics-statistics/GEMMA
anchorwave v1.2.1	https://doi.org/10.1073/pnas.2113075119	https://github.com/baoxingsong/AnchorWave
minimap2 2.26	https://doi.org/10.1093/bioinformatics/bty191	https://github.com/lh3/minimap2
Beagle 5.5	https://doi.org/10.1016/j.ajhg.2021.08.005	https://faculty.washington.edu/browning/beagle/beagle.html
BLAST	https://doi.org/10.1016/S0022-2836(05)80360-2	https://blast.ncbi.nlm.nih.gov/Blast.cgi
RNAfold WebServer	https://doi.org/10.1093/nar/gkn188	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We examined phenotypes or collected tissue from live plants of the following species: *Pterocarya stenoptera P. macroptera*, *P. rhoifolia*, *P. fraxinifolia*, and *Cyclocarya paliurus*. Plants were growing outdoors, not under controlled conditions. Metadata associated with sequenced plants are provided in Data S1A and S1C. We examined published fossil descriptions and sequence data from additional species in Juglandaceae as described in STAR Methods.

METHOD DETAILS

Phenotyping

Flower structures were examined in spring of 2024 and dichogamy types were assigned based on the relative developmental stages of pistillate and staminate flowers. We studied *P. stenoptera* trees, all offspring from open-pollination, at four localities. Brief descriptions of each site follow. (1) Davis, CA. A subset of trees occurred on UC Davis campus and were evidently planted near the mid-20th century. One of these, at the Chancellor's residence, is attributed to Knowles Ryerson (appointed at UC Davis 1937-1977) and is thought to be among the first trees of this species planted in North America (Hunter Wilson, UC Davis, personal communication). Another stand of trees was identified in a marginal agricultural habitat at the outskirts of campus; these were not likely intentionally planted, and appear to be closely related to the campus trees (Figure S1C), perhaps even direct descendants. (2) Winters, CA. These trees are part of the USDA National Clonal Germplasm Repository housed at Wolfskill Experimental Orchard. Records indicate that many of these trees are derived from wild-direct seed collections, in several cases with multiple accessions being collected from different seeds from the same maternal tree. (3) Sacramento, CA. A handful of trees were located among the city-maintained trees. These were clearly intentionally planted and were plausibly several decades old; we could not find further historical information. (4) Sonoma Botanical Garden, Glen Ellen, CA. These trees consisted of two sets of wild-direct seed collections (from two different maternal trees) from different localities in Sichuan Province, China. In addition, a single tree each of *P. macroptera*, *P. rhoifolia*, and *C. paliurus* were observed for flowering phenotype at Sonoma Botanical Garden. Accession numbers of trees, where available, are given in Data S1A and S1C.

Most trees in flower that were observed over multiple time points across the flowering season could be confidently assigned to one dichogamy type. Others were observed on a single day during flowering, and for these dichogamy type was often nonetheless clear but not in all cases. Certainty in the phenotype was greater for trees growing nearby others of the same species experiencing the same local conditions. Binoculars were sometimes required to view flowering structures in the canopy, and here the pink tinge of stigmas was found to be a reliable indicator of the pistillate flowering phase. In concert with an overall assessment of dichogamy type, we took two quantitative measurements for a subset of trees observed. Pistillate catkin lengths of trees in Winters and Davis, CA were measured using a hand ruler within a two-day period in early April near the middle of the flowering season. Staminate catkin buds on the same set of individuals were measured on Dec. 17 of the same year (during dormancy) under a dissecting microscope, and both length and width were measured using the software ToupView. As a measure of male catkin size, we used each tree's value along the first principal component from a principal components analysis of both length and width (Figure 1C). We found this to be a more useful measure of size at this developmental stage, when variation in bud length and width are comparable in magnitude (bud length was on average 2.2x bud width).

Genome assembly and annotation

We assembled chromosome-level genomes of one individual each of *Pterocarya stenoptera*, *P. fraxinifolia*, *P. macroptera*, and *Cyclocarya paliurus* (Data S1A). The *P. macroptera* and *P. fraxinifolia* assemblies represent the first complete genome assemblies of these species to our knowledge.

Fresh leaf tissue was collected in the field and immediately flash frozen in liquid nitrogen. High Molecular Weight DNA extraction, library preparation, and sequencing on a PacBio Revio machine was performed by MedGenome Inc. (Foster City, CA). For our *C. paliurus* assembly the DNA was extracted instead from staminate catkin tissue prior to anther dehiscence. HiFi reads for each individual were assembled into contigs using *hifiasm* 0.19.9.[97] As *Cyclocarya paliurus* can be either diploid or polyploid, we first visualized a k-mer count distribution and saw only 2 clear peaks, one at the expected coverage for a haploid complement given the amount of sequence data generated and previous estimates of the haploid complement genome size, and another peak at half of this count value corresponding to heterozygous regions (additional peaks would be expected for a tetraploid). Thus, we concluded this individual was diploid and proceeded with the same assembly pipeline as for *Pterocarya*.

Contig assemblies were then scaffolded using a reference-guided approach[98] using a recent chromosome-level assembly of *P. stenoptera* which incorporated Hi-C data in scaffolding.[41] It is possible that the use of *P. stenoptera* to scaffold assemblies of *P. fraxinifolia* and *P. macroptera* could lead to assembly errors if there have been large-scale chromosome rearrangements between these species. We expect these to be minimal as chromosome number and synteny over broad chromosomal regions is strongly conserved within Juglandaceae.[52,99] For example, our *P. stenoptera* assembly shows few rearrangements relative to the *J. hindsii* walnut genome (Figure S2A).

We assessed the quality of assemblies using continuity measures (e.g. N50) and completeness metrics (e.g. BUSCO). Assembly metrics are reported in Data S1A. We also checked for canonical telomeric repeats (CCCTAAA/TTTAGGG) within 5kb of the chromosome-scale scaffolds, which is also reported in Data S1A.

We used the BRAKER3 pipeline[100] to annotate our *P. stenoptera* assemblies. The BRAKER3 pipeline incorporates both external evidence from RNA-seq alignments and protein sequences from related species together with *ab initio* predictions. We first identified and soft-masked repetitive sequence in assemblies using the combination of RepeatModeler2 and RepeatMasker.[101]

We used as input RNA-alignments generated from 12 sequencing libraries derived from male and floral tissue from 10 individuals (this study), combined with 4 published RNA-seq libraries from *P. stenoptera* derived from leaf, stem, root, and fruit.[41] We first used STAR[102] to perform splice-aware alignment of RNA-seq reads; these alignment files were used as input for BRAKER3. As additional external evidence we used the Viridiplantea partition of the protein OrthoDB v11 protein sequence database[103] hosted at https://bioinf.uni-greifswald.de/bioinf/partitioned_odb11/. We note this database includes protein coding sequence from some close relatives, including two species of *Juglans*. To account for the possibility of differential gene content at the locus of interest, we performed annotation separately for our two haplotype resolved assemblies. The annotation pipeline does not predict UTR regions. For the candidate gene *GFAFL*, we manually annotated the UTRs by viewing the coverage extent of RNA-seq reads from our floral tissue libraries. We manually inspected the annotations across haplotypes at the focal locus using gene expression and comparison to other haplotypes in other species STAR Methods. For the annotation shown in Figure 2B, we note one small predicted gene in gray, other than *GFAFL* in red, shown within the G-haplotype which appeared to be an annotation error STAR Methods.

We further annotated transposable elements (TEs) in each assembly using the EDTA pipeline.[104] Spatial distributions of TE classes along a subset of chromosomes are visualized in (Figure S2B) and identify putative centromeric regions. We used bedtools[105] to merge overlapping repeat annotations to compute repetitive sequence content in predefined genomic windows. For example in Figure S2C we compare the repetitive content of G-locus haplotypes to their local genomic context calculated content across 50 flanking windows of equivalent size on either side of each haplotype.

Species tree inference

We inferred a well-resolved fossil-calibrated phylogeny using whole-genome sequence data for twenty species within Juglandaceae.

Taxon sampling

Our sample includes five out of six recognized species of *Pterocarya* (according to Plants of the World Online database, https://powo.science.kew.org/, accessed January 2025). A recent study of relationships within *Pterocarya* using reduced representation sequencing of the nuclear genome[45] recovered monophyly of two sections traditionally recognized on the basis of morphology.[53] Here, we included five *Pterocarya* species for which whole genome sequencing data are available. Three have long-read genome assemblies presented here, *P. rhoifolia* was shotgun sequenced in Groh et al.,[19] and *P. hupehensis* was shotgun sequenced in Geng et al.[54]

In addition to our sampling of *Pterocarya*, we selected taxa with available whole-genome-sequence data from across Juglandaceae in order capture the phylogenetic diversity within the family. For *Juglans*, we chose species representing four sections traditionally recognized within the genus.[53] (1) Sect. *Juglans*: *J. regia*[106]; (2) sect. *Rhysocaryon* (black walnuts): *J. hindsii*,[19] *J. californica*,[107] *J. microcarpa*,[108] and *J. nigra*[109]; (3) sect. *Cardiocaryon*: *J. mandhsurica*[110]; (4) sect. *Trachycaryon*: *J. cinerea*.[111] For *Carya*, we chose taxa spanning the North American - East Asian divergence.[46] Species representing the North American clade were pecan (*C. illinoinensis*)[99] and *C. ovata*,[19] and species representing the East Asian clade were *C. sinensis* and *C. cathayensis*.[112] Other genomes were from the following sources: *Cyclocarya paliurus*,[42] *Platycarya strobilaceae*[113]; *Engelhardia roxburghiana* and *Rhoiptelea chiliantha*.[52]

Data preparation

The input to our phylogeny inference was a set of orthogroups of protein-coding sequences. For species with published annotations that incorporated RNA-seq data, we used annotations of protein-coding genes directly. For species with no published annotations or having annotations that did not incorporate RNA-seq data from multiple tissues, we used *liftoff*[114] to identify coding sequence coordinates of protein-coding genes in target assemblies using a reference annotation from a closely related species. For genes with multiple annotated slice variants, we filtered coding sequences so that each gene model was represented by the primary transcript. The final number of transcripts used as input for the orthology search across all species ranged from 27962 to 35370.

For the two species of *Pterocarya* for which we relied on short read resequencing datasets, we constructed pseudo-haploid sequences across annotated exons in our *P. stenoptera* reference genome. To do this, we generate a vcf containing both variant and invariant sites filtered for minimum genotype quality 30 (variant sites only) and a sequencing minimum depth of 5 (all sites). We then used the vcf to extract genotypes within the CDS coordinates of the primary transcripts for all gene models of our *P. stenoptera* annotation using *bcftools consensus*, encoding sites with zero coverage after filtering (-e 'FMT/GT=="./."' -a "N" -M "N" –mark-del "N") as 'N' in the output files and using IUPAC ambiguity codes to represent heterozygous sites. Consensus sequences were then reverse complemented if the gene annotation was present on the reverse strand in the reference sequence. We then used a custom python script to replace ambiguous IUPAC codes representing heterozygote sites with an appropriate nucleotide chosen at random from the possible bases (except for 'N's, which were kept in place). The same procedure was used to construct a pseudohaploid sequence of *Carya ovata* using the *C. illinoinensis* reference genome.

We then ran OrthoFinder 2.5.5[115] with DNA sequence as input (option -d) to identify ortholog groups. 96.0 % of genes were classified into 35,260 orthogroups with a median orthogroup size of 19. OrthoFinder identified 5,089 single-copy orthologs, and 13,652 orthogroups in which all species were represented. The first set of single-copy orthologs was used for the basis of the phylogeny results presented in the main text, and the second set was used as input for a complementary phylogenetic method applied.

CellPress

Phylogeny inference

We used both coalescent-based and concatenation approaches to infer the phylogeny. Trees were rooted with *Rhoiptelea chiliantha* as the outgroup. In order to run ASTRAL, for each single copy ortholog group, we performed alignment using muscle,[116] and inferred gene trees in IQ-tree,[117] specifying a GTR+I+G substitution model and 1000 ultrafast bootstrap replicates. We collapsed nodes with low bootstrap support (<10%) to polytomies[118] using newick tools https://github.com/xflouris/newick-tools, and then ran ASTRAL-IV v1.19.4.6[119] on the resulting gene trees to infer a species tree under a model consistent with incomplete lineage sorting. We report Local Posterior Probabilities for internal nodes computed by ASTRAL in the main text.

Separately, to perform maximum-likelihood inference, we concatenated alignments of single-copy orthologs to obtain an alignment with 443,387 parsimony-informative sites. The maximum likelihood phylogeny was found in IQ-tree 2.2.5[117] and in the maintext we report bootstrap confidence values for internal nodes from 1000 replicates of the ultrafast bootstrap implemented in IQ-tree.

Finally, as a complementary approach for phylogenetic inference, we examined the tree inferred by STAG,[120] which is incorporated into the OrthoFinder pipeline. STAG is a gene tree summary method using all ortholog groups in which all species are represented.

All three of these methods inferred the same topology (Figure 1) with high confidence. The ASTRAL tree had Local Posterior Probabilities of 1 at each internal node, and the maximum likelihood tree had boostrap values of 100% at all internal nodes. Given the strong support for this topology, we treated the topology as fixed in a second step of time-calibration.

Time-calibration

We performed Bayesian estimation of divergence times using MCMCTree.[49,121] Briefly, this method allows the user to specify priors on node ages, and obtains posterior distributions through Markov-Chain-Monte-Carlo (MCMC) sampling that makes use of an approximation to the likelihood around the maximum likelihood estimate (MLE). We specified a HKY85 + Γ substitution model with 5 discrete rate categories, and an independent-rates relaxed-clock model of substitution rates. The MCMC algorithm was run with a burnin of 50,000 steps, a sampling frequency of 100 steps, and a sample size of 100,000. We ran two separate chains to check for convergence. We also simulated under the prior (Figure S1D) and compared posterior distributions to prior distributions for nodes of interest (Figure S1E). We use fossil information to specify prior distributions on internal nodes of the Juglandaceae phylogeny. In each case, distributions were specifically tailored reflect our uncertainty in both fossil age and assignment. We also implemented MCMCTree's ability to use 'soft' bounds by placing a small prior mass below the minimum and/or above the maximum bounds, accounting for the possibility that nodes are actually younger (i.e. due to inaccurate fossil assignment) or older (i.e. due to a sparse preservation/detection of fossils) than the specified bounds. MCMCTree then uses a birth-death model to generate a joint prior on all internal node ages lacking manually specified priors. Seven fossil-calibrated nodes are listed below in order from oldest to youngest.

We note that in general, specifying a minimum age constraint for a daughter node implies the same minimum age constraint for a parent node. In this sense, a fossil constraining the minimum age of the parent node, which is no older in age to a fossil constraining the daughter node, would provide no additional information. However, if the minimum bound on a daughter node is 'soft', this relaxes the constraint on parent nodes, and a younger fossil constraining the parent node with a 'hard' minimum bound can provide non-redundant information. Following this logic, here we use minimum age constraints on both the Juglandoideae-Engelhardiodeae divergence (parent node) and the *Pterocarya-Cyclocarya* divergence (daughter node) based on fossils of similar age, but treat the minimum constraint on the former as 'hard' and the minimum age constraint on the latter as 'soft' (justification provided below). We used the R package MCMCTreeR[122] to identify parameters specifying the desired prior distributions. The joint prior on all internal nodes is shown in Figure S1D, with a subset of nodes numbered in correspondence with the list of calibration points below.

Crown age of Juglandaceae.

Min: 70 Mya
Max: 96.5 Mya

Here, we follow the Angiosperm Phylogeny Group IV classification of Juglandaceae which includes *Rhoiptelea*.[48] Elsewhere, *Rhoiptelea* is regarded as the sole member of the family Rhoipteleaceae; unlike other Juglandaceae it possesses morphologically cosexual flowers (and pistillate flowers) and a superior ovary.

Juglandaceae is considered to have likely derived from members of the Normapolles complex, a diverse fossil assemblage of the Late Cretaceous and tertiary with morphological affinities to the modern order Fagales.[123] Normapolles pollen and associated flower structures first appear in the Upper Cenomanian and are abundant throughout the remainder of the Cretaceous.[123,124] Considering Fagales (which includes 6 other families in addition to Juglandaceae) to have derived from the members of the Normapolles complex, pollen of which is common in late Cretaceous sediments, we set the Upper Cenomanian as a maximum age for this node, as this represents the earliest occurrence of Normapolles fossils.

Fossil flowers of *Budvaricarpus* from the Late Turonian-Santonian (minimum age 83.6 Mya) as described by Heřmanová et al.[125] bear close resemblance to those of extant *Rhoiptelea*. These authors suggested these flowers and those of *Caryanthus*, an affiliated member of the Normapolles group, as possible basal lineages of Juglandaceae (here, including *Rhoiptelea*). Other workers have noted close affinity of pollen from the Late Cretaceous to that of *Rhoiptelea* e.g. Wolfe.[124] Reliable fossils of other genera within Juglandaceae have not been confirmed prior to the Teriary but nonetheless appear close in time to the Cretaceous-Paleogene boundary.[47] Thus, we consider it highly likely that the divergence between *Rhoiptelea* and other Juglandaceae genera occurred prior to the

end of the Cretaceous, and set 70 Mya as a minimum age for this node. We specify a modified uniform prior on the node, with 2.5% of the prior mass below the minimum and similarly 2.5% of the mass above the maximum bound.

MRCA of (Juglandoideae, Engelhardioideae).

Min: 65 Mya (hard bound)

Fossils of *Polyptera manningii* and *Cyclocarya brownii* are recorded from the Paleocene.[29,32,126] Both are known from the Fort Union Formation, which has associated radiometric date estimates ranging from 61.03-64.73 Mya.[127] Both of these taxa are members of Juglandoideae, judged on the basis of well-preserved vegetative and floral structures.[126,128] Similarly, the earliest known fossil fruit assigned to *Cyclocarya* is recorded within a stratigraphic series encompassing the first 1 million years after the K-Pg mass extinction event with multiple associated radiometric dates.[29] We thus set 65 Mya as a hard minimum bound for the divergence between Juglandoideae and Engelhardioideae, specifying a truncated Cauchy prior distribution in MCMCTree with a heavy right tail.

*MRCA: ((*Juglans*, (*Pterocarya*,* Cyclocarya*)),* Platycarya*)*.

Min: 47 Mya

The fossil genus *Cruciptera* from Eocene and Oligocene sediments of Western North America has been assigned with reasonable confidence to the crown group of Juglandeae (extant *Juglans*, *Pterocarya*, *Cyclocarya*)[30,128] on the basis of morphology. Radiometric dating of associated formations range from 43 to 47 million years old.[30] We set this as the minimum age constraint for the stem age of Juglandeae (divergence from *Platycarya*) using a truncated Cauchy distribution with a hard lower bound at 43 Mya and a heavy right tail.

*MRCA of (*Cyclocarya*,* Pterocarya*)*.

Min: 65 Mya (soft bound)

Cyclocarya possess distinctive radially symmetric winged fruits, which are recognizable within the first one million years after the Cretaceous–Paleogene mass extinction.[29] However, as reliable fossils of *Pterocarya* are not known prior to the Oligocene (30 million years later than the earliest records of *Cyclocarya* fruits), we considered it plausible that early *Cyclocarya* fossils could represent stem lineages of extant *Cyclocarya* and *Pterocarya*. To reflect this uncertainty, we used a truncated Cauchy distribution with a 'soft' minimum bound at 65 Mya, a heavy right tail, and a mode at approximately 66 Mya. We allowed 20% of the prior mass for this node to fall below 65 Mya using a decaying exponential function over younger possible ages. Thus, the prior still strongly favors an older divergence for these taxa, as implied by the current taxonomic treatment of fossils assigned to *Cyclocarya*.

Juglans *crown age*.

Min: 28 Mya
Max: 56 Mya

Manchester[47] reviews multiple fossil fruits attributable to sections *Rhysocaryon* (black walnuts, e.g. *J. siouxensis*) and section *Cardiocaryon* (butternuts, e.g. *J. lacunosa*) from the Early to Mid Oligocene. The divergence of these two sections represents the deepest divergence within the crown group of *Juglans. Juglans hindsii* represents a member of the black walnuts, while *J. mandshurica* represents a member of the butternuts. We therefore consider that the divergence of these sections occurred prior to the middle of the Oligocene period, and set a minimum age constraint for the MRCA of extant *Juglans* at 28 Mya. The first unambiguous *Juglans* fossil fruits are from the Mid Eocene, so we set the start of the Eocene as a maximum age constraint for this node.

We used a gamma distribution for the prior on this age, numerically finding the shape and scale parameters such that the prior mean was placed at 42 Mya, 2.5% of the prior mass was below the lower bound, and 10% of the prior mass was above the upper bound.

Carya *crown age*.

Min: 23 Mya
Max: 56 Mya

As discussed in Manchester,[47] pollen of *Caryapollenites* appears in the Upper Paleocene and is similar to that of modern *Carya*, however its comparatively smaller size, together with the absence of definitive *Carya* macrofossils, suggest the crown *Carya* emerged after the Paleocene. We thus set the end of the Paleocene as an upper bound on this node. *Carya* is traditionally divided into four sections (*Carya*, *Apocarya*, *Sinocarya*, and *Ramphocarya*), with extant members of the former two distributed in North America, and extant members of the latter two distributed in East Asia. Molecular phylogenies have supported monophyly of North American and East Asian clades.[19,46,112] Manchester[47] reviews several fruit fossils which bear close affinity to recognized sections. For example, *C. quadrangula*, known from the Middle Oligocene appears to represent section *Apocarya*, while *C. strychnina*, known the Upper Oligocene, appears to represent section *Sinocarya*. The presence of fossils assigned to these two sections by the end of the Oligocene implies a divergence before the end of the Oligocene, so we set 23 Mya as a lower bound on this node. We specified a uniform prior between the minimum and maximum with soft bounds (2.5% prior mass below the lower and above the upper bound).

MRCA: (*Pterocarya* sect. *Pterocarya*, *Pterocarya* sect. *Platyptera*)

Min: 5.3 Mya
Max: 56 Mya

Traditional classification of *Pterocarya*[53] groups species into two sections, *Pterocarya* (e.g. *P. stenoptera*) and *Platyptera* (e.g. *P. macroptera*). These two sections were recently supported as monophyletic in what we consider the most comprehensive molecular phylogeny of the genus to date based on multiple nuclear loci.[45] Pollen similar to that of modern *Pterocarya* occurs in Eocene, and becomes more abundant through the Oligocene and Miocene.[47] There are some isolated reports of *Pterocarya*-like pollen from the Paleocene, but these warrant reinvestigation.[47] Unambiguous fossil fruits of *Pterocarya* are not known prior to the Oligocene. Fossils that bear close affinity to the two modern sections (e.g. *P. smileyi*, sect. *Pterocara* and *P. asymmetrosa*, sect. *Platyptera*) are known from the Miocene.[47] We thus specify a uniform prior for the age of this node, with soft lower and upper bounds between the start of the Eocene and the end of the Miocene.

Short read sequence alignment, variant identification and filtering

For short read data, reads were mapped to the haplotype assemblies of *P. stenoptera* using *bwa* 0.7.17.[129] Variants were called in *bcftools* 1.17[130] and genotypes were filtered depending on the purposes of the analysis. SNPs tested for an association with dichogamy type were filtered using criteria of GQ 30 and minimum depth 10. For direct ascertainment of segregating sites in *GFAFL* copies within individuals, we used allele counts directly, setting a frequency cutoff of 0.05 for reads matching the within-individual minor allele.

Investigation of candidate genes

To investigate possible functional candidates within the G-locus, we manually cross referenced our denovo automated annotations of the *P. stenoptera* haplotypes, with RNAseq data from all available tissues of *P. stenoptera* from this study (multiple developmental stages of both male and female flower tissue) and previous studies including vegetative tissues.[19,41] We used the following criteria for considering any predicted protein-coding gene model as a causal determinant for the control of heterodichogamy. (1) Unambiguous evidence that the gene is transcribed and encodes a functional protein and (2) conserved presence in two haplotypes of the same allele in different species. Our analysis here is focused on expressed protein-coding genes, and we cannot rule out that non-coding regulatory sequences not captured by our sequencing strategy could be causal factors. The possibility of small RNAs transcribed from the dominant haplotype is discussed in the main text.

GFAFL is the only annotated gene that is conserved across all G-locus haplotype assemblies. In addition, the automated annotation predicted one unique gene in each of the *P. stenoptera* haplotypes. Because hemizygous genes are known to control other mating type systems, we carefully examined these for possible expression and putative function.

In the *G* haplotype, the automated annotation pipeline had predicted an additional 93 amino-acid open reading frame that was not present in the recessive haplotype annotation. A BLAST search revealed sequences with high similarity in variable numbers (~2-6 copies across *Pterocarya* assemblies), but no other assemblies showed similar sequence at the syntenic location on chromosome 11. Given that this predicted gene model is unique to this assembly (specifically, it is not found in the dominant haplotype of *P. macroptera*), it does not meet criterion (2) above. Examining RNAseq data from protogynous individuals, we found that only 5 out of 17 samples showed any evidence of reads aligning here. These 5 samples were not from the same tissue, and there was no tissue where all protogynous samples showed read alignments over this predicted gene. Among these 5 samples, we saw at most 3 non-duplicate properly-aligned reads that did not show significant read clipping. All of these were aligned to the same subregion of the predicted exon with numerous sequence mismatches. In none of these samples did we see expression spanning the entirety of the predicted exon. All of the reads that overlapped one particular location within the first half of the predicted single-exon gene showed a 2-bp indel that would render an early stop codon if translated. Thus, this predicted gene does not satisfy criterion (1). Most importantly, we saw these same patterns of read alignments to this region from protandrous individuals, which we confirmed do not possess this haplotype using whole genome sequence data. Therefore these RNAseq reads cannot have originated from this predicted gene that is unique to the *G* haplotype. These facts led us to conclude that this predicted gene model was a false positive, which we note is a widespread problem in genome annotation. We therefore omitted this predicted gene in the main text (Figure 1B).

Turning to the expression data for the recessive *g* haplotype of *P. stenoptera*, we investigated a predicted gene that was not annotated in the dominant haplotype. This predicted gene contains a single predicted exon with a 116 codon reading frame. A BLAST search identified homologous sequences in *Juglans* and *Carya* annotated as 'glutathione S-transferase T3-like'. We did not see any evidence of expression of this gene in any available floral or vegetative tissues from *P. stenoptera*. Therefore this gene does not satisfy criterion (1). Using BLAST we identified homologous sequence in both the dominant and recessive G-locus haplotypes of *P. macroptera*, and the recessive haplotype of *P. fraxinifolia*. However, each of these contained stop codon mutations when translated, so these do not encode functional proteins. Therefore it cannot satisfy criterion (2). We consider this gene prediction to most likely reflect a false positive and that this sequence is an unlikely candidate for the causal basis of heterodichogamy.

Lastly, using IsoSeq transcripts of *P. stenoptera* from Groh et al.,[19] we did find compelling evidence for an additional gene, not annotated by the automated pipeline, within the recessive haplotype of *P. stenoptera*. IsoSeq transcripts were detected in two replicate samples of bark tissue, but we did not see any expression in floral tissue. A BLAST search of this sequence identified homologs

across eudicots, monocots, and magnoliids with the annotation 'AtMg00810-like' based on homology to an Arabidopsis gene. Closely homologous sequences were detected in over 10 copies in each *Pterocarya* assembly, but importantly, none were found within the respective G-locus regions. Because this gene is not present in other recessive haplotypes in other species, it does meet criterion (2), and the pattern of bark-specific expression (and lacking in any floral tissue) does not suggest a probable link to flowering.

In contrast, *GFAFL* shows a conserved reading frame across all *Pterocarya* G-locus assemblies, and shows clear expression in staminate floral buds and mature staminate catkins, making it strong candidate for the regulation of heterodichogamy.

In RNAseq data of staminate catkin buds of *P. macroptera*, we detected negligible transcription of the *GFAFL* ortholog (<5 sequencing reads). However, we note that this particular sample was collected in late November, well after the growth season had completed. Expression of *GFAFL* is noticeably reduced in dormancy in *Cyclcoarya* (Figure 4C), and transcript counts in staminate buds in *P. stenoptera* collected in late summer were fairly low to begin with. Thus, it seems likely that we did not detect meaningful expression of *GFAFL P. macroptera* simply due to having sampled the tissue at the wrong time.

Supporting this idea, we did not detect biologically meaningful expression of *GFAFL* in dormant staminate buds of *Cyclocarya paliurus* collected on the same day in late November (Figure 4C), however we did detect clear expression of the *GFAFL* ortholog in mature staminate catkins prior to anthesis from the same individual tree (Figure 4C). Analysis of published datasets of *Cyclocarya* confirmed that *GFAFL* is expressed in flowers, but not during dormancy (Figures 4C and S4C) (we later determined this represented expression of two divergent alleles, described in main text). We also note that the *GFAFL* ortholog is clearly expressed in staminate floral tissue of two *Juglans* species (Figures S4D and S4E). In summary, *GFAFL* shows clear evidence of flower-specific expression in several systems, and together with its conserved presence in more than one of the same haplotype in different species, appears to be the strongest functional candidate for heterodichogamy among protein-coding sequences within the G-locus.

Genetic structure and GWAS

We explored genetic structure in our association panel using Principal Components Analysis (PCA) on biallelic SNPs after pruning pairs of sites with genotypic linkage disequilibrium of $r^2 > 0.2$. Samples cluster by sampling locality and according to known family relationships (Figure S1C). Both protandrous and protogynous morphs were observed at each sampling locality.

GWAS was performed in GEMMA 0.98.5, which implements a linear mixed effect model to control for relatedness, filtering out sites with MAF < 0.10 or with over 10% of samples having missing data. GWAS was done for alignments of short reads to both haplotypes of the diploid *P. stenoptera* individual which were assembled with hifiasm. Initial results appeared markedly different in that the GWAS against haplotype 2 revealed a single clear association peak (Figure 2A), while the GWAS against haplotype 1 revealed many scattered peaks throughout the genome. In the course of our previous work,[19] we observed a similar pattern when performing GWAS against assemblies containing the dominant vs. recessive haplotypes of the pecan G-locus, and determined this was a result of differential transposable element content between the haplotypes at the causal locus. Consistent with this hypothesis, for the GWAS against haplotype 1 where there were multiple peaks, we filtered SNPs that fell within annotations of repetitive sequence and observed that this removed the majority of peaks, while the peak at the syntenic location to that observed against haplotype 2 remained the most consistent with a true association (Figures S3B–S3E). A small number of high association SNPs remained at other locations but these lacked a gradual decline in association strength moving outwards from the associated SNP as would be expected based on linkage disequilibrium around a true causal locus embedded within a chromosome. As a further test of the differential TE hypothesis, we extracted local regions around the putative spurious associations in haplotype 1 and used BLAST to detect homology to regions in the haplotype 2 assembly. As expected, those regions we tested showed homology to sequence at the G-locus identified in haplotype 2. All of our subsequent analyses examining the G-locus shown in Figures 2 and 3 gave us high confidence that this peak is the true causal locus for heterodichogamy in *Pterocarya*.

Analysis of polymorphism and divergence within and between haplotypes

We performed whole-genome alignment using anchorwave[131] with settings '-R 1 -Q 1'. We calculated nucleotide divergence in 1 kb windows from the anchorwave alignments using a custom R script. In visualizing the divergence in the G-locus region, we filtered out divergence windows containing fewer than 500 aligned bp in or where >50% of the window was annotated as repeat content in the reference coordinates. We also removed windows that fell within the boundaries of the repeat-rich structural variant after identifying the boundaries using a dotplot from BLAST alignments.

We used minimap2[132] to obtain alignments of focal haplotypes in fixed coordinates of one haplotype assembly (i.e. Figure 3A) and called SNPs in bcftools setting ploidy to 1. For SNPs identified between long-read assemblies, we filtered out any singletons in the sample to focus on sites containing information about the topology of the local genealogy. To precisely locate the boundary of trans-species polymorphism, we manually examined SNPs near the border of the region where polymorphism transitions between a species tree topology to one of trans-species polymorphism. A small number of SNPs that occurred within 3-bp of indels were considered errors and filtered out.

We extracted a set of trans-species polymorphic SNPs between the *P. stenoptera* and *P. macroptera* haplotypes (Figure 3) and measured the fraction of these sites at which individuals were heterozygous in a set of independent samples from several *Pterocarya* species from.[19,54] As many of these sites were missing in the short read variant calls we considered these fractions as estimates from a larger number of sites and computed confidence intervals using the Agresti-Coull method.

To compare levels of polymorphism within and between G-locus haplotypes, we phased short-read sequencing data using Beagle 5.4.[133] Phasing was done within species for three species represented by multiple individuals. Phasing results were manually checked for phasing errors across the coding sequence of *GFAFL* by checking for phase switch errors in heterozygotes. We computed pairwise diversity using all pairs of phased haplotypes within each species sample for each haplotype using a custom R script (Figure 2F).

We computed pairwise dN/dS values bewteen the coding sequence regions of *GFAFL* using the R package ape. We used a custom R script to perform a MacDonald-Kreitman test, where nonsynonymous and synonymous divergence between long-read assemblies of *GFAFL* were compared to the levels of nonsynonymous and synonymous polymorphism seen within *g* haplotypes after phasing the data. We used polymorphism within *g* haplotypes as limited polymorphism within *G* haplotypes restricted the power of the test, nonetheless the result was not significant.

In the above analyses, we omitted a single sample of *P. stenoptera* which showed evidence of chimerism. This sample was derived from a protandrous tree and showed non-zero coverage over the hemizygous region whereas all other protandrous showed effectively zero coverage, but still lower coverage than protogynous samples (Figure 2C). Outside of the CNV, we directly examined SNPs with the aligned reads for this sample and observed numerous sites where the sample appeared heterozygous for SNPs associated with dichogamy type, but allelic depths at these SNPs deviated from what would be expected for an individual heterozygous for the two G-locus haplotypes. To formally test this, we first ascertained a set of high confidence SNPs where all protogynous *Gg* heterozygotes were heterozygous. We then measured the average across sites of the relative read depths of the reference vs. alternative alleles within each individual and compared these values among samples. Consistent with our hypothesis that the above protandrous sample was chimeric, we observed it showed a bias in allelic depth distinct from all other samples which showed heterozygosity at these SNPs (see Figures S3G and S3H and associated caption).

To examine segregation of *GFAFL* alleles in *Cyclocarya*, we combined our denovo assembly with assembly and resequencing data of *C. paliurus* from Qu et al.[42] We first identified *GFAFL* syntenic regions in these assemblies and plotted average read depth in genomic windows from resequencing samples across these regions, observing a segregating CNV in both diploids and tetraploids (Figures S6A–S6C). We used two complementary approaches to specifically test for segregation of the *G1* and *G2* alleles of *GFAFL*. First, we aligned all reads of both diploids and tetraploids to a chimeric haploid assembly containing both alleles. In this competitive mapping approach, given the high divergence between the alleles we observed *GFAFL* reads mapped to one or the other alleles with high affinity. For each sample, we took the average depths across the transcribed region of these two alleles and measured the relative depth across the two alleles to infer copy number in each sample. Second, we aligned all reads to an assembly containing only a single haploid copy of *GFAFL* to identify segregating sites within the *GFAFL* transcribed region, among all *GFAFL* copies in each sample, in order to test whether samples contained similar or highly diverged copies of *GFAFL*. We ascertained a site as segregating within an individual if the within-individual frequency of the minor allele within at least 0.05. As these samples varied widely in sequencing depths (<5 to >100) and we observed that the number ascertained segregating sites correlated with read depth, we applied a depth correction for the number of segregating sites as follows. The observed number of ascertained segregating sites S_{obs} is divided by the average across L sites (invariant and variant) of the probability of ascertainment at site ℓ, conditioned on site ℓ segregating and the observed read depth at site ℓ:

$$\widehat{S} = \frac{S_{obs}}{\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}\left[\mathbb{1}_{\ell ascertained} \middle| \mathbb{1}_{\ell segregating}, D_\ell\right]} \quad \text{(Equation 1)}$$

The denominator is the probability of ascertainment for site ℓ conditioned on site ℓ segregating and the observed read depth at site ℓ. This is computed under a binomial sampling model of sequencing reads where p is the within-individual allele frequency (across all copies aligned to the single-copy *GFAFL* region). For diploids, $p = 0.5$. For tetraploids, $p \in (0.25, 0.5)$ and will vary across sites depending on whether a variant exists between homologs or homeologs. For tetraploids with one *G2* allele which is highly divergent from *G1*, there will be only a small proportion of sites where $p = 0.5$ as most segregating sites will be variants in one of four copies, such that $p = 0.25$. For tetraploids with zero copies of *G2*, there will be fewer segregating sites but a greater proportion of these will have $p = 0.5$. However we do not see a strong signal of divergence between homeologs in the phylogeny and for simplicity we use $p = 0.25$ for all tetraploids, noting that this choice of p did not qualitatively impact our results.

$$\mathbb{E}\left[1_{\ell ascertained} \middle| 1_{\ell segregating}, D_\ell\right] = \begin{cases} 0 & D_\ell \in (0,1) \\ \sum_{k=\lceil 0.05 D_\ell \rceil}^{\lfloor 0.95 D_\ell \rfloor} \binom{D_\ell}{k} p^k (1-p)^{D_\ell - k} & D_\ell > 1 \end{cases} \quad \text{(Equation 2)}$$

Haplotype age estimation

To infer the divergence times between the haplotypes in *Pterocarya* and *Cyclocarya*, we used two complementary approaches. First, we used a calibration based on relative numbers of substitutions at fourfold degenerate sites within the coding sequence of *GFAFL*. We first obtained estimates of the number of substitutions at fourfold degenerate sites from codon alignments between pairs of haplotypes using ape, corrected these for multiple hits using the Jukes-Cantor substitution model, averaged these across multiple

pairwise comparisons, and then applied the molecular clock formula (i.e. $t_1 = t_2\left(\frac{s_1}{s_2}\right)$). For *Pterocarya*, we used the average divergence between the comparison of *P. stenoptera* haplotypes and the comparison of *P. macroptera* haplotypes. *Pterocarya-Cyclocarya* divergence estimates were averages across both comparisons of both alleles in *P. stenoptera* with both alleles in *Cyclocarya*, and the *Pterocarya-Juglans* estimate was an average of comparisons between both *P. stenoptera* haplotypes and *J. regia*.

Using *Pterocarya-Cyclocarya* and *Pterocarya-Juglans* as values s_2, together with generic divergence times inferred from our genome-wide phylogeny as values for t_2, gave estimates for the *Pterocarya* haplotype divergence at 40 Mya (95% CI 31-49) and 44 Mya (95% CI 35-52), respectively. Confidence intervals were obtained by propagating uncertainty from the 95% CI upper and lower bounds on generic divergence times for values of t_2.

For *Cyclcocarya*, the estimate for haplotype divergence at fourfold degenerate sites was an average of two different comparisons involving *G1* and *G2* alleles from different assemblies. Using the same generic divergence estimates as above (*Cyclcoarya-Pterocarya* and *Cyclcoarya-Juglans*) we obtained age estimates for the *Cyclocarya* haplotypes at 51 Mya (95% CI 40-61) and 55 Mya (95% CI 44-66).

As a complementary approach, we used a calibration based on the intron of *GFAFL*. We first produced an alignment of the intron containing 281 parsimony informative sites across orthologs in Juglandoideae (note the fourfold degenerate substitution estimates above ranged from 5-12 substitutions). We first inferred a maximum likelihood tree in IQtree, then used the least-squares dating algorithm implemented in IQtree,[134] specifying the age of an internal node. We calibrated this tree using the inferred divergence time of *Pterocarya* and *Cyclocarya* from our genome-wide species tree inference. Confidence intervals were obtained by repeating this calibration using the 95% credible interval bounds for the genus divergence time. This gave estimates of 51 Mya (95% CI 40-61) and 45 Mya (95% CI 35-54) for the *Pterocarya* and *Cyclocarya* haplotypes, respectively.

To summarize across these estimates, we take the mean of the point estimates, and report the minimum and maximum for the lower and upper bounds, respectively, of the credible intervals across these estimates.

Finally, we revisited our date estimates of the divergence ages between haplotypes at both the *Juglans* and *Carya* loci. We used the same molecular clock approaches described in Groh et al.,[19] recalibrated using posterior means and 95% credible intervals for the respective species divergence times that were used in previous calibrations. We obtain estimates of 40.48 (95% CI 33.44-47.12) for the *Juglans* G-locus and 59.47 Mya (95% CI 49.13-69.23) for the *Carya* G-locus, consistent with our previous estimates.

Relatedness and investigation of function of *GFAFL* homologs

We ran the predicted amino acid sequence of GFAFL through InterProScan, which identified a conserved FAF domain (PFAM entry PF11250) in amino acid residues 112-170. The FAF domain is a plant-specific domain of unknown molecular function named for a group of *FANTASTIC FOUR* paralogs in Arabidopsis that were originally identified as regulators of meristem size.[59] Since then, *FAF*-like genes have been identified in diverse vascular plants including gymnosperms and lycophytes.[135] While very few *FAF*-like genes have been functionally characterized, they have been found to regulate diverse developmental processes including meristem growth,[59] seed germination, drought tolerance, and primary root growth[60] and endosperm starch metabolism.[62]

To examine whether *GFAFL* showed close homology to any *FAF*-like genes of known function, we inferred a maximum likelihood phylogeny of *FAF*-like paralogs. We first identified paralogs within our newly assembled *Pterocarya* genomes using a protein blast search with the FAF domain of GFAFL as the query sequence. This identified 11 clear paralogs in *P. stenoptera* (blastp E Value <=8e-04). We repeated this paralog search procedure in *Juglans regia* and identified 11 paralogs using the same E value cutoff. We used a similar procedure targeting species in which *FAF*-like genes have been functionally characterized (Arabidopsis, maize, and tomato) and species selected to represent the diversity of embryophytes (*Selaginella* and *Amborella* and *Marchantia*) in order to provide information about the relatedness of *GFAFL* to functionally characterized *FAF*-like genes. We found a single copy *FAF*-like gene in *Marchantia polymorpha*, indicating that this gene family was present in the common ancestor of embryophytes (land plants). This sequence was used as an outgroup to root the tree. We performed codon-based alignment of nucleotide sequence encoding the FAF domain (~150 nt) using kc-align (https://github.com/davebx/kc-align) and inferred a maximum-likelihood phylogeny in IQtree using a GY+F model of codon evolution and 1000 ultrafast bootstrap replicates.[117] We also visualized a pre-constructed gene tree for this gene family in the Panther database (pantherdb.org) which suggested *GFAFL* is a member of *FAF* subfamily 17 (SF17, F2E2.18-related).

We then analyzed expression of orthologs of *GFAFL* in closely related species to provide further information about a conserved biological function. We examined expression data from multiple tissues of the sister genus *Cyclocarya*.[64] In this dataset which contains vegetative tissues and floral tissue, the *GFAFL* ortholog is uniquely expressed in flowers (Figure S4C). We note that we later identified expressed *GFAFL* transcripts in this dataset to originate from both *G1* and *G2 Cyclocarya* alleles. We then looked in two species from the next most closely related genus, *Juglans*. In *J. regia* we examined expression in 20 tissues using data from Chakraborty et al.,[136] provided by Houston Saxe, UC Davis. The *J. regia* ortholog is predominantly expressed is staminate catkins (Figure S4D). We also examined a dataset from *J. nigra*[137] and found a similar pattern of staminate flower-specific expression (Figure S4E).

We then looked further out to the two closest homologs of *GFAFL* in Arabidopsis. We note that in the phylogeny shown in Figure 3 constructed from only the FAF domain, *AT1G22110* and *AT1G77932* appear equally related to *GFAFL*. However, alignment over a broader region indicates that *AT1G22110* shows much greater sequence similarity (Figure S4A). In the expression atlas available at https://www.arabidopsis.org/ showing data from Klepikova et al.,[138] *AT1G22110* shows high expression in anthers and weak

expression in other tissues (Figure S4F). Expression data for this gene from a separate developmental map study[139] also show high expression in mature pollen and comparatively low or no expression other tissues (Figure S4G). This suggests that orthologs of *GFALF* might have a conserved functional role in relation to pollen development. The other homolog *AT1G77932* shows high expression in flower petals and little or no expression in other tissues (Figure S4H). Expression data for this gene were not found in the Schmid et al.[139] dataset.

Gene expression analysis

Sampling

We sampled mature or nearly mature flower tissues from near the middle of the flowering season from multiple individuals of each type of *Pterocarya stenoptera* in Davis, CA on March 20, 2024. Separately, we sampled pre-dormant staminate catkin buds from multiple individuals of each type from the same species on August 2nd, 2024. This latter sample was comprised of 10 individuals from Wolfskill Experimental Orchard in Winters, CA, plus a single tree from Davis, CA. We also sampled dormant staminate catkin buds from a single individual each of *Cyclocarya paliurus* and *Pterocarya macroptera* (the same individuals used for genome assembly) on November 24, 2024, and mature floral tissue from the same *Cyclocarya paliurus* individual on May 20, 2024. Locations, sampling dates, and tissue descriptions are given in Data S1D.

Analysis

RNAseq reads were first trimmed using skewer.[140] We directly aligned RNA-seq reads to the transcriptome using salmon.[141] To study allele-specific expression of *GFAFL* alleles and mitigate impacts of reference bias, we used a competitive mapping approach by including both *GFAFL* alleles in the reference transcriptome. To validate this approach, we separately used STAR[102] to perform splice-aware alignment of RNAseq reads to both assembly haplotypes and examining SNPs within these reads at sites differing between the two *GFAFL* alleles.

We first visualized transcript count estimates by tissue and flowering type, and finding that *GFAFL* was only expressed in staminate flower tissue proceeded to test for differential expression only for pre-dormancy staminate catkin buds. We used DESeq2[142] to obtain transcript counts under a generalized linear model that corrects for library size and other RNAseq biases. To test for differential expression of the *g* allele between *gg* homozygotes and *Gg* heterozygotes, we required a separate significance test to account for the difference in copy number. To obtain transcript estimates for a single copy of *g* we divided the normalized counts for *gg* homozygotes by two, then fit a generalized linear model to these counts using a negative binomial distribution model and dichogamy type as a predictor. We used a likelihood ratio test from a comparison of this model to one without dichogamy type included as a predictor.

We lacked biological replicates of mature staminate catkin tissue for protandrous morphs of *P. stenoptera*, as this tissue was sampled on a date when the majority of protandrous individuals had already completed pollen shed. Thus, it is possible that overall levels of *GFAFL* expression may differ between morphs in this tissue at other points during the season.

To test for allele-specific expression in heterozygotes, we used the R function prop.test to test whether the proportion of estimated read counts of each allele differed from 0.5 (two-tailed test). For staminate bud samples, we ran the test for the two G-locus heterozygotes with sufficient transcript counts for the test to have power. The result was significant in both cases. We report the more conservative *P value* in the main text, while also noting that the sample with the highest raw transcript count (values shown in the main text apply a library size correction) had 30 estimated transcript counts for *GFAFL*, all of which were derived from the *G* allele. In the other two heterozygotes with low transcript counts for the *G* copy of *GFAFL*, we still see the same direction of effect, with zero counts of the *g* copy of *GFAFL*. The same procedure was done for the mature staminate catkin tissue, and the result was significant for two out of three heterozygote samples (in the reverse direction).

In the *Cyclocarya* expression data from Qu et al.,[35] we used salmon to quantify allele-specific transcript counts and DESeq2 to obtain counts corrected for variation in library size.

Analysis of the *GFAFL* region in *Cyclocarya*

To investigate the possibility that heterodichogamy in *C. paliurus* may be controlled by variation at the *GFAFL* locus, we examined short read sequencing depth against the diploid protandrous and protogynous assemblies from[42] and our assembly of a diploid protogynous individual. This analysis revealed a segregating copy number variant (CNV) at the location syntenic with the *Pterocarya* G-locus. Depth of long reads from individuals of known type indicates that this CNV shows an association with dichogamy type (Figures S6A and S6C).

We found that the diploid protandrous assembly from Qu et al.[42] possessed an annotated copy of *GFAFL* (*CpaM1st14242*), and an unannotated copy ca. 100 kb away on the same chromosome. The latter was soft-masked in the original assembly, but we verified it encodes a functional copy of *GFAFL* using RNAseq (see below). This copy was not found in the assembly of our protogynous individual from the same study, nor in our assembly from a protogynous individual in this study. We determined on the basis of read depth and heterozygosity patterns from short read samples that this haplotype was likely incorrectly assembled and appears to have stitched adjacently together homologous portions of two divergent haplotypes. To further support this, we remapped the long reads from Qu et al.[42] to this regions and found discontinuities in coverage between the duplicates, with zero reads spanning a 3-gene cluster that appeared to have a tandem duplication in the assembly (Figures S6A and S6B). As this diploid protandrous assembly was generated using a methodology which returns a single collapsed assembly, we concluded that this region was assembled as a chimera from the two divergent haplotypes. We then examined the tetraploid protandrous assembly from Qu et al.,[42] which was generated using a haplotype-specific assembly methodology. We found 4 copies of *GFAFL*, one of which clustered phylogenetically

with the unannotated copy from the diploid protandrous assembly (Figures 4A and S4B). This haplotype in the tetraploid showed no evidence of tandem duplication here, further supporting that *GFAFL* is single copy in the two haplotypes which are ancestral to the divergence of diploid and tetraploid lineages.

We next examined in detail four published RNAseq expression datasets including flower tissue from *C. paliurus*. We found that the samples sequenced in Chen et al.[38] represent a single time point of the samples reported in the larger time series of Qu et al.,[42] using PCA to confirm sample identities. Likewise, we noted that the samples sequenced in Qu et al.[143] appear to be a subset of the data from Qu et al.,[42] so we did not treat this dataset as independent.

RNAseq data sets from Zhang et al.[41] were not associated with flowering type metadata, and it is not specified whether the sampled tissue originated from staminate or pistillate flowers. However these data showed clear expression of both *GFAFL* alleles and confirmed that the expression of *GFAFL* is flower-specific.

Analysis of haplotype structure

We used BLAST[144] to align G-locus regions of different haplotypes to infer haplotype structure. We extracted the coordinates of repeat regions in *G* haplotypes and their homologous sequence in *g* haplotypes from the BLAST alignments, aided by a between-haplotype dotplot showing the aligned regions. The same procedure was done for the *Cyclocarya* haplotypes. These extracted regions were used as BLAST queries to assemblies of *Juglans* and *Carya* to identify orthologous sequence at syntenic regions in other assemblies. As in *Pterocarya* the sequences identified in *Juglans* and *Carya* are located approximately 1 kb downstream of the 3' UTR of the *GFAFL* orthologs. We used muscle[116] to align these sequences and then manually curated these alignments to remove aligned regions where homology was dubious, namely in regions with indels and at the edges of the alignments. A maximum likelihood phylogeny was inferred from this alignment using IQ-tree[117] with 1000 ultrafast bootstrap replicates.

Given the apparent *trans*-acting allele-specific regulation associated with the presence of the *G* haplotype repeats in *Pterocarya* and the *G2* repeats in *Cyclocarya*, we considered the possibility of post-transcriptional silencing, and so tested for possible sequence homology over at least 20 bp between the repeated regions in the dominant haplotypes of both *Pterocarya* and *Cyclocarya* and the transcribed regions of *GFAFL* on both haplotypes using a blastn-short search with lenient alignment parameters to detect homology between short sequences. In *Cylocarya*, we found no compelling matches between the entire hemizygous region and the transcribed region of either *GFAFL* allele.

In *P. stenoptera*, we identified no regions that showed sequence matching between any part of the transcribed region of the dominant haplotype copy of *GFAFL* and the entirety of the *G*-repeat region. We saw that several *G*-repeats showed exact sequence matching over 25-30 bp to a site within the 5' UTR intron of only the recessive copy of *GFAFL*, roughly 770 bp upstream of the start codon. At the same local location within the *GFAFL* intron in *P. macroptera*, we found several *G*-repeats matching the recessive but not dominant *FAFL* intron over 20-21 bp. This sequence motif (TTTCACATAAAATATAATAAATAATTCAC) is not unique to this genomic region, occurring in 5099 copies in *P. stenoptera*, and it is significantly enriched ($p < 2.2e-16$) within both upstream and downstream 2 kb flanking regions (but not 300 bp flanking regions) of genes. However, despite similarly high copy number in genomes of two *Juglans* species, we detected few or no sequences matching the motif in small RNA sequencing libraries derived from the homologous tissue[19] in these species.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses are described in the STAR Methods, method details section, in the main text and Figure/Table legends. Scripts for all analyses are available as detailed in the data and code availability statement.

Current Biology, Volume *35*

Supplemental Information

Distinct haplotypes and reversed dominance at a single-gene balanced polymorphism controlling heterodichogamy in two genera of wingnuts

Jeffrey S. Groh, Gracie Ackerman, Michael Wenzel, and Graham Coop

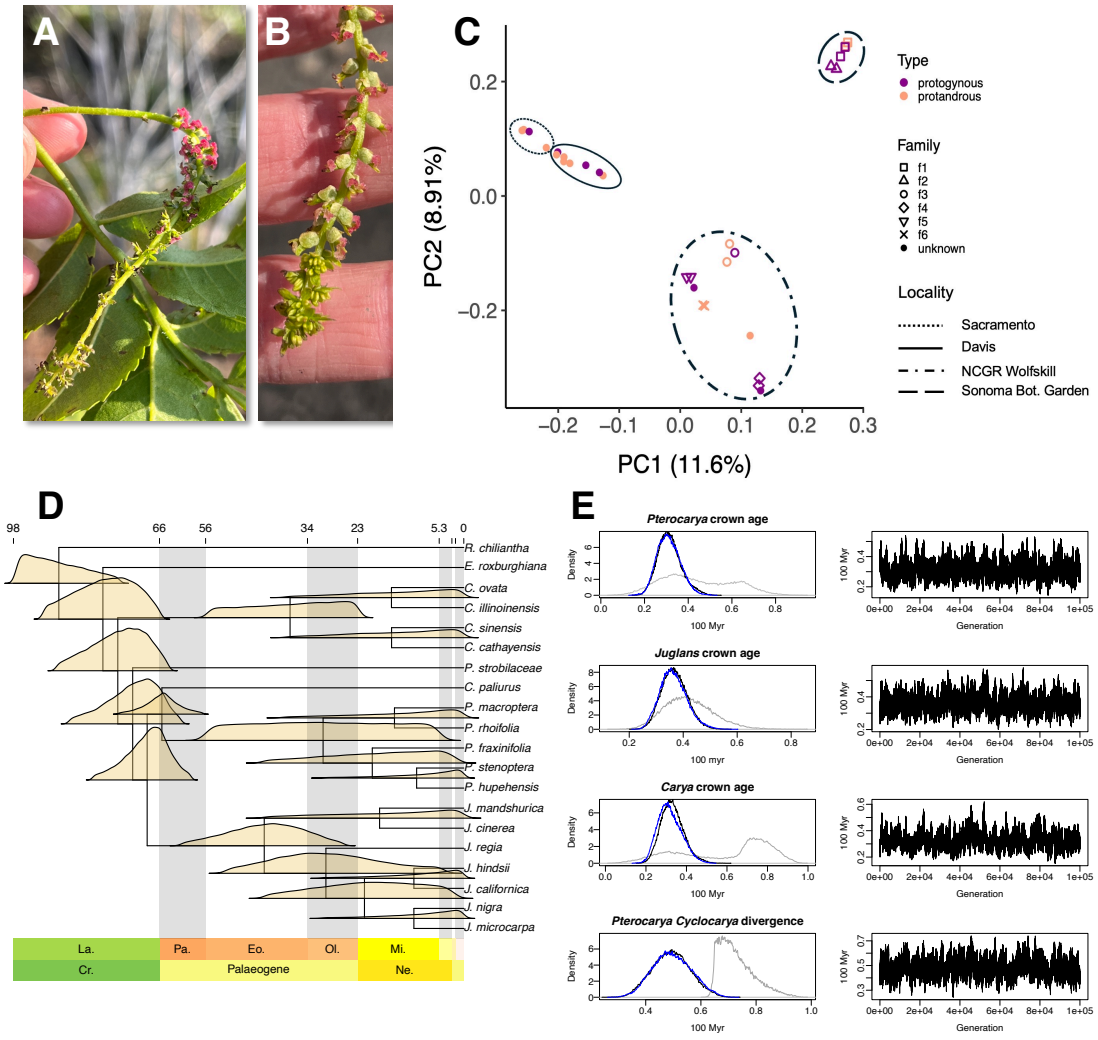


Figure S1. Inflorescences and genetic structure of the *P. stenoptera* mapping population, and Juglandaceae time calibration. Related to Figure 1, Figure 2, and STAR Methods. A,B) Examples of bisexual inflorescences of *Pterocarya stenoptera* uncommonly observed during the course of this study (inflorescences are typically u nisexual). Staminate flowers are borne at the distal end of the rachis while pistillate flowers (visible by pink stigmas) are borne near the proximal end. The inflorescences in (A) and (B) show protandrous and protogynous flowering, respectively. Photos by J. S.G. **C)** Principal Components Analysis (PCA) of genome-wide SNP variation in our mapping population of *P. stenoptera*. Families here are sets of offspring from the same maternal tree. **D)** Visualization of the joint prior distribution on divergence times used in the MCMCTree analysis. Prior distributions were manually specified for a subset of nodes based on fossil information (STAR Methods), and the joint prior on all nodes is subsequently generated by a birth-death model of speciation in MCMCTree. Note that the marginal prior for a given node may deviate slightly from its specified distribution due to constraints of the birth death model that is used to specify the joint prior. **E**) Diagnostics of Bayesian estimation of divergence times in Juglandaceae using MCMCTree. We show the marginal prior density (from sampling, in gray) vs. marginal posterior density (black and blue show two chains run separately) and traces of a single MCMC chain for four nodes of interest. The priors for *Juglans* crown age and *Cyclocarya, Pterocarya* divergence were manually set, while the priors for *Pterocarya* crown age and *Carya* crown age are derived from the birth-death model within MCMCTree. The molecular data appear highly informative for a more recent divergence between *Cyclocarya* and *Pterocarya* than what is reflected in the prior.

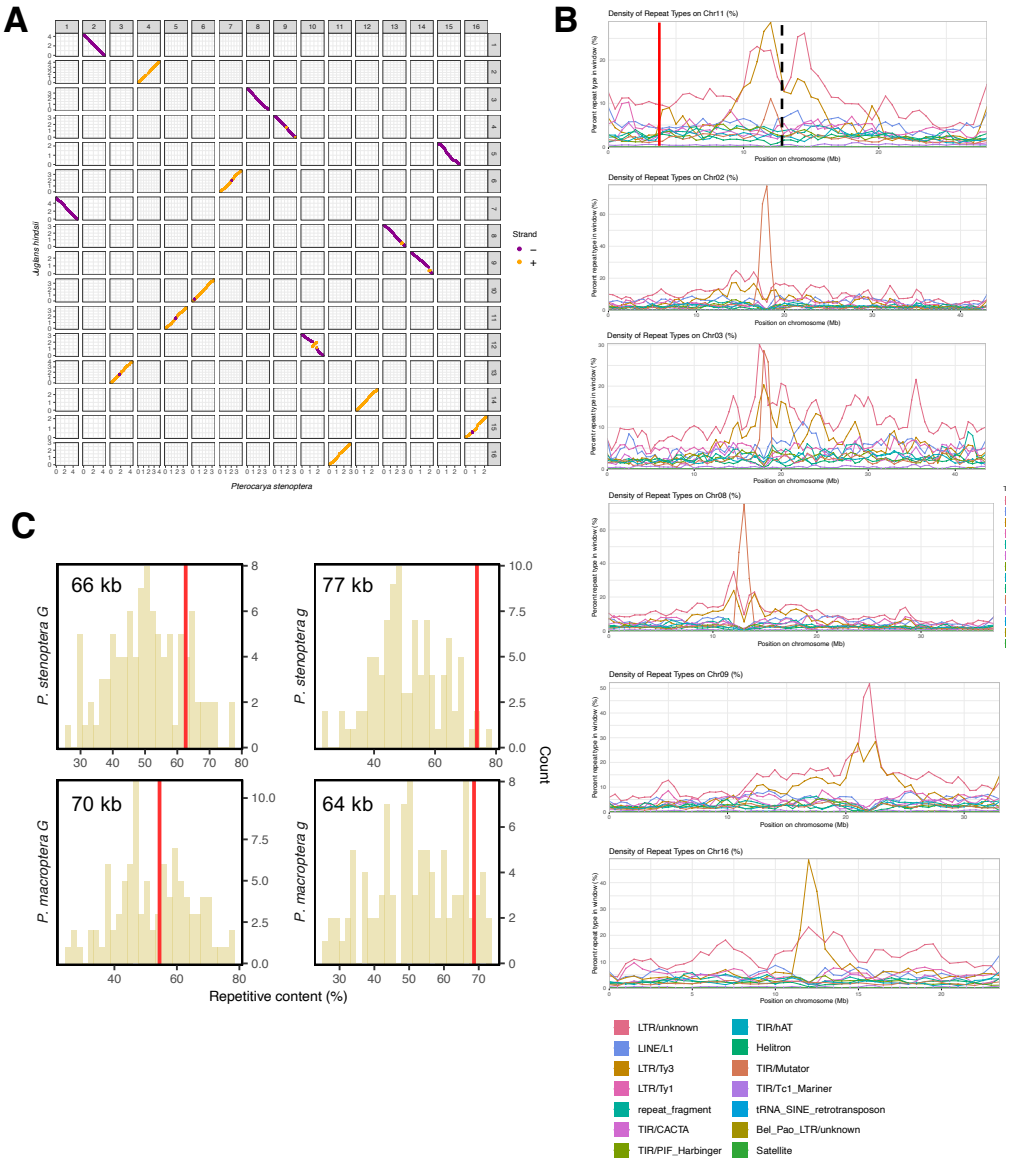

A
Juglans hindsii
Pterocarya stenoptera
Strand
B
Density of Repeat Types on Chr11 (%)
Density of Repeat Types on Chr02 (%)
Density of Repeat Types on Chr03 (%)
Density of Repeat Types on Chr08 (%)
Density of Repeat Types on Chr09 (%)
Density of Repeat Types on Chr16 (%)
Percent repeat type in window (%)
Position on chromosome (Mb)
LTR/unknown
LINE/L1
LTR/Ty3
LTR/Ty1
repeat_fragment
TIR/CACTA
TIR/PIF_Harbinger
TIR/hAT
Helitron
TIR/Mutator
TIR/Tc1_Mariner
tRNA_SINE_retrotransposon
Bel_Pao_LTR/unknown
Satellite
C
66 kb
77 kb
70 kb
64 kb
P. stenoptera G
P. stenoptera g
P. macroptera G
P. macroptera g
Count
Repetitive content (%)

Figure S2. Genome synteny and repeat content. Related to Figure 2. A) Chromosome-level synteny between our denovo assembly of *Pterocarya stenoptera* and an assembly of *Juglans hindsii* from Groh et al.[S1] . *P. stenoptera* was scaffolded by alignment to an assembly from the same species [S2] which incorporated Hi-C data, and chromosome names (shown in dark gray facet labels) match those of the reference used in scaffolding. The *J. hindsii* was reference-scaffolded to the high-quality assembly of *J. regia* which also incorporated Hi-C data, and follows the chromosome naming scheme for that species. The broad-scale 1:1 synteny between chromosomes reflects a known pattern of strong synteny conservation in Juglandaceae (e.g. [S3]). Axis units are 10 Megabases. **B)** Distributions of annotated repetitive sequences (from EDTA) along 6 chromosomes from the *P. stenoptera* haplotype 1 assembly. In this assembly, chromosomes 2, 3, 8, 9 and 16 were assembled as single contigs and all contained telomeric sequence at both ends. Chromosome 11 (top panel) in this assembly was comprised of 2 contigs (contig break shown with black dashed line) and showed telomeric sequence at both ends. The position of the G-locus is indicated by the red line. **C)** Repeat content of G-locus regions (red) compared with 100 neighboring genomic windows of equivalent size.

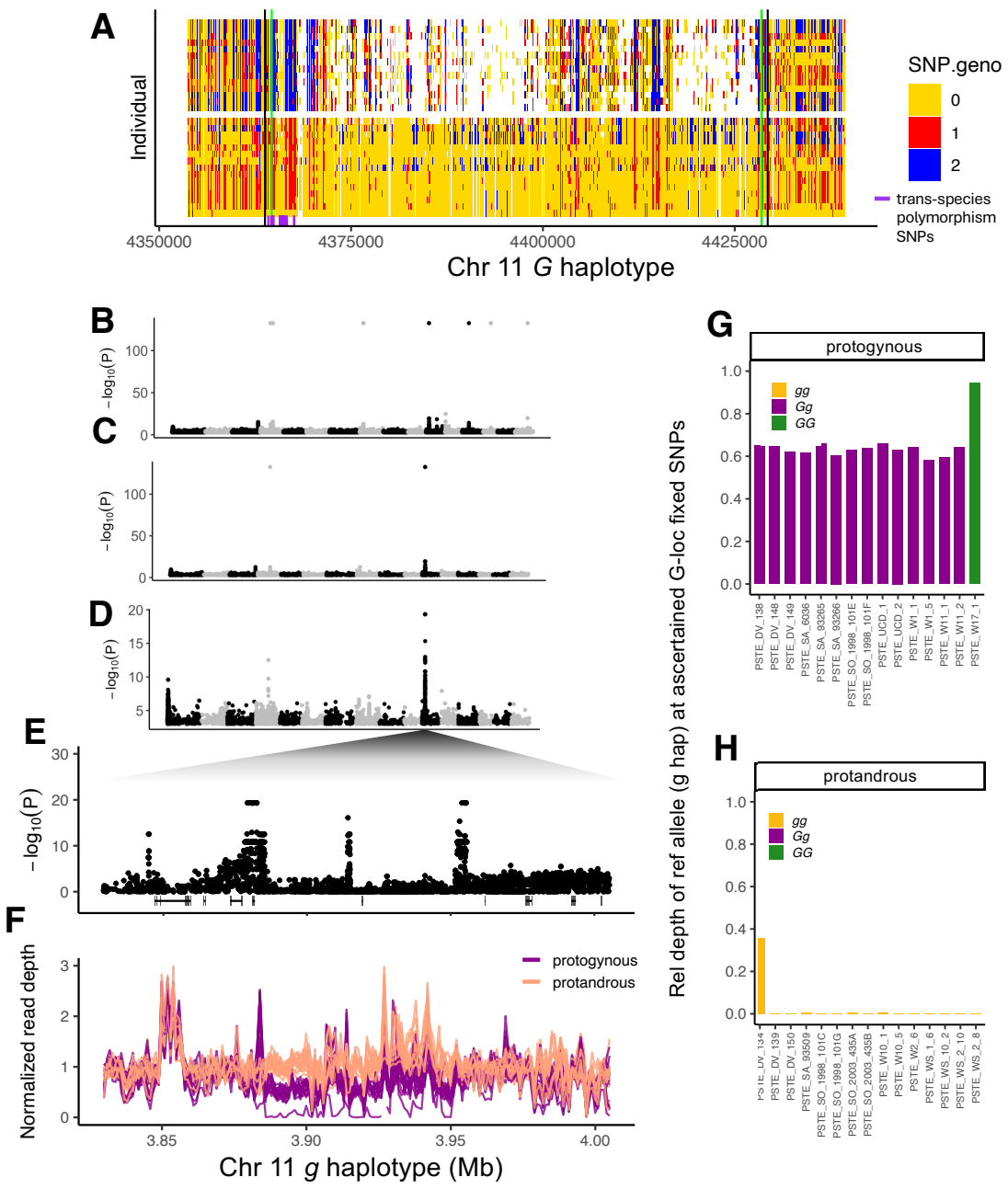

A
Individual
SNP.geno
0
1
2
trans-species polymorphism SNPs
4350000
4375000
4400000
4425000
Chr 11 G haplotype
B
C
D
E
F
$-\log_{10}(P)$
Normalized read depth
protogynous
protandrous
3.85
3.90
3.95
4.00
Chr 11 g haplotype (Mb)
Rel depth of ref allele (g hap) at ascertained G-loc fixed SNPs
G
protogynous
gg
Gg
GG
H
protandrous

Figure S3. Genomic identification of the *G-locus* in *P. stenoptera*. Related to Figure 2. A) SNP genotypes across the G-locus region in the GWAS sample. Each row represents an individual and each column a biallelic SNP. Individuals are grouped by dichogamy type: top - protandrous, bottom - protogynous. Sites are colored according to whether they are homozygous reference (gold), heterozygote (red), or homozygous alternate (blue). The reference assembly here is the dominant haplotype of *P. stenoptera*. Missing genotypes are not filled (white space). Vertical green lines show the outermost positions of the GWAS peak shown in Figure 2B in the main text. Black lines show the region we identified to contain SNPs perfectly associated with dichogamy type. The minor discrepancy between these lines can be explained by our SNP ascertainment criteria allowing for a variable and greater extent of missingness across samples while retaining perfect SNP associations (at least 5 individuals of each type required). We also note that this ascertainment excluded one sample which we identified to be likely contaminated or chimeric (see 'Analysis of Polymorphism and Divergence Within and Between Haplotypes' in STAR Methods) and panels G,H. Purple tick marks at bottom show positions of SNPs where the dominant haplotypes of *P. stenoptera* and *P. macroptera* share one allele, and the recessive haplotypes share the other allele. The individual in the lowermost row we identified as the only dominant allele homozygote in our GWAS sample. **B)** GWAS for dichogamy type using variants called against haplotype 1 of *P. stenoptera* (recessive). The GWAS was considerably noisier in this case with many spurious peaks (vs. variant calls against haplotype 2, see Figure 2). This pattern is not expected for a binary trait, as these peaks all show perfect SNP associations across many different chromosomes that segregate independently. Furthermore, these strongly associated sites show P values with considerably stronger associations than what we saw for variants called against haplotype 2. Examining the set of SNPs with the strongest associations, we identified that they all showed zero missing data, and an allele frequency of exactly 0.25 (our GWAS sample was 50/50 of each type). Given that our coverage analysis indicated that at individual was homozygote for the dominant allele, this result was unexpected. Many of these peaks occurred at isolated SNPs with little association signal at surrounding loci, as would be expected for a true causal locus in linkage disequilibrium with closely physically linked alleles. We therefore hypothesized these patterns could be the result of reads derived from portions of the associated region in haplotype 2 that do not have homologous sequence at the syntenic position in haplotype 1. **C)** Consistent with this, after filtering out any positions within regions annotated as repetitive sequence by either RepeatMasker or EDTA, many of these strong associations disappear, leaving only two on chromosome 4 and chromosome 11. Furthermore, we extracted the local region within haplotype 1 around the strongly associated SNP on chromosome 4, and used this as a BLAST query against haplotype 2. This sequence showed a strong match to the G-locus region on chromosome 11 of haplotype 2, but we did not see a match to this region when BLASTing this query sequence against haplotype 1. **D)** Here, the y-axis is truncated, hiding the most extreme association values on chromosome 4 and chromosome 11 seen in the above panel. **E)** Zoomed in view of the GWAS peak with gene annotations at bottom. We found no evidence of expression for the predicted gene in the center of the associated region and considered it an annotation error (STAR Methods). **F)** Normalized average read depth in 1 kb windows for 30 phenotyped individuals. The individual with near zero read depth in the associated region was also identified as a *GG* homozygote on the basis of SNP genotypes and read depth against the *G* haplotype. **G,H)** Allelic depths at a set of 255 SNP sites with fixed differences between G-locus haplotypes in *P. stenoptera*. One protandrous sample (DV_134) was excluded in the ascertainment. We used this to investigate the possibility of this sample being chimeric as the basis for exclusion in analyses involving haplotype phasing. This sample showed non-zero depth of the reference (*G*) allele at many of these sites. Allele-specific read depths suggests this individual is not a heterozygote but rather was affected by sample contamination or the original sample contained material from two cell populations with different G-locus genotypes. Given the deep divergence between haplotypes, heterozygotes should show a bias toward expression of the reference allele (here, *G*), which we do see for heterozygous protogynous samples (purple). In contrast, DV_134 shows the reverse bias in allelic depth, which is not consistent with this individual being a heterozygote. This individual was confidently phenotyped as male-first so we consider this pattern to reflect either a contaminated or chimeric sample and excluded it from a subset of analyses.

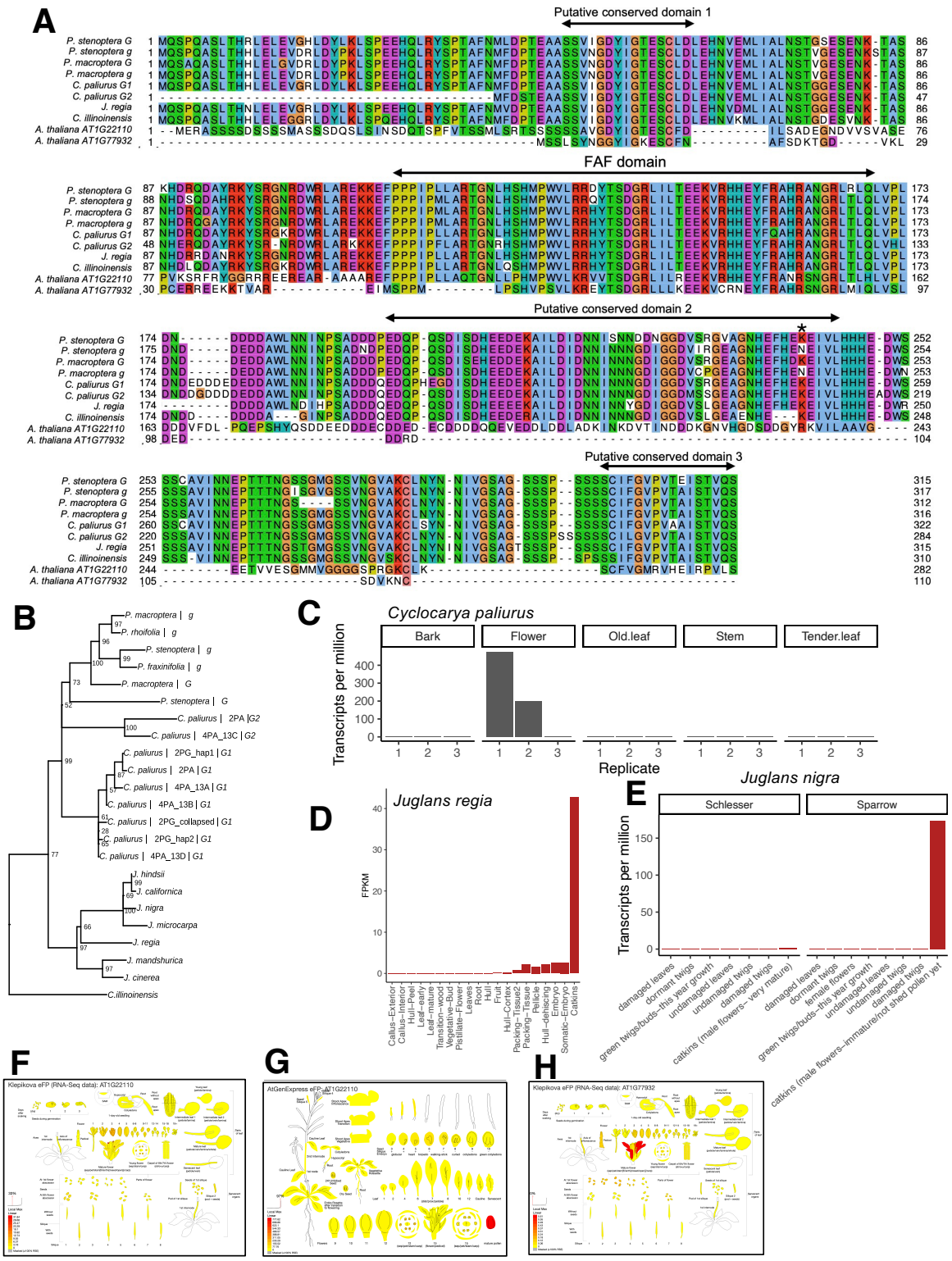

A
Putative conserved domain 1
FAF domain
Putative conserved domain 2
Putative conserved domain 3
B
C
Cyclocarya paliurus
Bark
Flower
Old.leaf
Stem
Tender.leaf
Transcripts per million
Replicate
D
Juglans regia
FPKM
E
Juglans nigra
Schlesser
Sparrow
Transcripts per million
F
G
H

Figure S4. Alignment, phylogeny, and expression of *GFAFL* orthologs. Related to Figure 3. A) Amino acid alignment of G-LOCUS-FAF-LIKE orthologs from *Pterocarya Cyclocarya* and outgroups. We highlight four regions showing amino acid sequence conservation to the closest ortholog in Arabidopsis, one of which is the FAF domain. An asterisk marks the location of a nonconservative substitution within a conserved region which is the only nonsynonymous polymorphism between G-locus alleles shared across *Pterocarya*. At each site, amino acids are colored according to the conservation of amino acid properties in the sample acid properties using the default coloring scheme for Clustal X. **B)** Maximum likelihood phylogeny of *GFAFL* orthologs in Juglandoideae constructed from an alignment of the coding sequence region. **C)** Expression of the annotated ortholog of *GFAFL* in *Cyclocarya paliurus*. Raw data from Zhang et al.[S4]. The original reference does not indicate whether the flower samples originated from pistillate or staminate flowers. **d)** Expression of the *GFAFL* ortholog across 20 tissues sampled from *Juglans regia*. Here, 'catkins' refers to the staminate inflorescences. Transcript quantifications were made available by Houston Saxe (UC Davis) (raw data from Chakraborty et al.[S5]). Expression is measured as Fragments Per Kilobase of transcript per Million mapped reads. These data are from a single individual ('Chandler') which is protandrous. **E)** Expression of *GFAFL* across multiple tissues from two individuals of *J. nigra*. Both individuals are protogynous heterozygotes. Raw data are from Hardwood Genomics Team [S6]. **F,G)** Expression atlas for the Arabidopsis gene *AT1G22110*, which is the closest Arabidopsis homolog of *GFAFL* (see panel A). Data are from Klepikova et al.[S7] (F) and Schmid et al.[S8] (G). **H)** Expression atlas for the Arabidopsis gene *AT1G77932*, a more distantly related homolog. Data are from Klepikova et al.[S7]. Results for (F-H) were visualized at https://bar.utoronto.ca/eplant/; details of the quantification methods used are found at the provided link and in the original papers.

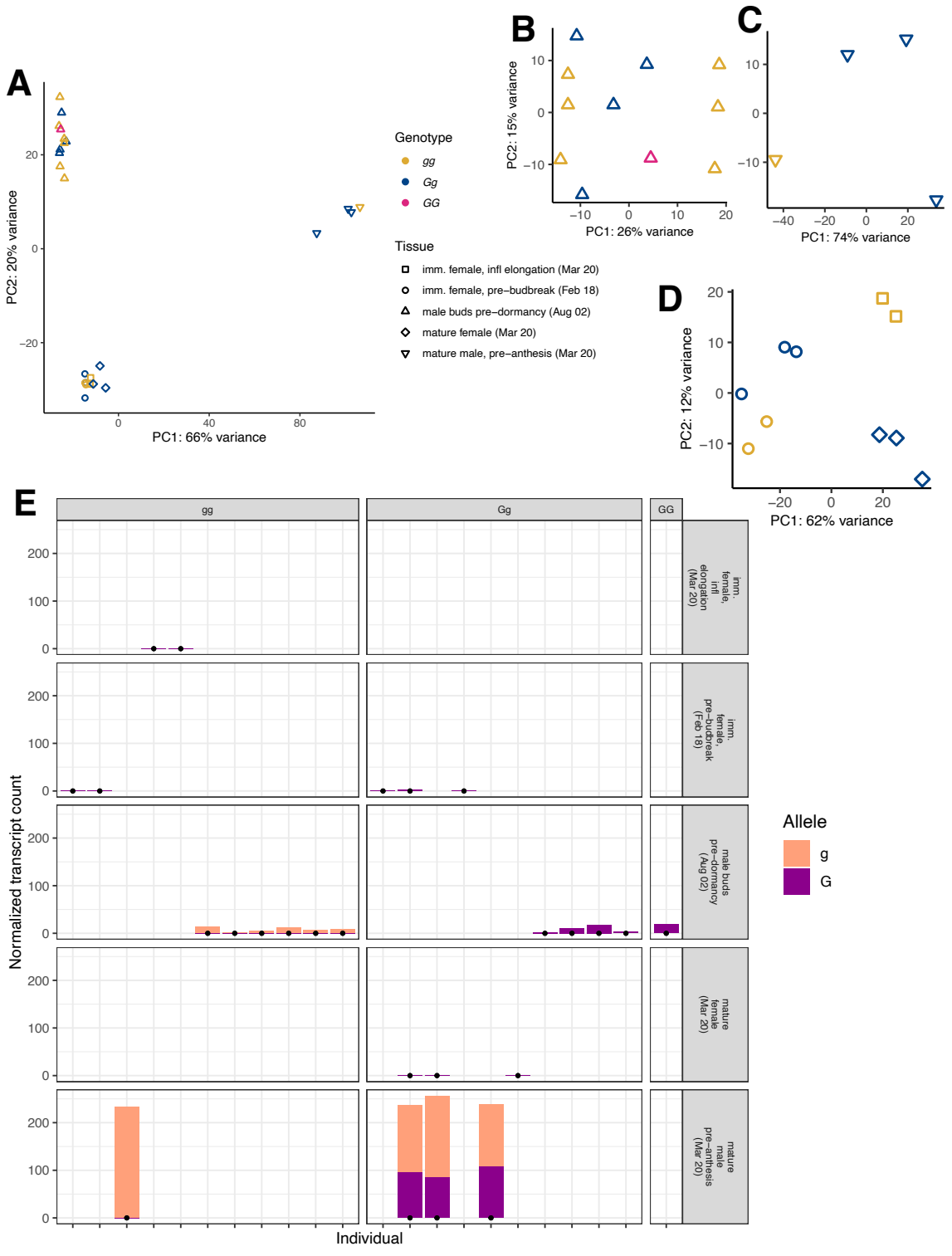

A
PC2: 20% variance
PC1: 66% variance
B
PC2: 15% variance
PC1: 26% variance
C
PC1: 74% variance
D
PC2: 12% variance
PC1: 62% variance
Genotype
gg
Gg
GG
Tissue
imm. female, infl elongation (Mar 20)
imm. female, pre−budbreak (Feb 18)
male buds pre−dormancy (Aug 02)
mature female (Mar 20)
mature male, pre−anthesis (Mar 20)
E
gg
Gg
GG
Normalized transcript count
Individual
Allele
g
G

Figure S5. Analyses of RNA-seq data from *P. stenoptera* flowers. Related to Figure 3. A) Principal Components Analysis (PCA) biplots of variance-stabilized transcript counts for biological replicates across three floral tissue types of *P. stenoptera*. Within each tissue type, each point represents a different individual; some individuals were sampled for multiple tissue types. Samples cluster group by tissue type. Samples of pistillate (female) catkins and mature staminate (male) catkins were sampled on the same day during the middle of the flowering season (March 20th, 2024), while samples of pre-dormant male catkin buds were sampled on August 2nd of the same year at the end of the growth season. **B-D)** Separate PCA plots for each tissue run separately, shapes same as in (A). There is some clustering by morph for the two floral tissues collected during flowering, particularly in (D), where samples also group by sampling date. **E)** Expression of *GFAFL* transcripts across all *Pterocarya stenoptera* flower samples in this study. See STAR Methods for details of allele-specific transcript count estimation. Note that each tissue is represented by a unique set of samples, although some individuals were sampled for multiple tissues. Each column on the x-axis represents a distinct individual. The sampled individuals for each tissue are indicated with black dots at bottom. Across all female flower tissue samples from multiple developmental stages, only one sample showed a single non-duplicate read pair aligning to the *GFAFL* coding region.

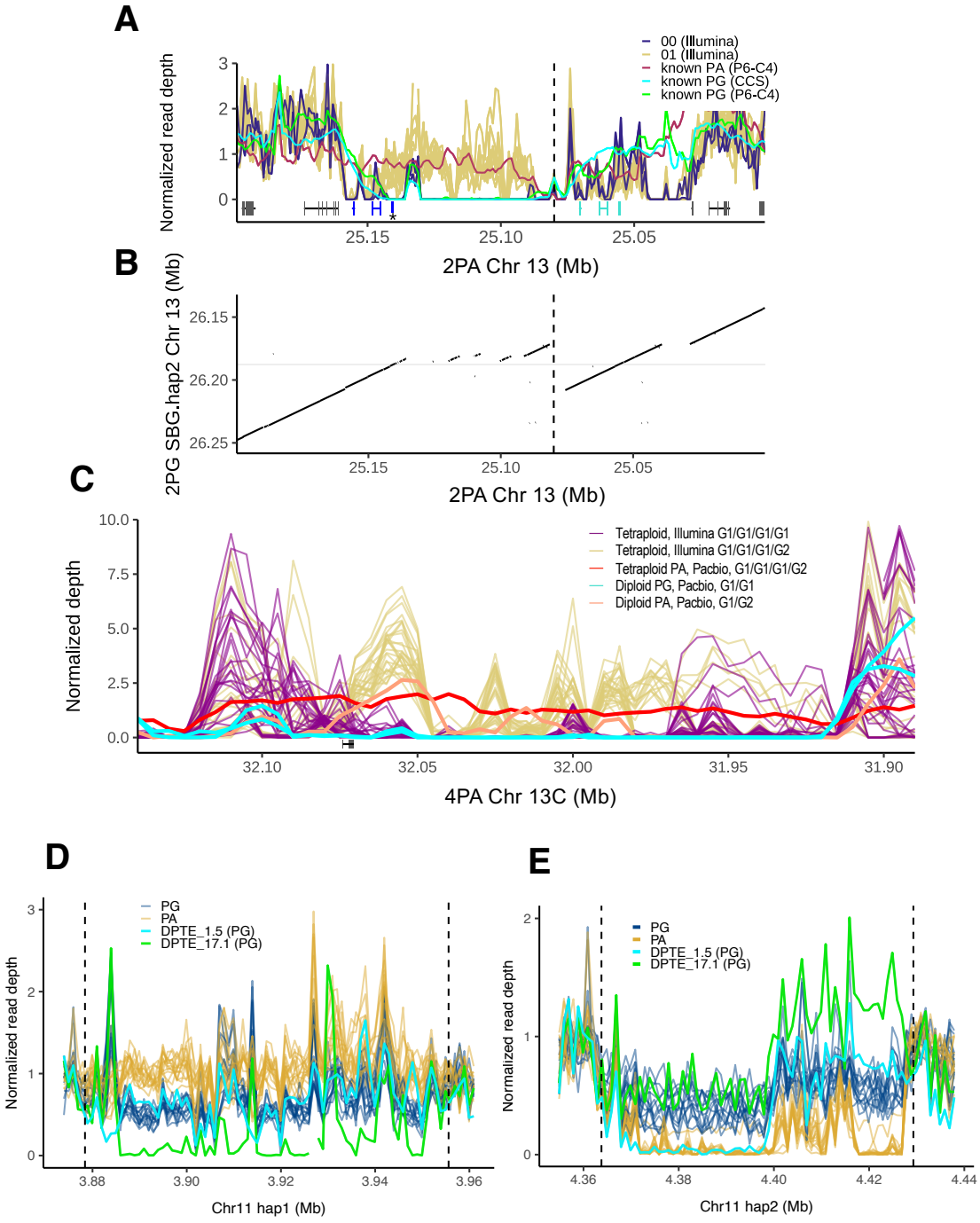

A
Normalized read depth
00 (Illumina)
01 (Illumina)
known PA (P6-C4)
known PG (CCS)
known PG (P6-C4)
2PA Chr 13 (Mb)
B
2PG SBG.hap2 Chr 13 (Mb)
2PA Chr 13 (Mb)
C
Normalized depth
Tetraploid, Illumina G1/G1/G1/G1
Tetraploid, Illumina G1/G1/G1/G2
Tetraploid PA, Pacbio, G1/G1/G1/G2
Diploid PG, Pacbio, G1/G1
Diploid PA, Pacbio, G1/G2
4PA Chr 13C (Mb)
D
Normalized read depth
PG
PA
DPTE_1.5 (PG)
DPTE_17.1 (PG)
Chr11 hap1 (Mb)
E
Normalized read depth
PG
PA
DPTE_1.5 (PG)
DPTE_17.1 (PG)
Chr11 hap2 (Mb)

Figure S6. Coverage analyses of the G-locus in *Cyclocarya* and *Pterocarya*. Related to Figures 2, 4, and 5. **A)** Read depth from 13 diploid *Cyclocarya paliurus* against the diploid collapsed assembly from a protandrous *C. paliurus* individual from Qu et al.[S9]. Ten individuals (Illumina samples, dark blue and gold lines) are of unknown type, and three individuals (PacBio samples, light blue, green and maroon lines) are of known type. This assembly appears to contain two tandem duplicated blocks of 3 genes each (turquoise and blue rectangles at bottom), including an unannotated duplicate of *GFAFL* (blue, asterisk). However, we determined that the assembly in this region is chimeric, with two heterozygous haplotypes placed adjacent to another, and these duplicate blocks are in fact orthologous single-copy alleles. The dashed vertical line indicates the inferred location of a contig break; no long reads from the individual used for assembly (red line) span this junction. **B)** Dotplot alignment between the protandrous diploid assembly (x-axis) and one haplotype of our protogynous diploid assembly (y-axis). Horizontal gray line indicates the position on the y-axis of *GFAFL*. **C)** Read depth of *C. paliurus* individuals over the *G2* haplotype from a haplotype-resolved assembly from a protandrous tetraploid individual of the same species. The short read samples in tan and purple (unknown phenotypes) correspond exactly to samples that fall into the two tetraploid genotypic clusters shown in Figure 4B, indicating that this copy number variant co-segregates with the *G2* copy of *GFAFL*. We also show sequencing depth from long-read Pacbio sequencing of four individuals of known type. The red line shows depth for the same individual used for assembly[S9]. Blue lines correspond to our PacBio HiFi assembly of a diploid protogynous individual and the assembly reads for the diploid protogynous assembly from Qu et al.[S9], the salmon line shows the protandrous diploid assembly reads from the same study. For each color, we indicate the inferred G-locus genotype. **D,E)** Read depth patterns of protogynous individuals suggest structural diversity of *G* haplotypes in *P. stenoptera*. The individual DPTE_1.5 lacks depth across a portion of the *G* haplotype in whole genome-resequence data, suggesting it may possess a *G* haplotype that is structurally distinct from others in the sample. The sample DPTE_17.1, which we determined to be a *GG* homozygote, shows a partial reduction in read depth across the same region, within the range of read depth seen in heterozygotes. This may indicate that one of its two *G* haplotypes lacks this portion, however it could simply reflect poor read mapping to this region. We additionally note that one protandrous sample shows non-zero read depth across this region of the *G* haplotype. We show elsewhere that this sample shows evidence of contamination or chimerism on the basis of skewed allele depth ratios across this region (see methods and Figure S3G,H).

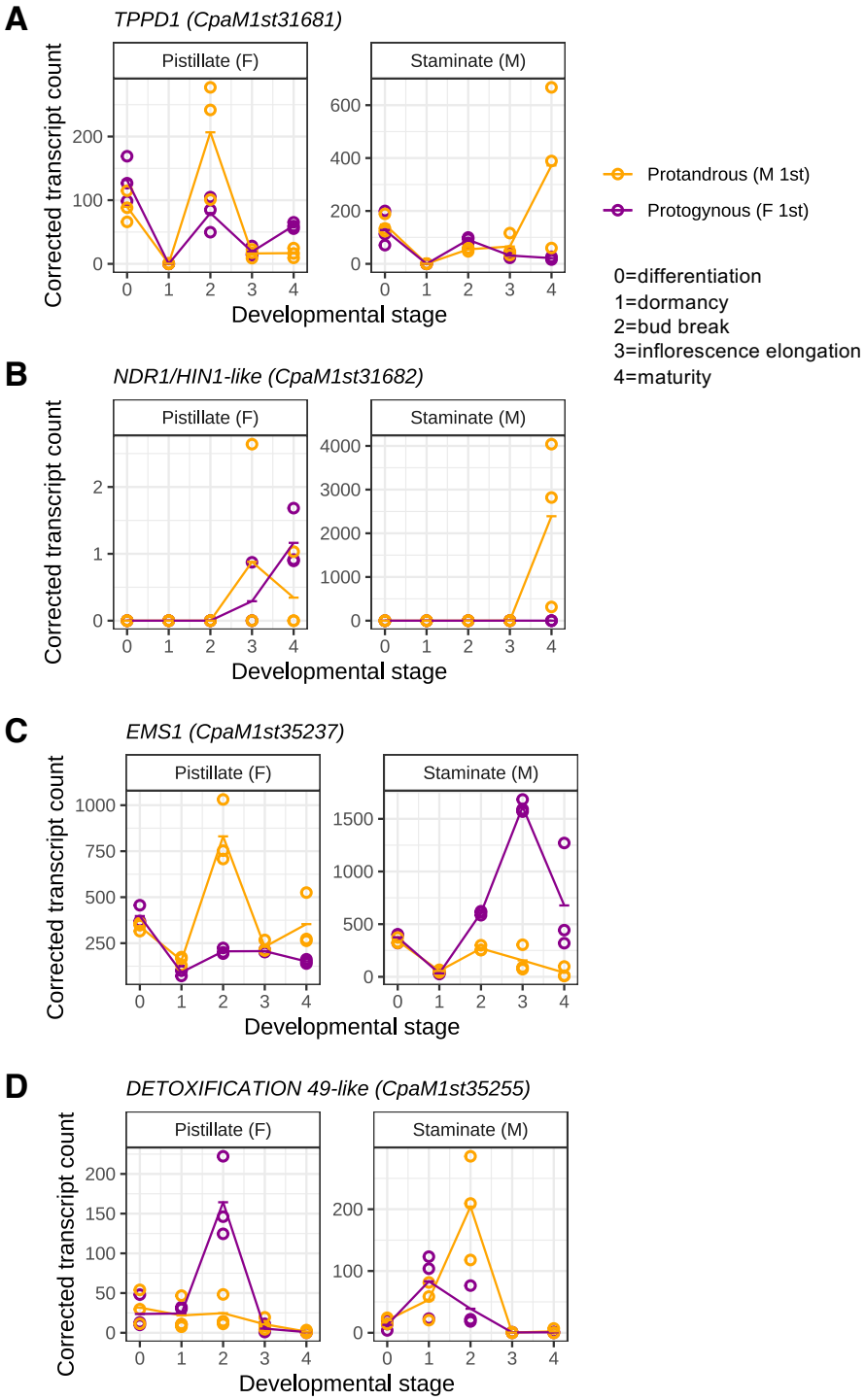

A
TPPD1 (CpaM1st31681)
Pistillate (F)
Staminate (M)
Corrected transcript count
Developmental stage
B
NDR1/HIN1-like (CpaM1st31682)
C
EMS1 (CpaM1st35237)
D
DETOXIFICATION 49-like (CpaM1st35255)
Protandrous (M 1st)
Protogynous (F 1st)
0=differentiation
1=dormancy
2=bud break
3=inflorescence elongation
4=maturity

Figure S7. Temporal expression of *Cyclocarya* orthologs of candidate genes found in other genera. Related to Figure 4. **A,B)** Expression of *Juglans* G-locus genes across a developmental time series dataset of gene expression from protandrous and protogynous morphs of *Cyclocarya paliurus* [S10]. Note that the x-axis shows developmental stage rather than time, so some difference between morphs could reflect either sampling noise or expression changes that depend on time in the season rather than developmental stage. **C,D)** Expression of two candidate genes from the *Carya* G-locus across a developmental time series dataset of gene expression from protandrous and protogynous morphs of *Cyclocarya paliurus* [S10]. These two genes show complementary patterns of expression between staminate flowers of one morph and pistillate flowers of the other morph. These data suggest that, in *Cyclocarya*, *EMS1* could be negatively regulating flowering, while *DETOXIFICATION 49-like* could be positively regulating flowering.

Supplementary References

S1 Groh, J.S., Vik, D.C., Davis, M., Monroe, J.G., Stevens, K.A., Brown, P.J., Langley, C.H., and Coop, G. (2025). Ancient structural variants control sex-specific flowering time morphs in walnuts and hickories. Science *387*, eado5578.

S2 Zhang, Z.Y., Xia, H.X., Yuan, M.J., Gao, F., Bao, W.H., Jin, L., Li, M., and Li, Y. (2024). Multi-omics analyses provide insights into the evolutionary history and the synthesis of medicinal components of the Chinese wingnut. Plant Diversity *46*, 309–320.

S3 Lovell, J.T., Bentley, N.B., Bhattarai, G., Jenkins, J.W., Sreedasyam, A., Alarcon, Y., Bock, C., Boston, L.B., Carlson, J., Cervantes, K. et al. (2021). Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding. Nature Communications *12*, 4125.

S4 Zhang, S.Y., Peng, Y.Q., Xiang, G.S., Song, W.L., Feng, L., Jiang, X.Y., Li, X.J., He, S.M., Yang, S.C., Zhao, Y. et al. (2024). Functional characterization of genes related to triterpene and flavonoid biosynthesis in *Cyclocarya paliurus*. Planta *259*, 50.

S5 Chakraborty, S., Britton, M., Martínez-García, P., and Dandekar, A.M. (2016). Deep RNA-Seq profile reveals biodiversity, plant–microbe interactions and a large family of NBS-LRR resistance genes in walnut (*Juglans regia*) tissues. AMB Express *6*, 1–13.

S6 Hardwood Genomics Team (2015). Comparative genomics of environmental stress responses in North American hardwoods. . https://www.ncbi.nlm.nih.gov/bioproject/273267.

S7 Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., and Penin, A.A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. The Plant Journal *88*, 1058–1070.

S8 Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nature Genetics *37*, 501–506.

S9 Qu, Y., Shang, X., Fang, S., Zhang, X., and Fu, X. (2023). Genome assembly of two diploid and one auto-tetraploid *Cyclocarya paliurus* genomes. Scientific Data *10*, 507.

S10 Qu, Y., Shang, X., Zeng, Z., Yu, Y., Bian, G., Wang, W., Liu, L., Tian, L., Zhang, S., Wang, Q. et al. (2023). Whole-genome duplication reshaped adaptive evolution in a relict plant species, *Cyclocarya paliurus*. Genomics, Proteomics & Bioinformatics *21*, 455–469.