

Supplemental information

“An updated perspective: What genes make a tree a tree?”

Siri Birkeland¹, Eduardo R. Soldado², Sonali S. Ranade³, M. Rosario G. Gil³, Shruti Choudhary³, Hannele Tuominen³, Ewa J. Mellerowicz³, Nathaniel R. Street^{2,4}, Torgeir R. Hvidsten¹

¹Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

²Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden

³Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden

⁴SciLifeLab, Umeå University, Umeå, Sweden

Correspondence: siri.birkeland@nhm.uio.no, torgeir.r.hvidsten@nmbu.no

Supplemental material and methods.....	2
Supplemental results	5
Overview of supplemental figures.....	6
Overview of supplemental tables	6
Overview of supplemental materials	9
Supplemental figures	10
Supplemental tables	18
References	29

Supplemental material and methods

Taxon Sampling, Transcriptome generation, and Final Data Set

We gathered coding sequences from woody and non-woody species within the Rosids, focusing on six major Fabid orders: Fabales, Cucurbitales, Fagales, Rosales, Malpighiales, and Oxalidales (Figure 3, Supplementary Table S1). In addition, we included two non-Fabid Rosid species (*Arabidopsis thaliana* and *Vitis vinifera*), one basal Eudicot species (*Macadamia integrifolia*), and one Magnoliid species (*Drimys angustifolia*) for comparison. Although there exists a continuum of woodiness from herbaceous annual to woody perennial, we chose to include only species that were herbaceous annuals or woody perennials to maximize the chances of picking up shared signals. The final data set consisted of coding sequences from 36 species. While a detailed phylogenetic justification for the sampling strategy is provided in the alignment generation section, our aim here was to maximize comparative power by including lineages with independent transitions in growth form. Data for most of the 36 species stemmed from already published genomes, but a iso-seq transcriptome was generated *de novo* for *Drimys angustifolia*. Square-shaped sections of tree trunks (from 12 individual trees), spanning from the bark to the xylem, were excised in the field in Brazil and immediately placed on dry ice for preservation. In the laboratory, total RNA was extracted from wood scrapings (mainly phloem, cambium and expanding xylem) using a protocol adapted from Chang et al. [S9] and the Qiagen RNeasy Kit, with modifications by Jarmo Schrader in 2002. RNA was quantified using Qubit RNA High sensitivity (Life Technologies, Carlsbad, California, USA), while RNA quality and integrity were assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) using RNA Pico chips. The five best RNA samples in terms of quality and quantity were pooled and sent to Novogene for SMRTbell library preparation, followed by HiFi long read sequencing on a PacBio Sequel II/Ile system (Pacific Biosciences, Menlo Park, California, USA). The resulting HiFi reads were clustered using the IsoSeq v3 workflow (<https://github.com/PacificBiosciences/IsoSeq/blob/master/isoseq-clustering.md>) and run through TransDecoder v.5.5.0 (<https://github.com/TransDecoder/TransDecoder>) using default settings. The completeness of the transcriptome was evaluated by estimating numbers of benchmarking universal single-copy orthologues (BUSCOs) using BUSCO version 5.4.3 and the embryophyta database [odb10; S10].

Ancestral State Reconstruction

As a basis for downstream comparative analyses we performed a simple ancestral state reconstruction of woodiness using the ace function (Ancestral Character Estimation) in the R package ape v5.6-2 [Analyses of Phylogenetics and Evolution; S1,S2], together with the R package phytools v1.0-3 [S3]. Using ace, we made a maximum likelihood reconstruction using discrete characters (woody/herbaceous), and equal rate of transition. Both an unequal rate of transition, as well as stepwise transitions could also be plausible [see e.g., S11], but the main point was to generate a hypothesis for which branches to treat as woody/herbaceous in downstream analyses. The input tree was obtained from already published phylogenies using <http://www.timetree.org/> [Figure S1; S4].

Alignment generation for downstream selection analyses

To reduce potential phylogenetic biases in our comparative analysis, we focused on closely related orders where transitions between woody and herbaceous growth forms have independently occurred. Specifically, we generated gene alignments for selection analyzes separately for four sets of orders: (i) Fabales, (ii) Rosales, (iii) Fagales and Cucurbitales, (iv) Malpighiales and Oxalidales (Fig 1). Fagales and Cucurbitales were grouped because they are closely related and represent contrasting growth forms - Fagales comprises exclusively woody species, while Cucurbitales includes only herbaceous ones. Oxalidales, which had only one representative genome, was combined with its close relative Malpighiales to ensure sufficient sampling. Including Malpighiales and centering sampling around this order also felt appropriate, as it contains the well-studied model tree genus *Populus*, which has been a reference point in many studies of tree biology. We prioritized clades with a relatively high number of available genome assemblies and aimed to include both woody perennials and herbaceous annuals within each set. Although herbaceous annuals are generally underrepresented in genomic resources compared to woody perennials within the Rosids, we strove to balance the representation of growth forms within each group by occasionally excluding genomes of woody species. Additionally, we chose to perform gene alignments within each order or order pair, rather than across the full data set, to reduce alignment errors associated with deep sequence divergence and missing data. For each order (or order pair) we ran Orthofinder v.2.5.4 [S8] with default settings, and the resulting orthogroups were aligned based on protein sequences using MAFFT v7.427 [S12] and the --auto option. Only orthogroups that included gene copies from all species were processed for downstream analyses. As there were very few single-copy orthogroups, we split multi-copy orthogroups into orthogroup subsets based on Kimura protein distances [S13] following the approach of Birkeland et al. [S14]. Briefly, we chose one woody species in each order (or order combination; i) *Dalbergia odorifera*, ii) *Prunus avium*, iii) *Betula pendula*, and iv) *Populus tremula*), and generated orthogroup subsets by picking one gene from the remaining species based on the smallest protein distance to each of its gene copies. Protein distances were calculated using the distmat algorithm in EMBOSS v6.6.0 [S15]. The new orthogroup subsets included only one gene copy from each species, and were realigned based on nucleotide sequence using PRANK v.170427 [S16] with ten bootstraps within the alignment filtering program GUIDANCE2 v2.02 [S17]. We also filtered out all columns scoring <0.80. The combination of a phylogeny aware aligner in combination with alignment filtering has been demonstrated to enhance the accuracy of (positive) selection detection, particularly by reducing false positives and alignment-related biases [S18-S19].

Testing for relaxed selection to identify potential “tree genes”

We hypothesized that genes experiencing convergent shifts in selection pressures in transitions from woody to herbaceous growth forms might be tied to wood development. We therefore ran the RELAX algorithm [S5] implemented in HyPhy v2.5.38 to detect whether a gene had undergone relaxation or intensification of selective strength in herbaceous compared to woody taxa (see Figure 2). RELAX is based on the branch-site test and requires a codon alignment and a predefined species or gene tree to infer a selection intensity parameter (k) for a set of test branches (here the herbaceous taxa) compared to a set of reference branches (the woody taxa). The test compares a null model where k is constrained to 1, with an alternative model where k is a free parameter. If the latter leads to a significant improvement in model fit, it is concluded that selection has been relaxed ($k <$

1) or intensified ($k > 1$) using a likelihood ratio test (LRT; $p \leq 0.05$). Note that relaxed selection in this case includes relaxation of both positive and purifying selection. Ancestral state reconstruction is incorporated as a part of the algorithm's inference framework. We ran RELAX for each of the four sets of gene alignments (Fabales (i), Rosales (ii), Fagales and Cucurbitales (iii), and Malpighiales and Oxalidales (iv)), and compared the genes that were found to be under relaxed selection based on orthology. Since we had thousands of alignments, we chose to run RELAX with predefined species trees (generated with timetree.org; Figures S2-S9) instead of individual gene trees for simplicity. The overlaps in orthology were visualized through an UpSet plot using UpSetR version 1.4.0 [S20]. We also tested the statistical significance of the intersections with the supertest function in the SuperExactTest package v1.1.0 [S21]. In addition to generating robust alignments and testing the significance of gene set intersections, we implemented additional measures to enhance the reliability of our RELAX analyses and reduce the risk of false positives. Specifically, we focused on genes that were consistently recovered across multiple independent analyses and evolutionary transitions, increasing confidence that the observed patterns reflect robust, recurrent evolutionary signals rather than artifacts of individual comparisons. Second, we prioritized genes with functional annotations related to woodiness, growth regulation, meristem function, or plant development, which supports their biological relevance and reduces the likelihood of spurious hits.

Gene function enrichment analyses on shared gene sets

To functionally characterize gene sets that were found to be under relaxed selection in more than one transition to herbaceousness, we performed Gene Ontology (GO) enrichment analyses on intersecting gene sets using topGO v2.52.0. [S23,S24]. Enrichment analyses were performed within the Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) domains using a Fisher's exact test in combination with the elim algorithm [recommended by the developers due to its simplicity; S23]. All analyses were performed based on gene copies from the model tree *P. tremula*, except when orthogroup intersections did not include Malpighiales and Oxalidales. In those cases, we used the gene copies of *B. pendula* (Fagales) or *P. avium* (Rosales) instead. The entire annotated gene sets of *P. tremula*, *B. pendula* or *P. avium* were used as backgrounds in the respective analyses, and a GO-term was considered significant when the p-value of the Fisher's exact test was ≤ 0.05 . We did not correct for multiple testing since GO-term enrichment tests are not independent.

Investigating gene family evolution in terms of copy number variation

We hypothesized that gene families which consistently had higher copy numbers in trees compared to herbaceous species to be tied to wood formation. To detect significant differences in copy numbers we used both phylogeny-independent and phylogeny-informed approaches. i) Phylogeny-independent: First we performed a statistical test on each individual family (negative binomial distribution as implemented in the R package DESeq2) and retained the families with a false discovery corrected p-value < 0.1 . Second we used a random forest classifier to predict growth form from copy numbers (R package *randomForest* with parameters *importance=TRUE*, *proximity=TRUE*, *ntree=1000*). Here the importance score (*MeanDecreaseGini*) of each family reflects to what degree that family, possibly in combination with other families, was important for the classifiers ability to predict growth form. P-values were calculated by comparing the importance

scores to those obtained from running the classifiers with randomly shuffled growth form labels, and families with a false discovery corrected p-value < 0.1 were retained.

Phylogeny-informed: We ran CAFE v.5.0.0 [S7] to reconstruct copy number variation and detect significant retractions/expansions of gene families across a species tree [generated with <http://www.timetree.org/>; S4]. CAFE identifies significant expansions or contractions ($p < 0.01$) by modeling gene family size evolution as a birth-death process and comparing observed family sizes in extant species to expectations under this model, which includes reconstructed ancestral gene copy numbers at internal nodes and accounts for variation in evolutionary rates among families [S7]. The input file to CAFE was made with Orthofinder v.2.5.4 [S8] and included species specific gene copy counts per orthogroup (here treated as gene families). Gene families with large copy number variation (i.e., where one or more species had ≥ 100 gene copies) were filtered out before estimating a single birth-death parameter (λ). A poisson distribution was used for the root frequency distribution (-p), and a model accounting for genome assembly error was included in the final run (-e). The final birth-death parameter (λ) of the species tree was estimated to be 0.005, with a -lnL of 486464. All resulting gene family expansions and contractions were visualized on the species tree using CafePlotter v0.1.1 (<https://github.com/moshi4/CafePlotter>). Gene families that had experienced significant retraction/expansions in more than one transition to herbaceousness were investigated in terms of gene identity and function. To aid interpretation where functional annotations were limited, we also examined gene family expression across wood-forming tissues using a co-expression atlas of wild-growing aspen [*P. tremula*, [Supplemental Materials S2; S25](#)].

Supplemental results

Functional characterization of genes found to be under relaxed selection

To functionally characterize genes found to be under relaxed selection in more than one transition to herbaceousness, we performed gene function enrichment analyses with topGO v2.52.0 [S23,S24; Table 3; Tables S5-S8]. The 19 genes that were found to be under relaxed selection in all transitions to herbaceousness were enriched for GO-terms tied to the “SWR1-complex” (CC domain), transcription and translation (e.g., “transcription preinitiation complex assembly”, “regulation of translational elongation”, both BP), various sugar related metabolic processes (e.g, “starch metabolic process” and “sucrose metabolic process”, both BP), as well as biogenesis of cellular components (“positive regulation of cellular component biogenesis”; BP). Genes found to be under relaxed selection in three out of four gene sets, were enriched for GO-terms related to cell morphogenesis (“cell morphogenesis involved in differentiation”; BP), “xylan biosynthetic process” (BP), as well as FACT and RNA polymerase complexes (e.g, “FACT complex”, “RNA polymerase II/IV/V complex”; CC). Finally, genes under relaxed selection in just two gene sets were enriched for GO-terms related to “cellulose metabolic/biosynthetic processes” (BP), telomere maintenance (e.g., “telomere maintenance via telomerase”; BP), cell growth (e.g , “multidimensional cell growth”; BP) and “RNA binding” functions (MF).

Overview of supplemental figures

Figure S1	this document	Species tree used in ancestral state reconstruction and phylogeny-informed analyses of gene copy numbers
Figure S2	this document	Species tree for Fabales which we used when testing for relaxed selection in hyphy
Figure S3	this document	Species tree for Rosales which we used when testing for relaxed selection in hyphy

Figure S4	this document	Species tree for Fagales & Cucurbitales which we used when testing for relaxed selection in hyphy
Figure S5	this document	Species tree for Malpighiales & Oxalidales which we used when testing for relaxed selection in hyphy
Figure S6	this document	Tree topologies for Fabales used in hyphy (with information on foreground and background branches)
Figure S7	this document	Tree topology for Rosales tested in hyphy (with information on foreground and background branches)
Figure S8	this document	Tree topology for Fagales & Cucurbitales tested in hyphy (with information on foreground and background branches)
Figure S9	this document	Tree topologies for Malpighiales & Oxalidales used in hyphy (with information on foreground and background branches)

Overview of supplemental tables

Table S1	excel spreadsheet	Published genomes used in Figure 1B
Table S2	excel spreadsheet	Dataset information
Table S3	this document	Number of genes with evidence for change in selection pressure on herbaceous branches compared to woody branches using RELAX in hyphy
Table S4	excel spreadsheet	<p>Genes repeatedly under relaxed selection with annotations</p> <ul style="list-style-type: none"> ★ S4A Genes found under relaxed selection in all four transitions based on orthogroup intersections: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales: <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S4B Genes found under relaxed selection in three transitions based on orthogroup intersections: i) Oxalidales & Malpighiales, iii) Fabales, iv) Rosales: <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S4C Genes found under relaxed selection in three transitions based on orthogroup intersections: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales; <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S4D Genes found under relaxed selection in three transitions based on orthogroup intersections: (i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iv) Rosales; <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S4E Genes found under relaxed selection in three transitions based on orthogroup intersections: ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales; <i>Betula pendula</i> annotations ★ S4F Genes found under relaxed selection in two transitions based on orthogroup intersections: (i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales; <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S4G Genes found under relaxed selection in two transitions based on orthogroup intersections: (i) Oxalidales & Malpighiales, iii) Fabales; <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S4H Genes found under relaxed selection only in two transitions based on orthogroup intersections: (i) Oxalidales & Malpighiales, iv) Rosales; <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S4I Genes found under relaxed selection only in two transitions based on orthogroup intersections: ii) Fagales & Cucurbitales, iv) Rosales; <i>Betula pendula</i> annotations ★ S4J Genes found under relaxed selection only in two transitions based on orthogroup intersections: ii) Fagales & Cucurbitales, iii) Fabales; <i>Betula pendula</i> annotations ★ S4K Genes found under relaxed selection two transitions based on orthogroup intersections: iii) Fabales, iv) Rosales; <i>Prunus avium</i> annotations
Table S5	this document	Results from the supertest. Intersections of genes under relaxed selection based on orthology.

Table S6	this document	Numbers of significantly enriched Gene Ontology (GO) terms ($p < 0.05$) in intersecting sets of genes found to be under relaxed selection
Table S7	excel spreadsheet	<p>topGO elim results for the Biological Process Domain (BP)</p> <ul style="list-style-type: none"> ★ S7A GO-terms enriched in the relaxed gene sets found in FOUR transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales ★ S7B GO-terms enriched in the relaxed gene sets found in THREE transitions: i) Oxalidales & Malpighiales, iii) Fabales, iv) Rosales ★ S7C GO-terms enriched in the relaxed gene sets found in THREE transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales ★ S7D GO-terms enriched in the relaxed gene sets found in THREE transitions: (i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iv) Rosales ★ S7E GO-terms enriched in the relaxed gene sets found in THREE transitions: ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales ★ S7F GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales ★ S7G GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, iii) Fabales ★ S7H GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, iv) Rosales ★ S7I GO-terms enriched in the relaxed gene sets found in TWO transitions: ii) Fagales & Cucurbitales, iv) Rosales ★ S7J GO-terms enriched in the relaxed gene sets found in TWO transitions: ii) Fagales & Cucurbitales, iii) Fabales ★ S7K GO-terms enriched in the relaxed gene sets found in TWO transitions: iii) Fabales, iv) Rosales
Table S8	excel spreadsheet	<p>topGO elim results for the Cellular Component Domain (CC)</p> <ul style="list-style-type: none"> ★ S8A GO-terms enriched in the relaxed gene sets found in FOUR transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales ★ S8B GO-terms enriched in the relaxed gene sets found in THREE transitions: i) Oxalidales & Malpighiales, iii) Fabales, iv) Rosales ★ S8C GO-terms enriched in the relaxed gene sets found in THREE transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales ★ S8D GO-terms enriched in the relaxed gene sets found in THREE transitions: (i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iv) Rosales ★ S8E GO-terms enriched in the relaxed gene sets found in THREE transitions: ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales ★ S8F GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales ★ S8G GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, iii) Fabales ★ S8H GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, iv) Rosales ★ S8I GO-terms enriched in the relaxed gene sets found in TWO transitions: ii) Fagales & Cucurbitales, iv) Rosales ★ S8J GO-terms enriched in the relaxed gene sets found in TWO transitions: ii) Fagales & Cucurbitales, iii) Fabales ★ S8K GO-terms enriched in the relaxed gene sets found in TWO transitions: iii) Fabales, iv) Rosales
Table S9	excel spreadsheet	<p>topGO elim results for the Molecular Function Domain (MF)</p> <ul style="list-style-type: none"> ★ S9A GO-terms enriched in the relaxed gene sets found in FOUR transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales ★ S9B GO-terms enriched in the relaxed gene sets found in THREE transitions: i) Oxalidales & Malpighiales, iii) Fabales, iv) Rosales ★ S9C GO-terms enriched in the relaxed gene sets found in THREE transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales ★ S9D GO-terms enriched in the relaxed gene sets found in THREE transitions: (i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iv) Rosales ★ S9E GO-terms enriched in the relaxed gene sets found in THREE transitions: ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales ★ S9F GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales ★ S9G GO-terms enriched in the relaxed gene sets found in TWO transitions: i)

		<p>Oxalidales & Malpighiales, iii) Fabales</p> <ul style="list-style-type: none"> ★ S9H GO-terms enriched in the relaxed gene sets found in TWO transitions: i) Oxalidales & Malpighiales, iv) Rosales ★ S9I GO-terms enriched in the relaxed gene sets found in TWO transitions: ii) Fagales & Cucurbitales, iv) Rosales ★ S9J GO-terms enriched in the relaxed gene sets found in TWO transitions: ii) Fagales & Cucurbitales, iii) Fabales ★ S9K GO-terms enriched in the relaxed gene sets found in TWO transitions: iii) Fabales, iv) Rosales
Table S10	excel spreadsheet	<p>Gene family intersections between all analyses: Statistical testing, random forest, and phylogeny-informed (CAFE)</p> <ul style="list-style-type: none"> ★ S10A All significant gene families/orthogroups based on differential analysis ★ S10B All significant gene families/orthogroups based on random forest ★ S10C All significant gene families/orthogroups based on CAFE ★ S10D Intersecting gene families/orthogroups between differential analysis and random forest ★ S10E Intersecting gene families/orthogroups between differential analysis and CAFE ★ S10F Intersecting gene families/orthogroups between random forest and CAFE ★ S10G Intersecting gene families/orthogroups between differential analysis, random forest, and CAFE
Table S11	this document	Summary of gene families with significant changes ($p < 0.05$) in gene copy numbers found in more than one transition to herbaceousness using CAFE.
Table S12	this document	Detailed overview of gene families with significant changes ($p < 0.05$) in gene copy numbers in more than one transition to herbaceousness using CAFE with gene family/orthogroup IDs. Gene families in bold show exclusively expansions or exclusively retractions in herbaceous transitions.
Table S13	this document	Overview of herbaceous transitions and their node number in CAFE (needed to understand Table S14).
Table S14	this document	Gene families with significant changes in more than one herbaceous transition with details on the exact overlaps. Note that these can be both expansions and retractions.
Table S15	excel spreadsheet	Base change in overlapping gene families from CAFE <ul style="list-style-type: none"> ★ S15A Gene families with significant changes in 5 transitions ★ S15B Gene families with significant changes in 4 transitions ★ S15C Gene families with significant changes in 3 transitions ★ S15D Gene families with significant changes in 2 transitions
Table S16	excel spreadsheet	Annotations of genes in intersecting significant gene families with Orthogroup ID based on <i>Populus tremula</i> and <i>Populus trichocarpa</i> <ul style="list-style-type: none"> ★ S16A significant in five transitions to herbaceousness <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S16B significant in four transitions to herbaceousness <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S16C significant in three transitions to herbaceousness <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2) ★ S16D significant in two transitions to herbaceousness <i>P. tremula</i> annotations (spreadsheet 1) and <i>P. trichocarpa</i> annotations (spreadsheet 2)

Overview of supplemental materials

Supplemental Materials S1	html file	html from R Markdown of differential analysis and Random Forest
Supplemental Materials S2	zip file with pdfs	AspWood expression of gene families with significant copy number differences from differential analysis and random forest

Supplemental figures

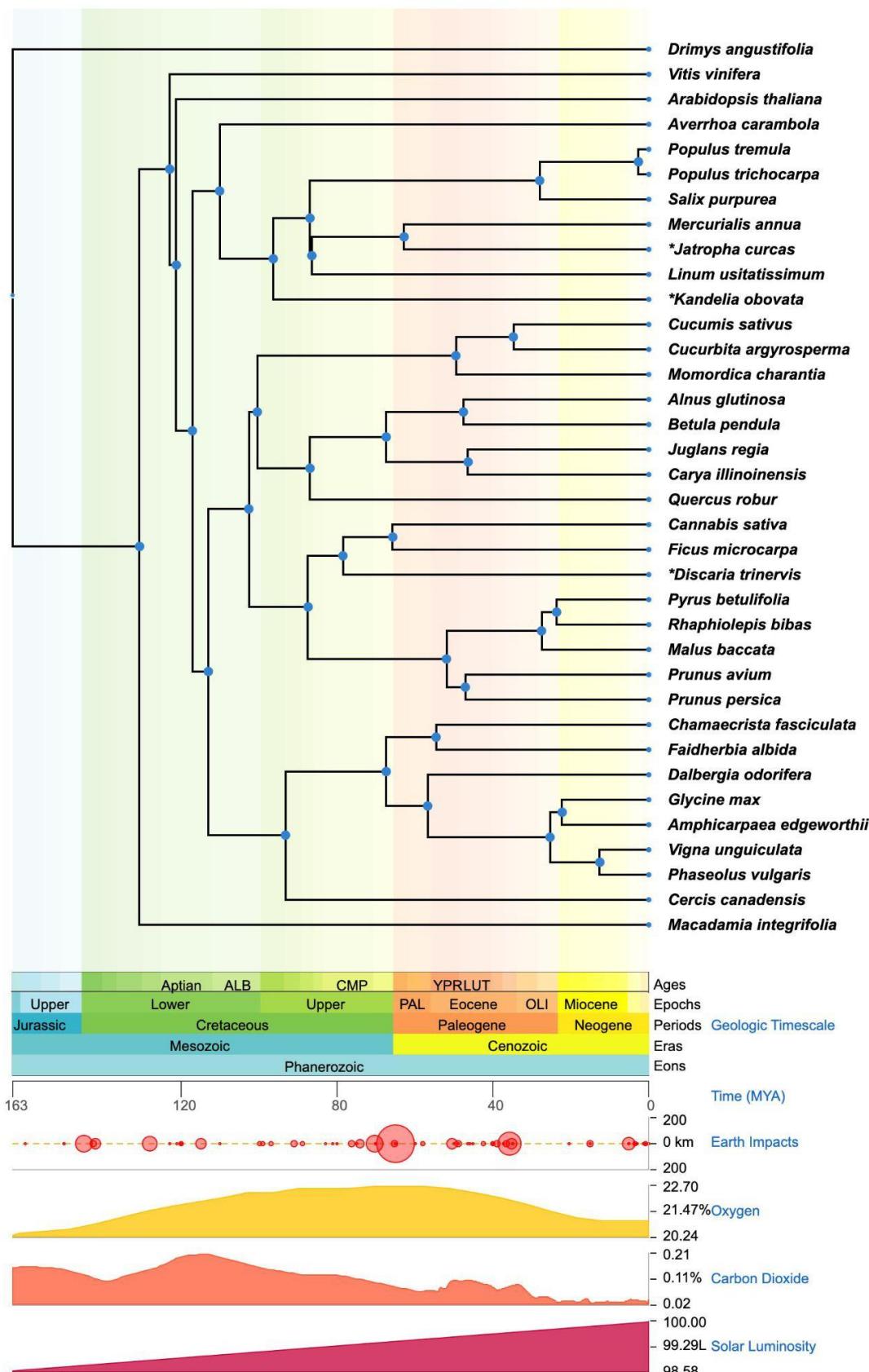


Figure S1. Species tree used in ancestral state reconstruction and phylogeny-informed analyses of gene copy numbers. The tree is based on already published phylogenies and is generated with timetree.org. Species marked with stars were replaced with congeneric species or species within the same family when constructing the tree: *Jatropha curcas* with *Jatropha integerrima*; *Kandelia obovata* with *Kandelia*, and *Discaria trinervis* with *Spyridium parvifolium*.

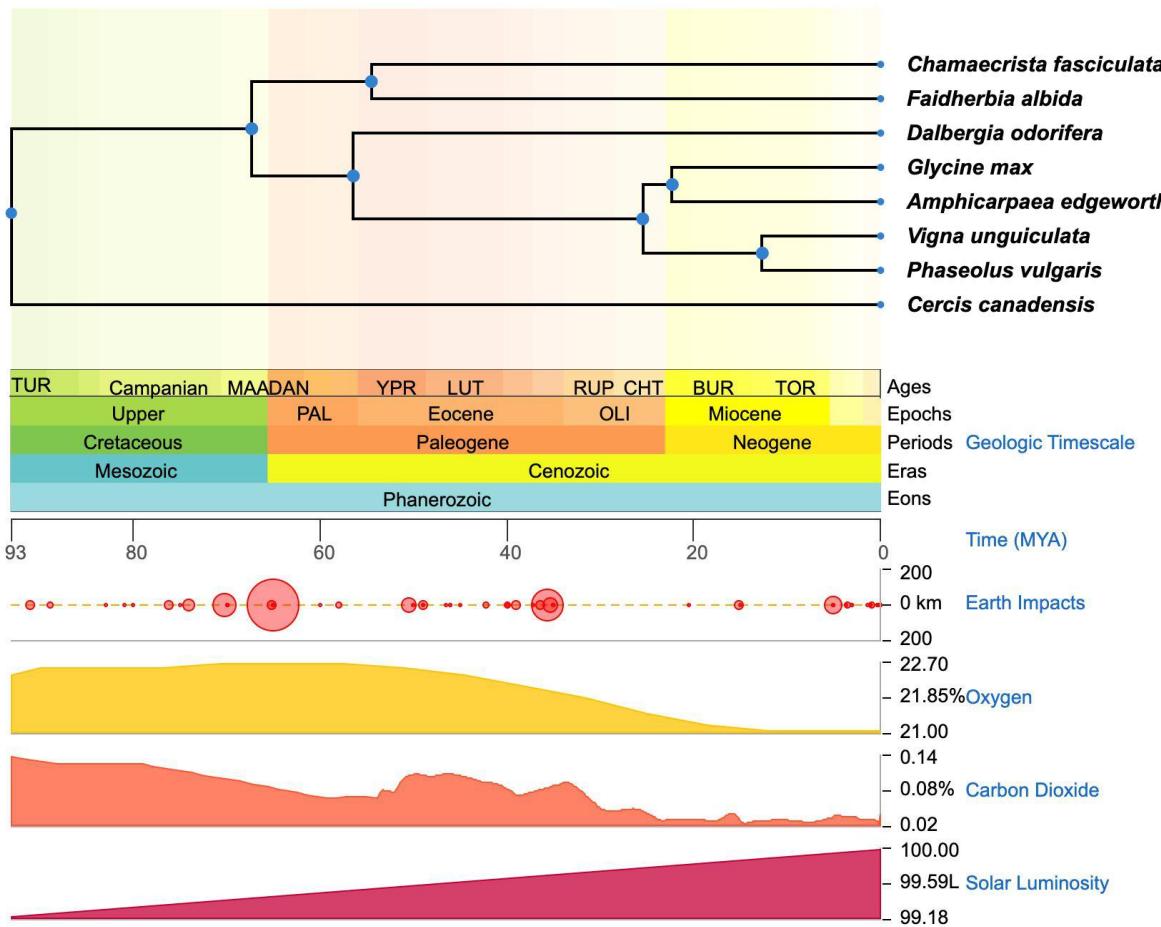


Figure S2. Species tree for Fabales which we used when testing for relaxed selection in hyphy. The tree is based on already published phylogenies and is generated with timetree.org.

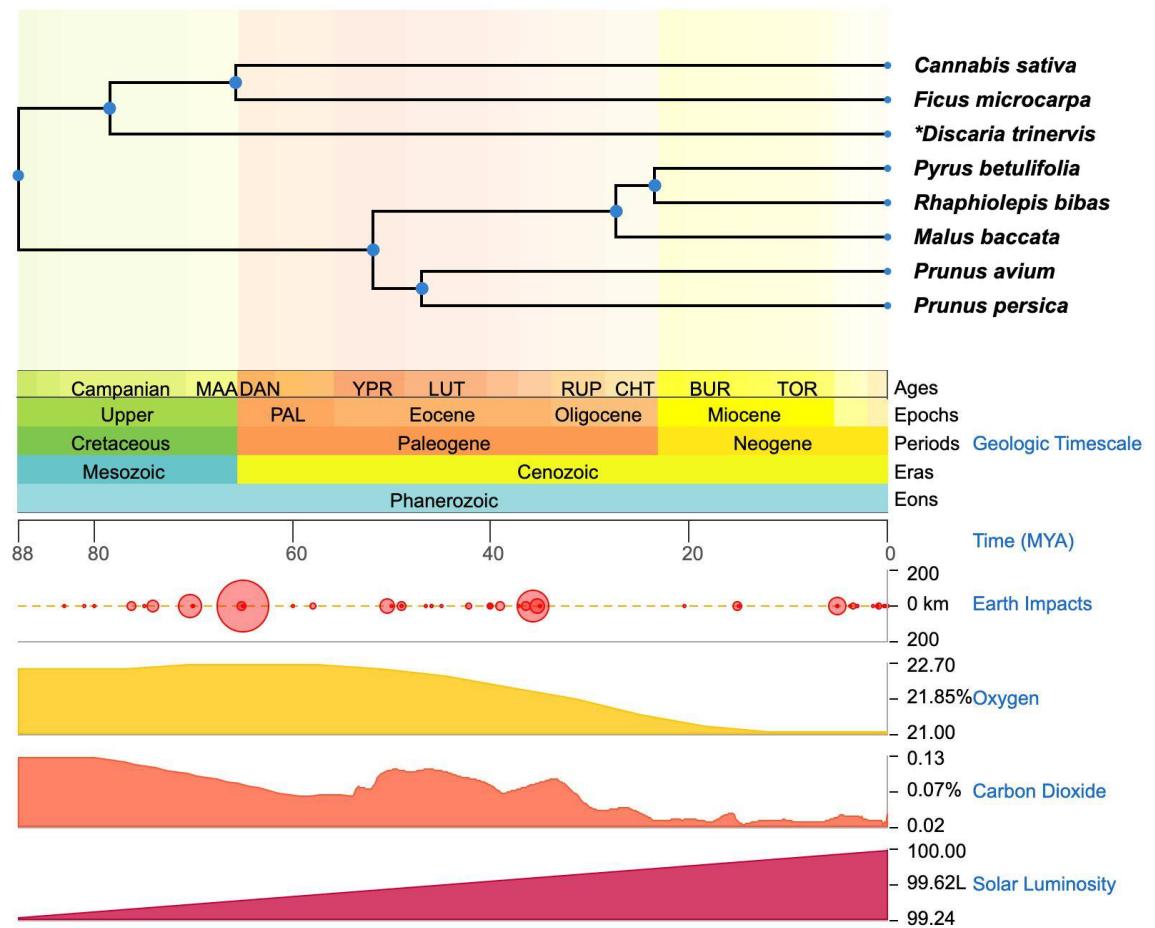


Figure S3. Species tree for Rosales which we used when testing for relaxed selection in hyphy. The tree is based on already published phylogenies and is generated with timetree.org. The species *Discaria trinervis* was replaced with *Spiridium parvifolium* when constructing the tree.

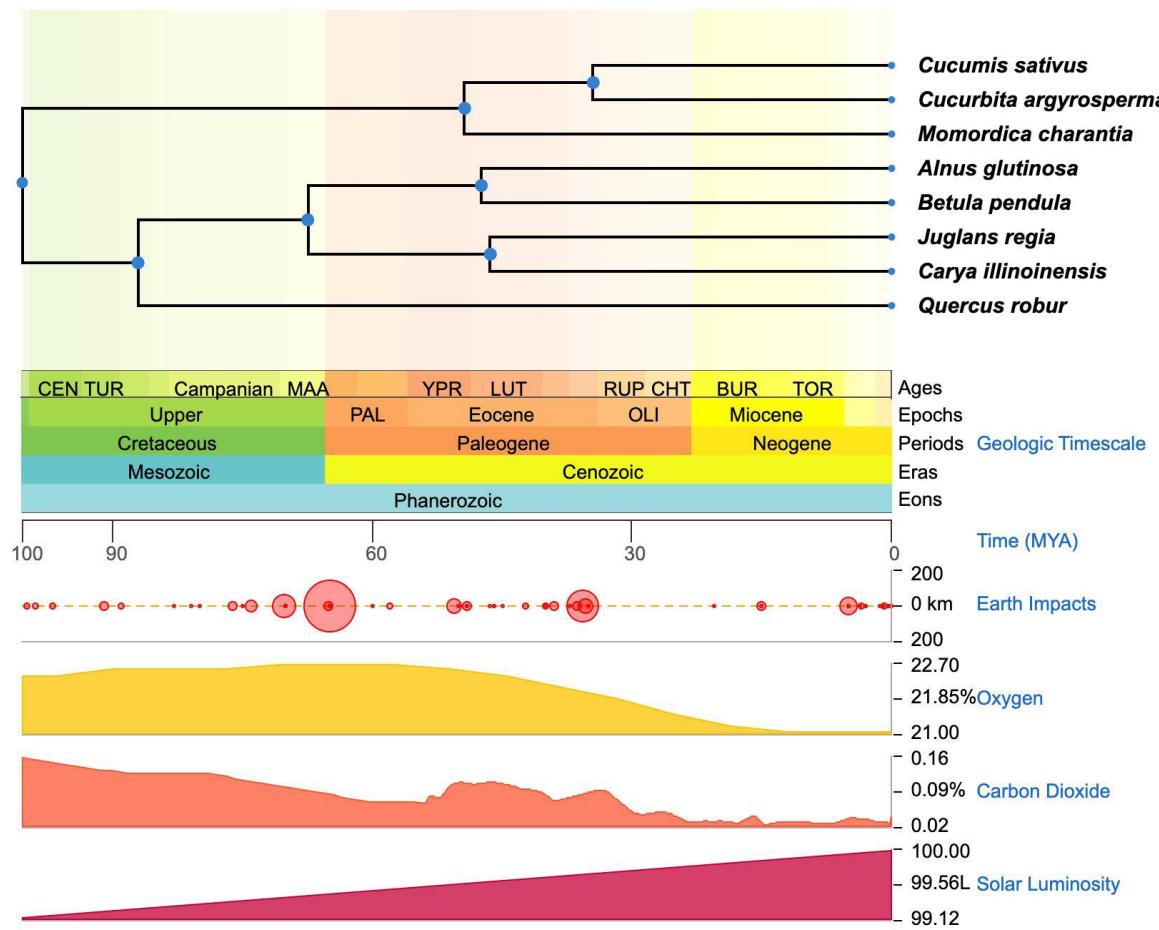


Figure S4. Species tree for Fagales & Cucurbitales which we used when testing for relaxed selection in hyphy. The tree is based on already published phylogenies and is generated with timetree.org.

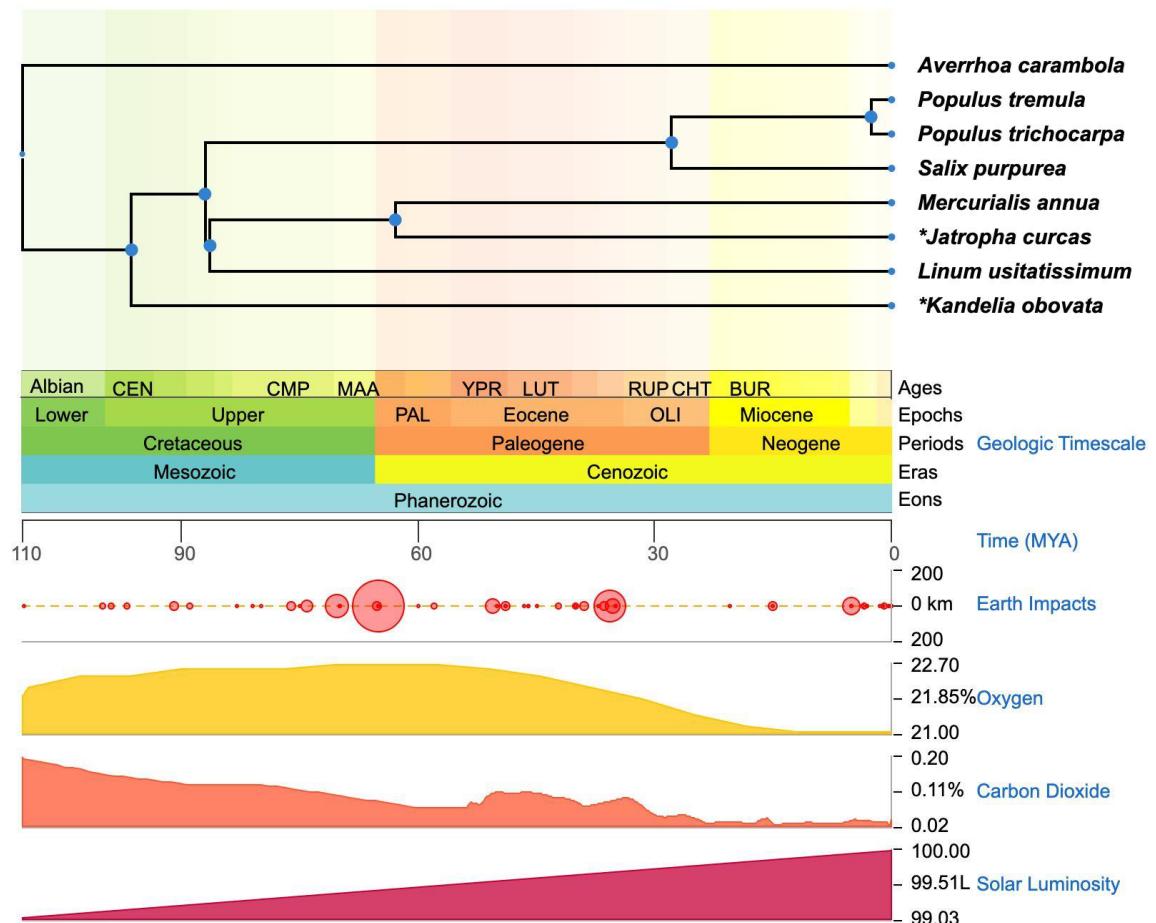


Figure S5. Species tree for Malpighiales & Oxalidales which we used when testing for relaxed selection in hyphy. The tree is based on already published phylogenies and is generated with timetree.org. *Jatropha curcas* was replaced with *Jatropha integerrima* and *Kandelia obovata* with *Kandelia* when constructing the tree.

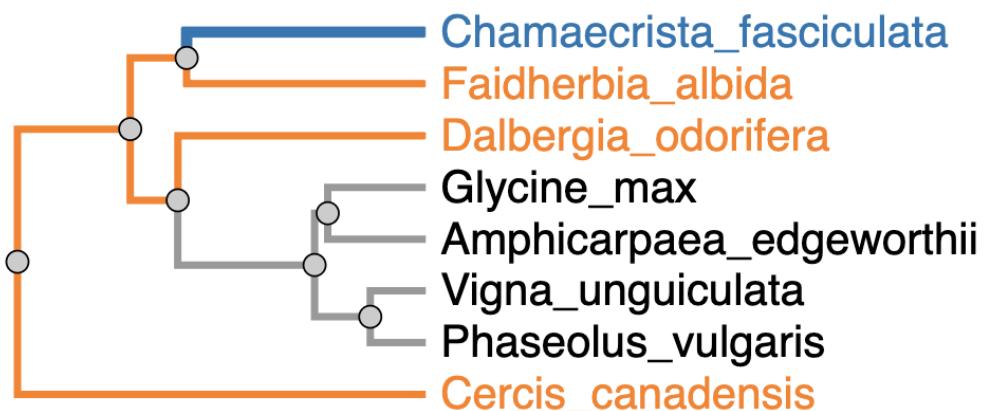
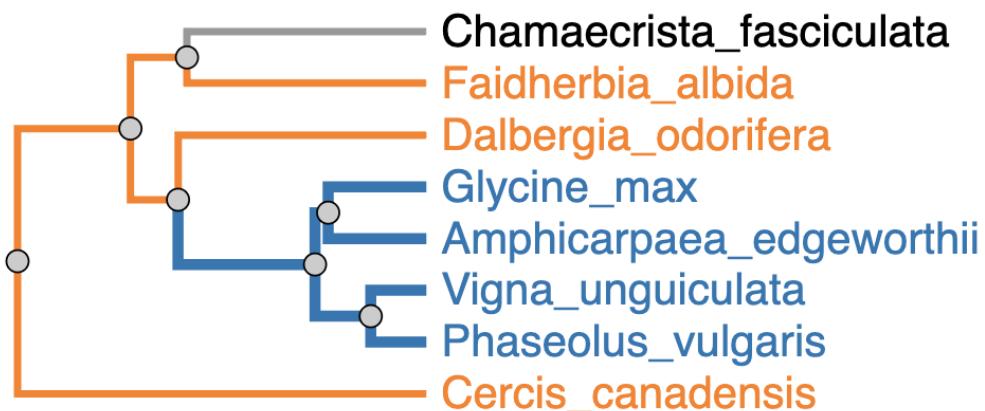
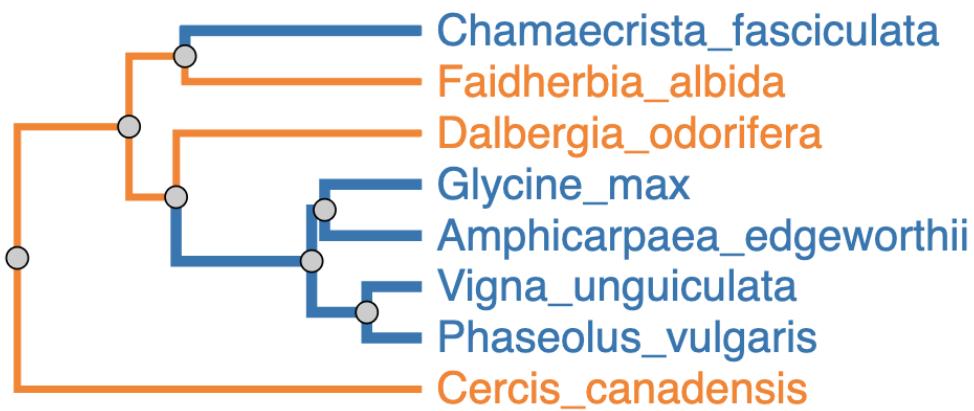


Figure S6. Tree topologies for Fabales used in hyphy. Blue branches were marked as foreground branches, while orange branches were marked as background branches. We tested both transitions together (top) but also tested them individually (middle and bottom).

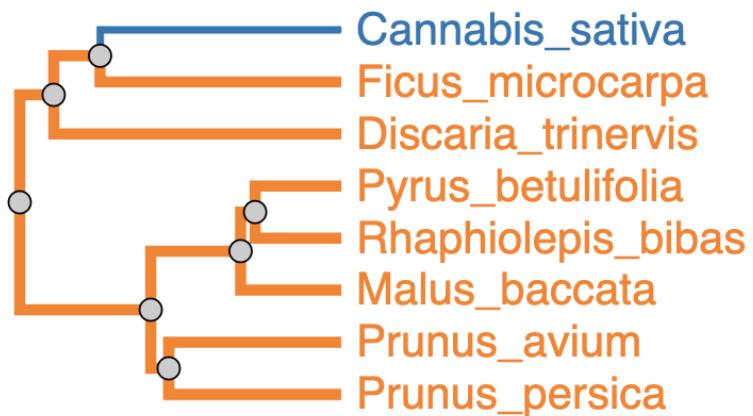


Figure S7. Tree topology for Rosales tested in hyphy. Blue branches were marked as foreground branches, while orange branches were marked as background branches.

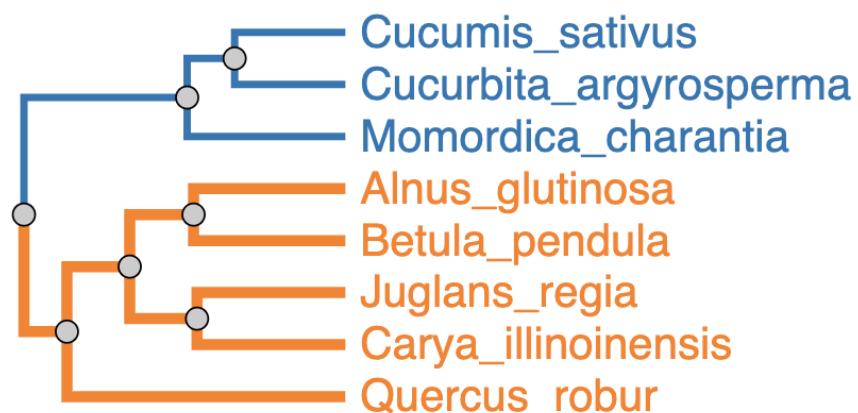


Figure S8. Tree topology for Fagales & Cucurbitales tested in hyphy. Blue branches were marked as foreground branches, while orange branches were marked as background branches.

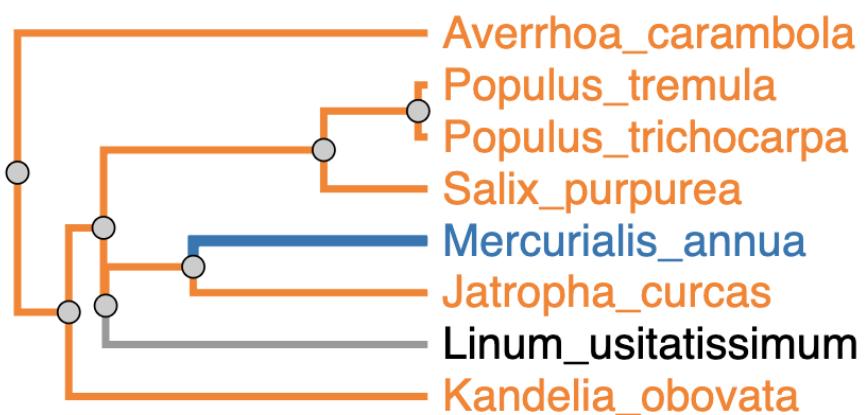
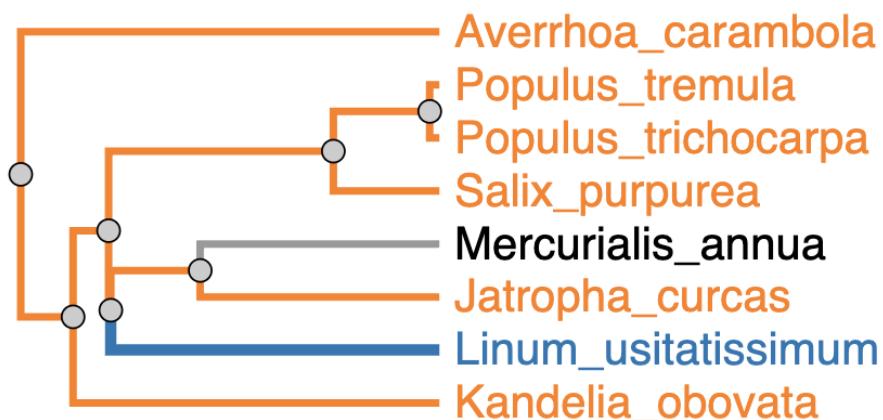
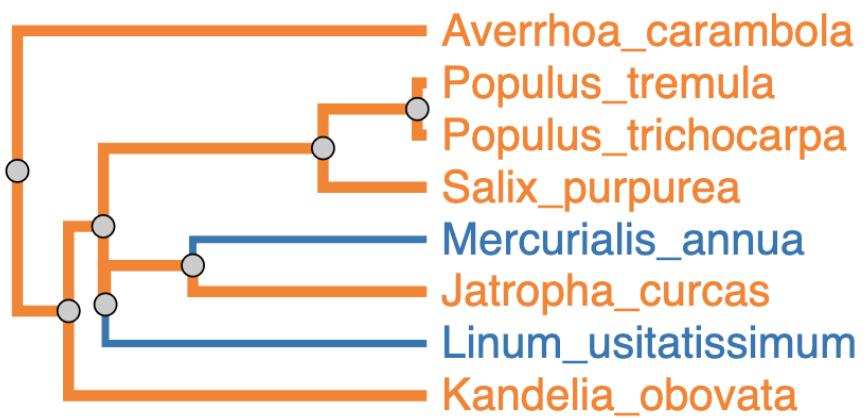


Figure S9. Tree topologies for Malpighiales & Oxalidales used in hyphy. Blue branches were marked as foreground branches, while orange branches were marked as background branches. We tested both transitions together (top) but also individually (middle and bottom).

Supplemental tables

Table S3. Number of genes with evidence for change in selection pressure on herbaceous branches compared to woody branches using RELAX in hyphy.

	SHIFT IN SELECTION PRESSURE	RELAXED	INTENSIFIED	TOTAL ALIGNMENTS
FABALES TRANSITION I + II^{a+b}	3287 [20.3 %]	1661 [10.2 %]	1626 [10.0 %]	16231
FABALES ONLY TRANSITION I^a	2293 [14.1 %]	1276 [7.9 %]	1038 [6.4 %]	16231
FABALES ONLY TRANSITION II^b	1894 [11.7 %]	559 [3.4 %]	1346 [8.3 %]	16231
OXALIDALES & MALPIGIALES TRANSITION I + II^{c+d}	4441 [26.7 %]	482 [2.9 %]	3959 [23.8 %]	16642
OXALIDALES & MALPIGIALES ONLY TRANSITION I^c	4495 [27.0 %]	501 [3.0 %]	3994 [24.0 %]	16642
OXALIDALES & MALPIGIALES ONLY TRANSITION II^d	2895 [17.4 %]	635 [3.8 %]	2260 [13.6 %]	16642
FAGALES & CUCURBITALES	3965 [34.6 %]	871 [7.6 %]	3094 [27.0 %]	11471
ROSALES	2579 [18.8 %]	470 [3.4 %]	2109 [15.4 %]	13724

Transition leading to: ^a*Phaseolus vulgaris*, *Vigna unguiculata*, *Glycine max*, *Amphicarpa edgeworthii*; ^b*Chamaecrista fasciculata*; ^c*Linum usitatissimum*; ^d*Mercurialis annua*

Table S5. Significance of relaxed gene set intersections based on the supertest.

Intersection	Degree	Background	Observed overlap	Expected overlap	FE*	p-value**
Fagales & Cucurbitales, Fabales, Malpighiales & Oxalidales, Rosales	4	11,471	19	0.07	275.57	1.822746e-40
Fabales, Malpighiales, Rosales	3	13,724	35	0.87	40.35	8.578879e-45
Fagales & Cucurbitales, Fabales, Malpighiales & Oxalidales	3	11,471	49	2.03	24.10	1.035167e-51
Fagales & Cucurbitales, Malpighiales & Oxalidales, Rosales	3	11,471	25	0.70	35.69	6.508010e-31
Fagales & Cucurbitales, Fabales, Rosales	3	11,471	55	2.13	25.87	9.446128e-60
Fagales & Cucurbitales, Malpighiales & Oxalidales	2	11,471	85	20.66	4.11	1.588451e-30
Fabales, Malpighiales & Oxalidales	2	16,231	133	25.88	5.14	1.238614e-60
Malpighiales & Oxalidales, Rosales	2	13,724	61	10.54	5.79	1.184605e-29
Fagales & Cucurbitales, Rosales	2	11,471	94	21.60	4.35	6.81242e-36
Fagales & Cucurbitales, Fabales	2	11,471	239	62.69	3.81	5.24757e-85
Fabales, Rosales	2	13,724	140	32.00	4.37	2.22731e-55

*FE = Fold Enrichment. **all less than p < 0.01. We used the lowest number of alignments tested in each intersection as background.

Table S6. Numbers of significantly enriched Gene Ontology (GO) terms ($p < 0.05$) in intersecting sets of genes found to be under relaxed selection.

	Genes intersecting between	Gene sets	BIOLOGICAL PROCESS	CELLULAR COMPONENT	MOLECULAR FUNCTION
A	all relaxed gene sets	i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales	39	14	22
B	three relaxed gene sets	i) Oxalidales & Malpighiales, iii) Fabales, iv) Rosales	25	6	14
C	three relaxed gene sets	i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iii) Fabales	30	6	14
D	three relaxed gene sets	(i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales, iv) Rosales	29	13	11
E	three relaxed gene sets	ii) Fagales & Cucurbitales, iii) Fabales, iv) Rosales	11	1	16
F	two relaxed gene sets	i) Oxalidales & Malpighiales, ii) Fagales & Cucurbitales	24	5	18
G	two relaxed gene sets	i) Oxalidales & Malpighiales, iii) Fabales	59	32	28
H	two relaxed gene sets	i) Oxalidales & Malpighiales, iv) Rosales	74	15	17
I	two relaxed gene sets	ii) Fagales & Cucurbitales, iv) Rosales	13	3	11
J	two relaxed gene sets	ii) Fagales & Cucurbitales, iii) Fabales	22	2	34
K	two relaxed gene sets	iii) Fabales, iv) Rosales	23	4	26

Table S11. Summary of gene families with significant changes ($p < 0.05$) in gene copy numbers found in more than one transition to herbaceousness using CAFE.

No. of transitions to herbaceousness*	No. gene families	No. of gene families with mainly retractions** (-)	No. of gene families with mainly expansions** (+)
6/6	0	0	0
5/6	5	1 (in at least 5)	0 (in at least 5)
4/6	12	6 (in at least 4)	3 (in at least 4)
3/6	40	22 (in at least 3)	7 (in at least 3)
2/6	122	23 (in at least 2)	27 (in at least 2)

*Here counting all transitions separately: ¹*Phaseolus vulgaris*, *Vigna unguiculata*, *Glycine max*, *Amphicarpa edgeworthii* (FABALES TRANSITION I); ²*Chamaecrista fasciculata* (FABALES TRANSITION II); ³*Linum usitatissimum* (OXALIDALES & MALPIGHIALES TRANSITION I); ⁴*Mercurialis annua* (OXALIDALES & MALPIGHIALES TRANSITION II); ⁵*Momordica charantia*, *Cucurbita argyrosperma*, *Cucumis sativus* (CUCURBITALES & FAGALES); ⁶*Cannabis sativa* (ROSALES); **Not counting gene families with equally many retractions and expansions.

Table S12. Detailed overview of gene families with significant changes ($p < 0.05$) in gene copy numbers in more than one transition to herbaceousness using CAFE with gene family/orthogroup IDs. Gene families in bold show exclusively expansions or exclusively retractions in herbaceous transitions.

No. of transitions to herbaceousness*	No. gene families	No. gene families with mainly retractions** (-)	Gene family IDs	No. gene families with mainly expansions** (+)	Gene family IDs
6/6	0				
5/6	5	1 (in at least 5)	OG0000551	0 (in at least 5)	
4/6	12	6 (in at least 4)	OG0000047 OG0000067 OG0000097 OG0000215 OG0000267 OG0000551	3 (in at least 4)	OG0000010 OG0000073 OG0000582
3/6	40	22 (in at least 3)	OG0000038 OG0000041 OG0000045 OG0000047 OG0000067 OG0000082 OG0000085 OG0000091 OG0000097 OG0000098 OG0000156 OG0000157 OG0000209 OG0000215 OG0000248 OG0000267 OG0000281 OG0000365 OG0000423 OG0000551 OG0000612 OG0001191	7 (in at least 3)	OG0000010 OG0000073 OG0000151 OG0000346 OG0000426 OG0000536 OG0000582
2/6	122	23 (in at least 2)	OG0000023 OG0000034 OG0000057 OG0000063 OG0000067 OG0000096 OG0000097 OG0000156 OG0000157 OG0000215 OG0000224 OG0000246 OG0000259 OG0000267 OG0000335 OG0000341 OG0000423 OG0000551 OG0000594	27 (in at least 2)	OG0000010 OG0000066 OG0000073 OG0000084 OG0000104 OG0000154 OG0000155 OG0000249 OG0000273 OG0000301 OG0000346 OG0000437 OG0000490 OG0000491 OG0000536 OG0000539 OG0000574 OG0000631 OG0000642

			OG0000612 OG0000640 OG0000881 OG0001190		OG0000646 OG0001184 OG0001753 OG0002864 OG0003045 OG0003289 OG0003744 OG0003834
--	--	--	--	--	---

*Here counting all transitions separately: ¹*Phaseolus vulgaris*, *Vigna unguiculata*, *Glycine max*, *Amphicarpaea edgeworthii* (FABALES TRANSITION I); ²*Chamaecrista fasciculata* (FABALES TRANSITION II); ³*Linum usitatissimum* (OXALIDALES & MALPIGHIALES TRANSITION I); ⁴*Mercurialis annua* (OXALIDALES & MALPIGHIALES TRANSITION II); ⁵*Momordica charantia*, *Cucurbita argyrosperma*, *Cucumis sativus* (CUCURBITALES & FAGALES); ⁶*Cannabis sativa* (ROSALES); **Not counting gene families with equally many retractions and expansions

Table S13. Overview of herbaceous transitions and their node number in CAFE (needed to understand Table S14).

Herbaceous transition	Node number in CAFE	Order	Herbaceous taxa
1	<17>	Rosales	<i>Cannabis sativa</i>
2	<23>	Malpighiales I	<i>Mercurialis annua</i>
3	<26>	Fabales I	<i>Phaselous vulgaris, Vigna unguiculata, Amphicarpaea edgeworthii, Glycine max</i>
4	<29>	Fabales II	<i>Chamaecrista fasciculata</i>
5	<38>	Malpighiales II	<i>Linum usitatissimum</i>
6	<47>	Cucurbitales	<i>Momordica charantia, Cucurbita argyrosperma, Cucumis sativus</i>

Table S14. Gene families with significant changes in more than one herbaceous transition with details on the exact overlaps. Note that these can be both expansions and retractions.

Herbaceous transitions	Node number	Overlapping gene families	Gene family IDs
6/6	17,23,26,29,38,47	0	
5/6	17,23,26,29,38	0	
5/6	17,23,26,29,47	1	OG0000016
5/6	17,23,26,38,47	0	
5/6	17,23,29,38,47	2	OG0000267 OG0000551
5/6	17,26,29,38,47	2	OG0000010 OG0000582
5/6	23,26,29,38,47	0	
4/6	17,23,26,29	1	OG0000016
4/6	17,23,26,38	0	
4/6	17,23,26,47	2	OG0000016 OG0000097
4/6	17,23,29,38	2	OG0000267 OG0000551
4/6	17,23,29,47	3	OG0000016 OG0000267 OG0000551
4/6	17,23,38,47	3	OG0000215 OG0000267 OG0000551
4/6	17,26,29,38	2	OG0000010 OG0000582
4/6	17,26,29,47	4	OG0000010 OG0000016 OG0000047 OG0000582
4/6	17,26,38,47	3	OG0000010 OG0000067 OG0000582
4/6	17,29,38,47	6	OG0000010 OG0000085 OG0000267 OG0000551 OG0000582 OG0003300
4/6	23,26,29,38	0	
4/6	23,26,29,47	1	OG0000016
4/6	23,26,38,47	1	OG0000073
4/6	23,29,38,47	2	OG0000267 OG0000551
4/6	26,29,38,47	2	OG0000010 OG0000582
3/6	17,23,26	3	OG0000016 OG0000097

			OG0001198
3/6	17,23,29	3	OG0000016 OG0000267 OG0000551
3/6	17,23,38	6	OG0000091 OG0000151 OG0000215 OG0000267 OG0000404 OG0000551
3/6	17,23,47	6	OG0000016 OG0000097 OG0000215 OG0000267 OG0000536 OG0000551
3/6	17,26,29	4	OG0000010 OG0000016 OG0000047 OG0000582
3/6	17,26,38	4	OG0000010 OG0000041 OG0000067 OG0000582
3/6	17,26,47	6	OG0000010 OG0000016 OG0000047 OG0000067 OG0000097 OG0000582
3/6	17,29,38	9	OG0000010 OG0000064 OG0000085 OG0000267 OG0000346 OG0000365 OG0000551 OG0000582 OG0003300
3/6	17,29,47	9	OG0000010 OG0000016 OG0000047 OG0000085 OG0000267 OG0000551 OG0000582 OG0001191 OG0003300
3/6	17,38,47	14	OG0000010 OG0000020 OG0000067 OG0000085 OG0000098 OG0000156 OG0000157 OG0000215 OG0000267 OG0000281 OG0000354 OG0000551 OG0000582 OG0003300
3/6	23,26,29	2	OG0000016 OG0004529
3/6	23,26,38	3	OG0000073 OG0000082 OG0000426
3/6	23,26,47	3	OG0000016 OG0000073 OG0000097
3/6	23,29,38	4	OG0000045 OG0000068 OG0000267

			OG0000551
3/6	23,29,47	5	OG0000016 OG0000248 OG0000267 OG0000551 OG0000612
3/6	23,38,47	8	OG0000021 OG0000038 OG0000073 OG0000209 OG0000215 OG0000267 OG0000423 OG0000551
3/6	26,29,38	2	OG0000010 OG0000582
3/6	26,29,47	4	OG0000010 OG0000016 OG0000047 OG0000582
3/6	26,38,47	5	OG0000010 OG0000067 OG0000073 OG0000162 OG0000582
3/6	29,38,47	6	OG0000010 OG0000085 OG0000267 OG0000551 OG0000582 OG0003300
2/6	17,23	16	OG0000016 OG0000091 OG0000097 OG0000151 OG0000154 OG0000215 OG0000267 OG0000404 OG0000418 OG0000490 OG0000536 OG0000551 OG0000631 OG0000646 OG0001198 OG0003744
2/6	17,26	11	OG0000010 OG0000016 OG0000041 OG0000047 OG0000067 OG0000097 OG0000539 OG0000582 OG0001036 OG0001198 OG0003289
2/6	17,29	18	OG0000010 OG0000016 OG0000047 OG0000064 OG0000085 OG0000267 OG0000273 OG0000346 OG0000363 OG0000365 OG0000437 OG0000551 OG0000574 OG0000582 OG0001184

			OG0001191 OG0002864 OG0003300
2/6	17,38	38	OG0000010 OG0000020 OG0000041 OG0000053 OG0000063 OG0000064 OG0000067 OG0000085 OG0000091 OG0000096 OG0000098 OG0000104 OG0000111 OG0000144 OG0000151 OG0000155 OG0000156 OG0000157 OG0000214 OG0000215 OG0000216 OG0000241 OG0000267 OG0000277 OG0000281 OG0000315 OG0000346 OG0000354 OG0000365 OG0000368 OG0000374 OG0000404 OG0000491 OG0000551 OG0000582 OG0003045 OG0003300 OG0003834
2/6	17,47	23	OG0000010 OG0000016 OG0000020 OG0000047 OG0000057 OG0000067 OG0000085 OG0000097 OG0000098 OG0000156 OG0000157 OG0000215 OG0000267 OG0000281 OG0000301 OG0000335 OG0000354 OG0000536 OG0000551 OG0000582 OG0001190 OG0001191 OG0003300
2/6	23,26	11	OG0000016 OG0000073 OG0000082 OG0000097 OG0000118 OG0000246 OG0000350 OG0000426 OG0001198

			OG0001742 OG0004529
2/6	23,29	8	OG0000016 OG0000045 OG0000068 OG0000248 OG0000267 OG0000551 OG0000612 OG0004529
2/6	23,38	30	OG0000021 OG0000038 OG0000045 OG0000068 OG0000073 OG0000077 OG0000082 OG0000091 OG0000137 OG0000151 OG0000180 OG0000193 OG0000209 OG0000215 OG0000224 OG0000267 OG0000341 OG0000352 OG0000396 OG0000404 OG0000423 OG0000426 OG0000500 OG0000551 OG0000594 OG0000929 OG0001046 OG0001141 OG0001795 OG0003049
2/6	23,47	18	OG0000016 OG0000021 OG0000023 OG0000034 OG0000038 OG0000059 OG0000073 OG0000097 OG0000112 OG0000209 OG0000215 OG0000248 OG0000267 OG0000423 OG0000536 OG0000551 OG0000612 OG0001143
2/6	26,29	6	OG0000010 OG0000016 OG0000047 OG0000582 OG0000758 OG0004529
2/6	26,38	14	OG0000010 OG0000041 OG0000067 OG0000073 OG0000074 OG0000082 OG0000150 OG0000162 OG0000164

			OG0000426 OG0000455 OG0000582 OG0000642 OG0001753
2/6	26,47	9	OG0000010 OG0000016 OG0000047 OG0000067 OG0000073 OG0000097 OG0000162 OG0000582 OG0000617
2/6	29,38	19	OG0000010 OG0000042 OG0000045 OG0000064 OG0000066 OG0000068 OG0000084 OG0000085 OG0000223 OG0000249 OG0000259 OG0000267 OG0000346 OG0000365 OG0000551 OG0000582 OG0000607 OG0002722 OG0003300
2/6	29,47	13	OG0000010 OG0000016 OG0000047 OG0000061 OG0000085 OG0000248 OG0000267 OG0000551 OG0000582 OG0000612 OG0000640 OG0001191 OG0003300
2/6	38,47	24	OG0000010 OG0000020 OG0000021 OG0000031 OG0000038 OG0000067 OG0000073 OG0000085 OG0000098 OG0000121 OG0000156 OG0000157 OG0000162 OG0000209 OG0000215 OG0000267 OG0000281 OG0000354 OG0000423 OG0000527 OG0000551 OG0000582 OG0000881 OG0003300

Supplemental References

- S1. Paradis, E. *et al.* (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289–290
- S2. Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528
- S3. Revell, L.J. (2012) phytools: An R package for phylogenetic comparative biology (and other things) *Methods in Ecology and Evolution*, 3217–223
- S4. Kumar, S. *et al.* (2022) TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol. Biol. Evol.* 39
- S5. Wertheim, J.O. *et al.* (2015) RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32, 820–832
- S6. Love, M.I. *et al.* (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 1–21
- S7. Mendes, F.K. *et al.* (2021) CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518
- S8. Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238
- S9. Chang, S. *et al.* (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11, 113–116
- S10. Manni, M. *et al.* (2021) BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc* 1, e323
- S11. Friedman, J. (2020) The Evolution of Annual and Perennial Plant Life Histories: Ecological Correlates and Genetic Mechanisms. *Annu. Rev. Ecol. Evol. Syst.* 51, 461–481
- S12. Katoh, K. *et al.* (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33, 511–518
- S13. Kimura, M. (1985) *The Neutral Theory of Molecular Evolution*, Cambridge University Press
- S14. Birkeland, S. *et al.* (2020) Multiple Genetic Trajectories to Extreme Abiotic Stress Adaptation in Arctic Brassicaceae. *Mol. Biol. Evol.* 37, 2052–2068
- S15. Rice, P. *et al.* (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277
- S16. Löytynoja, A. and Goldman, N. (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10557–10562
- S17. Sela, I. *et al.* (2015) GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43, W7–14
- S18. Jordan G, Goldman N. (2012) The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol Biol Evol.* 29(4):1125–1139.
- S19. Privman E, Penn O, Pupko T. (2012) Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol.* 29(1):1–5.
- S20. Conway, J.R. *et al.* (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940
- S22. Wang M, Zhao Y, Zhang B. (2015) Efficient test and visualization of multi-set intersections. *Sci Rep.* 5(16923):1–12.
- S23. Alexa, A. *et al.* (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22, 1600–1607
- S24. Gentleman, R.C. *et al.* (2004) *Bioconductor: open software development for computational biology and bioinformatics*
- S25. Sundell, D., Street, N. R., Kumar, M., Mellerowicz, E. J., Kucukoglu, M., Johnsson, C., Kumar, V., Mannapperuma, C., Delhomme, N., Nilsson, O., Tuominen, H.,

Pesquet, E., Fischer, U., Niittylä, T., Sundberg, B., & Hvidsten, T. R. (2017). AspWood: High-Spatial-Resolution Transcriptome Profiles Reveal Uncharacterized Modularity of Wood Formation in *Populus tremula*. *The Plant Cell*, 29(7), 1585–1604.