


Opinion

# An updated perspective: what genes make a tree a tree?

Siri Birkeland <sup>1,2,\*</sup>, Eduardo R. Soldado<sup>3</sup>, Sonali S. Ranade<sup>4</sup>, M. Rosario García-Gil<sup>4</sup>, Shruti Choudhary<sup>4</sup>, Vikash Kumar<sup>3</sup>, Hannele Tuominen<sup>4</sup>, Ewa J. Mellerowicz<sup>4</sup>, Nathaniel R. Street<sup>3,5</sup>, and Torgeir R. Hvidsten<sup>1,\*</sup>

**We learn early on how to tell trees apart from other plants. However, it has proved hard to distinguish trees from other plants at the genetic level, and it is believed that there are no unique ‘tree genes’. With the rapid increase in available tree genomes, we can perform new comparative and evolutionary analyses of plant life histories and growth forms. Here we provide a fresh perspective on the genetic foundation for woodiness and perenniality in angiosperms by analyzing selection pressures and gene family evolution in the rosids using genomic data. We examine genes distinguishing trees from herbs and discuss future directions for uncovering the genetic factors that define a tree in this new era of tree genomics.**

## A new era in tree genomics

Trees are not a single, unified group but a diverse collection of plants sharing a perennial life history and a woody growth habit (Figure 1A) [1,2]. Their woody stems allow them to grow larger and taller than other plants, making them some of the world’s largest and longest-living organisms. Notable examples include the 4855-year-old *Pinus longaeva* (‘Methuselah’) and the 115-m-tall *Sequoia sempervirens* (‘Hyperion’). The evolution of woody trunks made up of secondary vascular tissue predates the emergence of the seed and is observed in both gymnosperms and angiosperms [3]. In angiosperms, a woody growth form is considered to be ancestral [1,2,4–6] but has been lost and gained multiple times throughout their evolutionary history. Over 40% of today’s angiosperm species are estimated to be woody [7].

Although recent discoveries have shed light on the evolution of growth forms in flowering plants [8,9], our understanding of a potentially shared molecular basis remains limited. In 2005, Andrew Groover proposed a new perspective on the genes that make a tree a tree [1]. He observed that key genes regulating shoot apical meristem activity of herbaceous plants are also active in the vascular cambium during woody growth, suggesting that there are likely no unique ‘tree genes’ that would be absent in herbaceous plants [1]. Instead, the defining feature might lie in unique expression patterns of shared genes [1]. With the publication of the first tree genome in 2006 (*Populus trichocarpa*) [10] and a nearly exponential increase in the number of published tree genomes since then (Figure 1B; Table S1 in the supplemental information online), it seems timely to revisit whether there are genes defining a woody growth habit and explore whether a shared molecular basis exists for growth form evolution in angiosperms.

## Evolution of life history strategies and growth forms

Evolutionary shifts between trees and herbs have occurred frequently within many lineages of flowering plants, suggesting that few evolutionary steps separate the two growth forms [2,9]. Traditionally, it has been assumed that woody ancestors more often give rise to herbaceous plants [1,2]. However, recent data suggest that the reverse transition, the (re-)evolution of woody plants

## Highlights

In his 2005 paper ‘What genes make a tree a tree?’ Andrew Groover argued that it is unlikely that unique ‘tree genes’ are responsible for woody growth; rather, he suggested that the defining feature of trees lies in the distinct regulatory patterns of genes shared by both herbaceous and woody plants.

The first tree genome, *Populus trichocarpa*, was published in 2006, leading to an exponential growth in the number of available tree genomes since then. Using these new genomic data, we performed comparative analyses of selection pressures and gene copy numbers to uncover new insights into what genes make a tree a tree.

Our results highlight genes that may distinguish trees from herbs and suggest that factors beyond regulatory differences, such as pseudogenization and differences in gene copy numbers, may contribute to this distinction.

<sup>1</sup>Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

<sup>2</sup>Natural History Museum, University of Oslo, Oslo, Norway

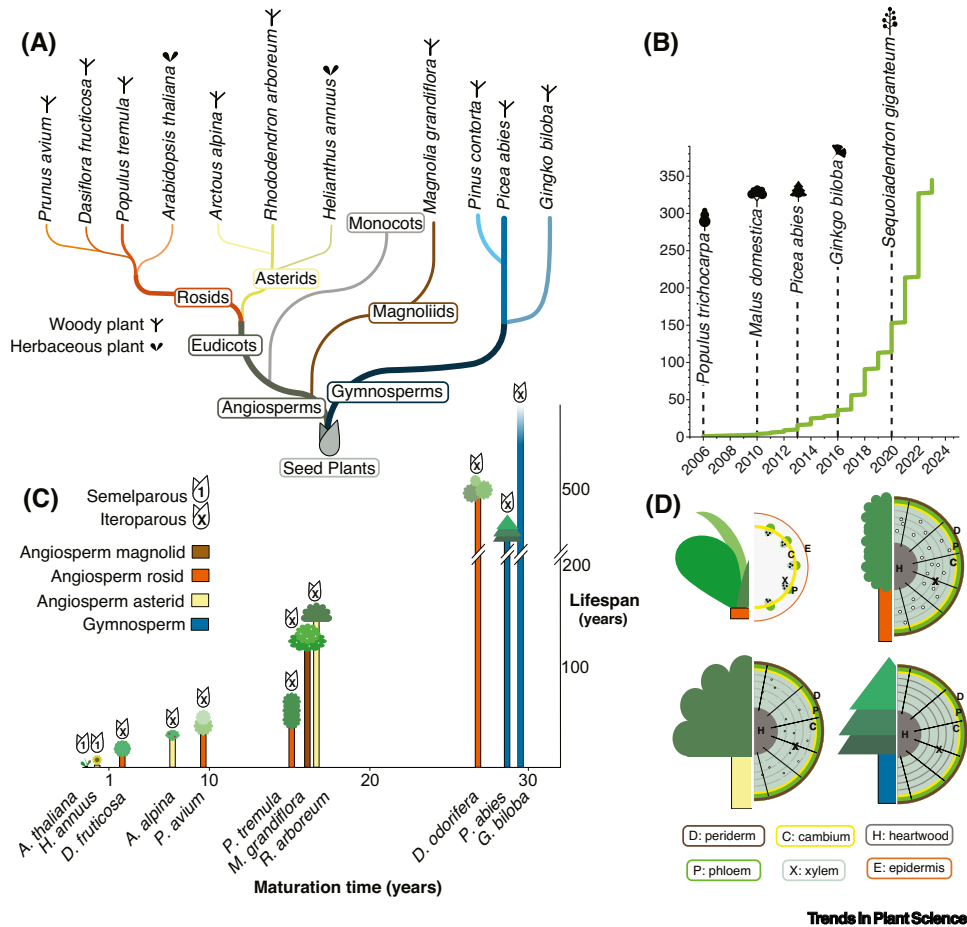
<sup>3</sup>Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden

<sup>4</sup>Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden

<sup>5</sup>SciLifeLab, Umeå University, Umeå, Sweden

\*Correspondence: [siri.birkeland@nhm.uio.no](mailto:siri.birkeland@nhm.uio.no) (S. Birkeland) and [torgeir.r.hvidsten@nmbu.no](mailto:torgeir.r.hvidsten@nmbu.no) (T.R. Hvidsten).





Trends in Plant Science

**Figure 1. Variation in growth form and plant life history traits.** (A) Simplified phylogeny of a selection of seed plants showing that trees are found in multiple plant groups. (B) Total number of arborescent species with published genomes (both angiosperms and gymnosperms). We included only species described as trees in online floras and excluded shrubs to avoid too much ambiguity because woodiness is a gradient rather than a presence/absence trait. However, woody shrubs that occasionally grow into small trees were included. Publication data collected from [https://plabipd.de/plant\\_genomes\\_pa.ep](https://plabipd.de/plant_genomes_pa.ep). (C) Plant species variation in maturation time and lifespan (note that both axes are related to the fast–slow continuum). (D) Examples of tissue types in species with secondary growth: *Arabidopsis thaliana* (top left), *Populus tremula* (top right), *Rhododendron arboreum* (bottom left), and *Picea abies* (bottom right).

from herbaceous ancestors, may be just as common [8]. A striking example of this is insular woodiness, where woody species repeatedly evolve from herbaceous ancestors on islands, likely due to increased drought, favorable aseasonal climate leading to continuous growth, and reduced competition [9]. This illustrates a close connection between plant growth form and environmental conditions [7,8].

The distinction between woodiness and herbaceousness is also closely tied to life history strategies, patterns of reproduction, growth, and survival that vary among plant species. Much of the variation in plant life histories can be understood through two main axes: (i) the fast–slow continuum, which ranges from fast-growing, short-lived species to slow-growing, long-lived ones (Figure 1C); and (ii) the number of reproductive cycles a species undergoes throughout its lifetime [11]. At one extreme of this life history spectrum are tall trees, which tend to grow slowly, live long lives, and reproduce many times (i.e., they are iteroparous). At the other end are fast-growing

herbaceous annuals, which grow quickly, reproduce once (i.e., they are semelparous), and then die. Although certain growth forms tend to occupy distinct areas within the 2D space defined by the fast–slow continuum and number of reproductive cycles, there are also many exceptions. For example, fast-growing small trees and shrubs may be positioned similarly to herbaceous perennials along the fast–slow continuum and exhibit comparable numbers of reproductive cycles [11].

### The genetic underpinnings of growth form and plant life histories

The genes distinguishing trees from herbs may lie in developmental programs controlling flowering and meristem determination. Woody perennials with extended juvenile phases accumulate biomass before flowering and maintain indeterminate meristems for repeated reproductive and vegetative cycles [12]. By contrast, herbaceous annuals have shorter juvenile phases, flower once, and die [1,2,12,13]. In *Arabidopsis thaliana*, disrupting flowering time genes *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *FRUITFULL* (*FUL*) delayed flowering, induced secondary growth, and reverted meristems to a vegetative state [14]. Perennials also use mechanisms such as endodormancy to survive extreme seasons [15] and may perhaps also have an extended ability to resist biotic and abiotic stresses [16]. This demonstrates that genetic switches between annual and perennial life histories involve not just timing of flowering and meristem determinacy but also the plants' resilience to environmental stress.

An important distinction between trees and herbs is the ability to produce wood through secondary growth. Wood, or secondary xylem, consists mainly of water-conducting cells (tracheids and/or vessel elements) and support cells (fibers) with walls reinforced with cellulose and lignin for mechanical support [17–19]. During secondary growth, wood cells are formed inward in the stem or root and phloem cells outward through cell division via the vascular cambium. Wood formation from a continuous cylinder of vascular cambium is generally lacking in monocots [1], but other types of wood formation exist within that group. Although most herbaceous plants exhibit secondary growth, trees achieve exceptional biomass accumulation through this process (Figure 1D) [1]. Key regulators of wood formation in poplar include *WUSCHEL-RELATED HOMEODOMAIN 4* (*WOX4*)-like genes, which control vascular cambium activity [20]; *LATERAL ORGAN BOUNDARIES DOMAIN* (*LBD*) genes, such as *LBD1*, driving vascular development [21,22]; and NAC-domain genes such as *PtVNS* (*Populus trichocarpa* VND-, NST/SND-, SMB-related proteins), which regulate xylem differentiation and secondary cell wall formation [23]. Genes controlling xylogenesis and secondary cell wall formation could be important in evolutionary transitions between trees and herbs.

Beyond gene-level patterns, there is an interesting association between genome size and plant life histories. Larger genomes are generally associated with slower cell division and longer generation times, whereas small genomes may allow faster development and greater flexibility in life cycle strategies [2,24]. Pioneering work by Bennett [24,25] demonstrated a positive correlation between angiosperm genome size and minimum generation time – the shortest possible time it takes for a plant to complete its life cycle. His findings revealed intriguing threshold effects: species with genomes above certain sizes were no longer capable of rapid life cycles. For instance, no ephemeral species were found with genome sizes over 3.4 pg, and all species with genomes above 27.6 pg were obligate perennials [24]. These patterns help explain why evolutionary shifts from perennial to annual life histories are often accompanied by reductions in genome size, especially in harsh or seasonal environments [24]. Despite this, evidence for consistent genome size differences between herbaceous and woody growth forms remains limited. In fact, a woody growth form may even constrain genome size evolution because of inherently long generation times [26]. Still, many gymnosperms, such as conifers, have very large genomes partly due to retrotransposon activity [27].

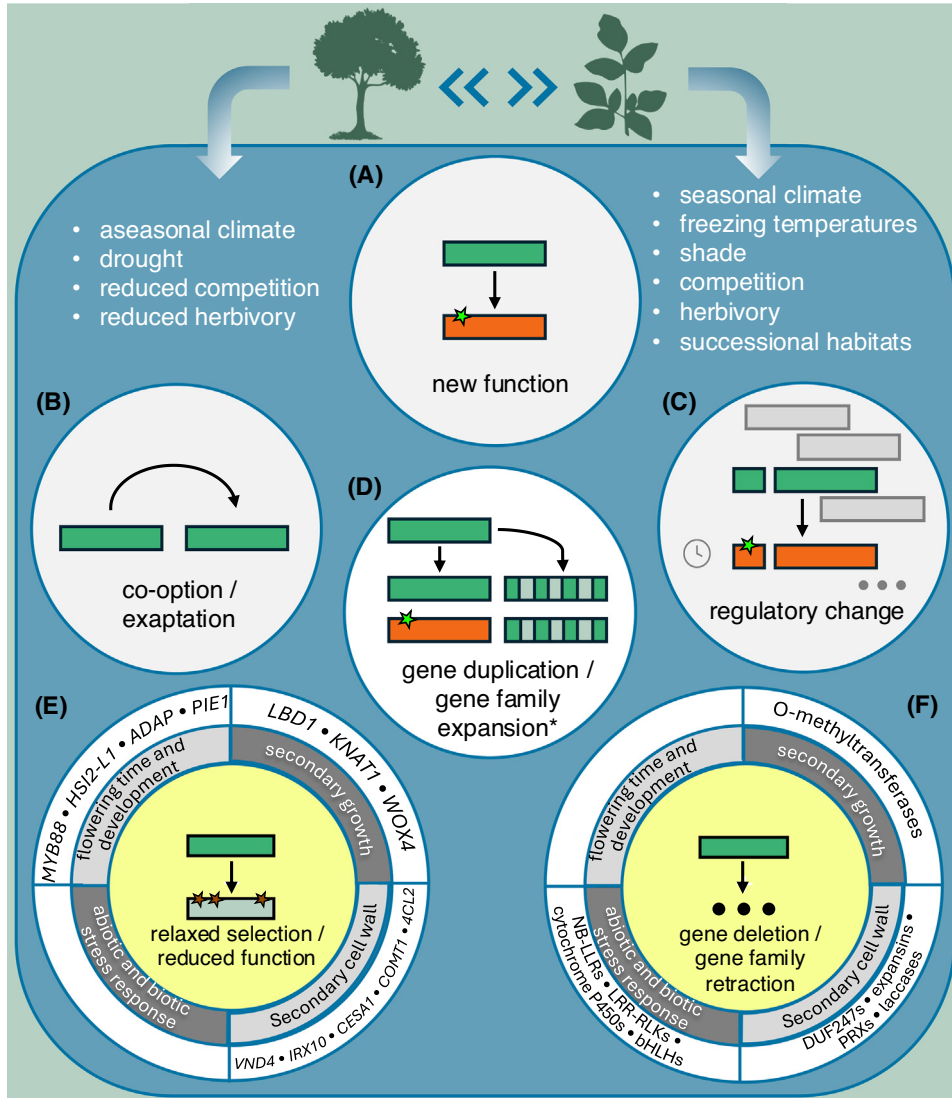
### A study of published tree genomes

There have so far been few comparative genomic studies exploring the evolution of growth forms and differences in longevity in angiosperms. To demonstrate the power inherent in new tree genomes, we performed analyses of selection pressures and gene family evolution within the rosids to give a fresh perspective on the evolution of woodiness and growth form. We focus on the rosids because this angiosperm clade contains most of our well-known forest trees and most of the assembled genomes and because it radiated at the same time as angiosperm forests became predominant on Earth [28]. We investigated two types of genetic change to uncover shared mechanisms of growth form evolution, with the goal of revealing specific genes that define the essence of woodiness. First, we looked at genes that have repeatedly experienced a reduction in selection pressures in herbaceous plants based on increased accumulation of mutations in evolutionary transitions from woodiness to herbaceousness (Figure 2A; but see also Figure 2B–F for an overview of different genetic changes). These genes may conversely be under purifying (or positive) selection in trees, indicating a role in woodiness. On the basis of a similar logic, we also expected gene families that repeatedly experienced a contraction (i.e., a reduction in gene copy number) in transitions from woodiness to herbaceousness to be associated with a woody growth form (Figure 2D). By design, our approach emphasizes genes and gene families that recur across independent lines of evidence, increasing the likelihood that these reflect robust evolutionary patterns linked to growth form. Note that we did not examine the evolution of genome size, because species with smaller and less complex genomes tend to be overrepresented among those with available genome assemblies. Moreover, we included primarily diploid species to maximize alignment reliability. Genome size evolution in angiosperms has been explored elsewhere, particularly in the context of trees [e.g., 26].

### Overview of approach and results

To investigate whether specific genes distinguish trees from herbs, we ran RELAX in HyPhy [29] to test for relaxed selection in transitions to herbaceousness in four groups: (i) Fabales, (ii) Rosales, (iii) Fagales and Cucurbitales, and (iv) Malpighiales and Oxalidales (11 471–16 642 alignments tested per order/group; Figure 3A; Table S2 in the supplemental information online). This analysis identified 470–1661 genes under relaxed selection and 1626–3959 genes under intensified selection ( $P \leq 0.05$ ; Tables S3–S9 in the supplemental information online). Among genes under relaxed selection, 19 intersected across all orders (Figure 3B; Table S4 in the supplemental information online). This overlap was higher than expected by chance (Table S5 in the supplemental information online) and included chromatin-remodeling proteins, transcription factors, and genes tied to wood development.

Next, we analyzed gene family size differences between woody and herbaceous species using phylogeny-independent and phylogeny-informed approaches (supplemental information online). Statistical tests (based on negative binomial distribution) revealed 26 gene families with significant differences, whereas a machine-learning approach based on random forest identified 101 families (Supplemental Materials S1 and S2 in the supplemental information online). Sixteen gene families were significant in both approaches (Table S10 in the supplemental information online) and included families involved in wood formation, growth, and abiotic stress responses. We also ran CAFE [30] to model gene family evolution across a phylogeny (Figure 3A) based on 36 517 orthogroups. This identified 8784 families present at the root, with 417 showing significant retractions or expansions ( $P < 0.05$ ; Figure 3A; Tables S11–S16 in the supplemental information online). Several gene families were retracted in herbaceous descendants compared with woody ancestors, including families involved in bacterial resistance and cell wall biosynthesis. Only one family, a group of leucine-rich repeat protein kinases (Figure 3C), was significantly retracted across all analyses (statistical, machine learning, and phylogeny-based approaches; Table S10 in the supplemental information online).

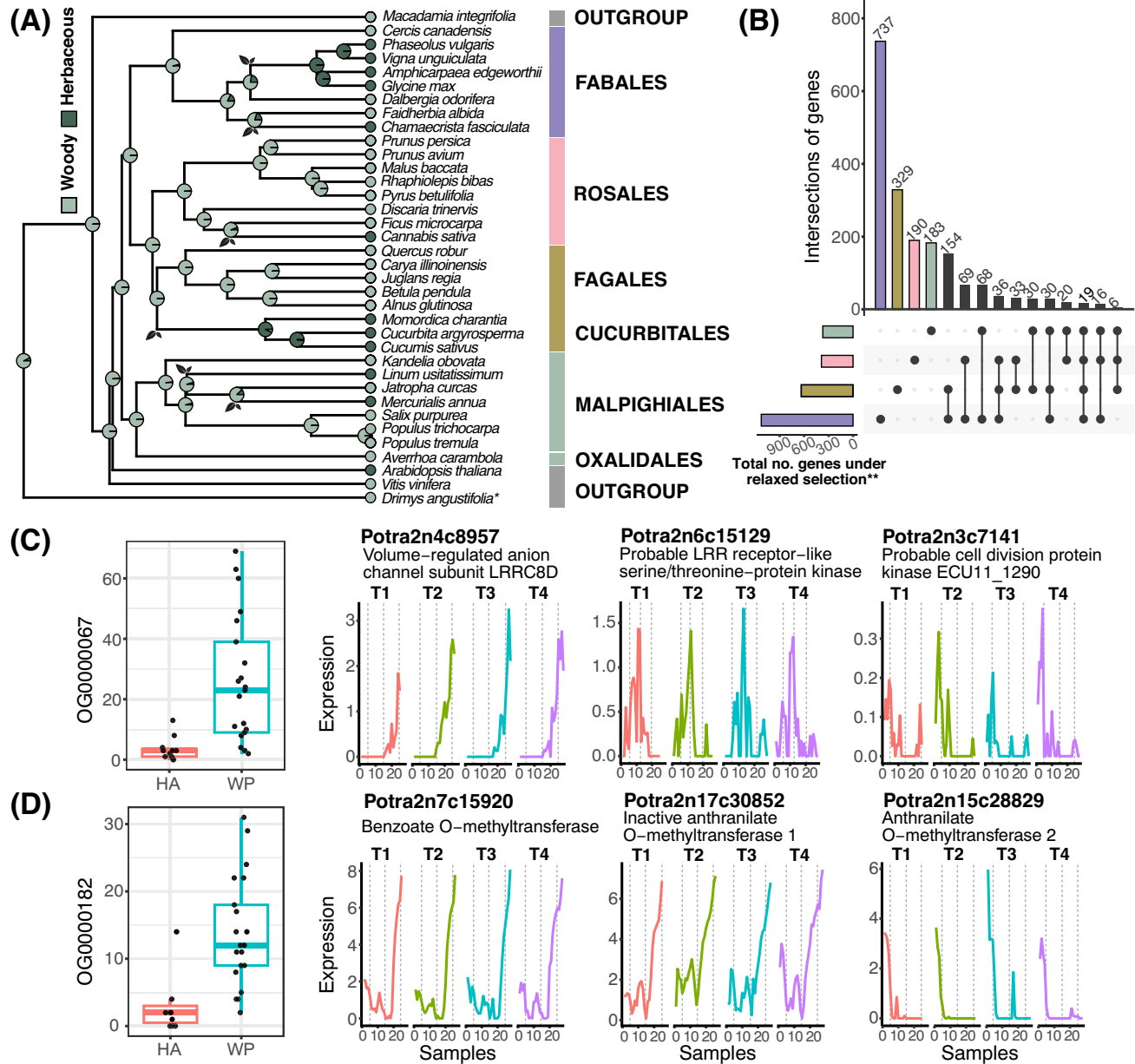


Trends in Plant Science

**Figure 2. Possible ecological drivers and underlying genetic changes of growth form evolution.** Examples of ecological drivers associated with growth form shifts are partly based on [8,9], followed by six potential types of underlying genetic change (A–F). Yellow circles (E, F) highlight the two types of genetic change supported by our data; examples from our analyses are shown in the surrounding outer circles. \*Although not discussed in detail in this opinion paper, gene family expansions were also analyzed (see supplemental information online). Stars indicate types of nonsynonymous substitutions with brown representing mutations under relaxed purifying selection and green showing mutations under positive selection. (A) A gene evolves a new function due to a new beneficial mutation. (B) An existing gene is co-opted for a new role (co-option or exaptation). (C) A noncoding mutation causes a regulatory change (e.g., altered transcription timing). (D) A gene is duplicated. One copy may evolve a new function (neofunctionalization) or both divide the original function (subfunctionalization), potentially leading to gene family expansion. (E) A gene under purifying selection accumulates nonsynonymous mutations due to relaxed constraint, potentially reducing or abolishing its function (pseudogenization). (F) A gene is deleted, often following functional erosion. This may lead to gene family contraction. Note that gene family expansion or contraction typically reflects multiple duplication or deletion events rather than single changes.

### Genes and gene families with important roles in flowering time and development

Several genes found to be under relaxed selection in all transitions to herbaceousness were tied to flowering time and development (Table 1). *PHOTOPERIOD-INDEPENDENT EARLY*



**Figure 3. Comparative genomic analyses of growth form and woodiness within the rosids.** (A) Ancestral state reconstruction of woodiness within the fabids done with ace in the R package ape [71,72] (supplemental information online). Pie charts indicate the scaled likelihoods of each ancestral state with herbaceous annuals in dark green and woody perennials in light green. Likely transitions from woodiness to herbaceousness are indicated with black leaves and used as foreground branches in the downstream selection analyses. A single star (\*) indicate that data were generated *de novo* for this study. (B) UpSet plot of genes found to be under relaxed selection in transitions from woodiness to herbaceousness (intersections based on orthogroup membership). All intersections were larger than expected by chance (Table S5 in the supplemental information online). The plot in the left corner shows total numbers of genes under relaxed selection, and the main plot shows the number of unique genes per orthogroup, followed by gene intersections between orthogroups (connected dots). Double stars (\*\*) indicate that genes can belong to the same orthogroup. (C) Gene family OG0000067 was found to have significantly fewer gene copy numbers in herbaceous annuals (HAs) than woody perennials (WPs) on the basis of statistical testing and random forest. Plot of copy number differences is followed by expression profiles from three genes within this family across wood-forming tissues in wild-growing aspen (*P. tremula*). Sample number corresponds to tissue type with ~1–5 = phloem, ~3–7 = dividing cambium cells (transition zone), ~6–12 = expanding xylem, and ~13–28 = lignified xylem. Four natural clonal replicates (T1–T4) of a single genotype in each plot [48]. (D) Gene family OG0000182 was found to have significantly fewer gene copy numbers in HAs than WPs in all analyses. Plot of copy number differences is followed by expression profiles from three genes within this family across wood-forming tissues in wild-growing aspen (*P. tremula*) as in (C).

Table 1. Examples of genes and gene families that may contribute to growth form evolution in angiosperms (woody perennial → herbaceous annual)

Gene/family	Name	Analysis	Description	Universality [1–3] <sup>a</sup>
Flowering time, developmental transitions, and lifespan				
Gene	<i>PIE1</i>	RELAX	Part of the SWR1 complex which regulates flowering and plant development	1
Gene	<i>ADAP</i>	RELAX	Positive regulator of ABA response; involved in germination and seedling growth in <i>Arabidopsis</i>	1
Gene	<i>HSI2-L1</i> (VP1/ <i>ABI3-LIKE 2</i> )	RELAX	Transcriptional repressor involved in the recruitment of PRC2 for genome-wide polycomb silencing	1
Gene	<i>MYB88</i>	RELAX	Transcription factor, which together with <i>FOUR LIPS</i> , is widely expressed throughout reproductive and vegetative development in <i>Arabidopsis</i> [5]	1
Regulation of secondary growth				
Gene	<i>LBD1</i>	RELAX	Regulates secondary growth in <i>Populus</i> ; downregulates <i>KNAT1</i>	2
Gene	<i>KNAT1</i>	RELAX	Transcription factor; promotes differentiation of cambial derivatives into xylem fibers	3
Gene	<i>WOX4</i>	RELAX	Part of the TDIF-TDR-WOX4 signaling pathway; regulates the maintenance and activity of the vascular cambium	3
Regulation and biosynthesis of secondary cell walls				
Gene	<i>VND4/NAC 007</i>	RELAX	Transcription factor; over-expression causes ectopic secondary cell wall growth in <i>Arabidopsis</i>	2
Gene	<i>IRX10</i>	RELAX	Involved in xylan biosynthesis	2
Gene	<i>CESA1</i>	RELAX	Involved in cellulose biosynthesis	2
Gene	<i>4-CL2</i>	RELAX	Involved in lignin biosynthesis	1
Gene	<i>COMT1</i>	RELAX	Involved in lignin biosynthesis	3
Family	S-adenosyl-L-methionine dependent O-methyltransferases (OG0000182)	Phylogeny-independent	Xylem-specific expression patterns	2
Family	Expansins (OG0001268)	Phylogeny-independent	Aid in loosening the cell wall during cell growth	2
Family	DUF247 transmembrane proteins ~clade 2 (OG0000881)	Phylogeny-independent & phylogeny informed	Proposed function in cell wall biosynthesis	2
Family	Lignin-polymerizing peroxidases e.g., <i>PRX52</i> (OG0000082)	Phylogeny informed	Involved in lignin biosynthesis	3
Family	Lignin-polymerizing laccases e.g., <i>LAC4</i> , <i>LAC11</i> , <i>LAC17</i> (OG0000021)	Phylogeny informed	Involved in lignin biosynthesis	3

Table 1. (continued)

Gene/family	Name	Analysis	Description	Universality [1–3] <sup>a</sup>
Disease resistance and abiotic stress responses				
Family	LRR-RLKs; leucine-rich repeat Protein kinases e.g., <i>FLS2</i> ; <i>EFR</i> , <i>GSO2</i> , <i>RLP38</i> (OG0000067)	Phylogeny-independent & phylogeny informed	Involved in innate immunity	1
Family	NB-LLRs; disease resistance genes in the ETI-pathway (OG0000551)	Phylogeny informed	Involved in resistance to bacteria	3
Family	Cytochrome P450s ~family 82, subfamily C (OG0000047)	Phylogeny-independent & phylogeny informed	Involved in abiotic and biotic defense	2
Family	bHLH-transcription factors including <i>IBH1</i> <sup>b</sup> (OG00006126)	Phylogeny-independent	<i>IBH1</i> improves salt and stress tolerance; also involved in cell elongation and plant development	2
Family	bHLH-transcription factors including <i>ICE1</i> <sup>b</sup> (OG0001234)	Phylogeny-independent	<i>ICE1</i> is involved in cold response and ABA-signaling during seed germination	2

<sup>a</sup>Universality is scored on a scale from 1 to 3. For genes, a score of 1 indicates that the gene was found to be relaxed in all transitions to herbaceousness, 2 indicates relaxation in three out of four transitions, and 3 indicates relaxation in two out of four transitions. For gene families, a score of 1 indicates that the family was significantly retracted in herbs across all analyses of copy number variation (both phylogeny-independent and phylogeny-aware), 2 indicates retraction in two analyses, and 3 indicates retraction in one analysis. In some cases, OrthoFinder grouped genes into subfamilies instead of entire gene families.

<sup>b</sup>Different subfamilies.

*FLOWERING 1 (PIE1)*, a chromatin-remodeling protein, affects multiple flowering pathways [31]. In *A. thaliana*, mutations in *PIE1* cause weak apical dominance and early flowering through reduction in expression of *FLOWERING LOCUS C (FLC)* [32]. The transcriptional repressor *HSI2-LIKE 1 (HSI2-L1)*; also named *VIVIPAROUS1* recruits Polycomb repressive complex 2 (PRC2) for genome-wide silencing [31]. *HSI2-L1* is involved in seed maturation [33,34] and vernalization-mediated *FLC* silencing in *A. thaliana* [35]. Other genes under relaxed selection in all transitions include transcription factors *ARIA-INTERACTING DOUBLE AP2 DOMAIN PROTEIN (ADAP)* and *MYB DOMAIN PROTEIN 88 (MYB88)*. *ADAP* enhances abscisic acid (ABA) response, and mutant lines germinate more efficiently and grow faster [31,36]. *MYB88* and its paralogue *FOUR LIPS (FLP)* are expressed during multiple developmental stages from embryo to flowering [37] and influence resistance to salt and drought stress [38]. Although it remains unclear whether these genes influence the evolution of annuality *per se* or generally contribute to the evolution of herbaceousness and a shortened life cycle, they highlight the role of flowering time genes, epigenetic changes, and transcriptional regulators in the evolution of new growth forms and life history strategies.

### Genes and gene families linked to secondary growth and secondary cell wall formation

Key genes and gene families involved in wood development and secondary cell wall formation showed relaxed selection in transitions to herbaceousness (Table 1). These included *LBD1*, which regulates secondary growth in *Populus* [21] and its target *KNOTTED-LIKE FROM ARABIDOPSIS THALIANA (KNAT1)*, promoting xylem fiber differentiation [21,39]. *WOX4*, essential for cambial cell maintenance and division [20], also experienced relaxed selection. Genes involved

in secondary cell wall formation included *VASCULAR RELATED NAC-DOMAIN PROTEIN 4* (*VND4/NAC 007*), a transcription factor inducing of secondary cell wall formation [40]; *IRREGULAR XYLEM 10* (*IRX10*), involved in xylan biosynthesis [41]; and *cellulose synthase 1* (*CESA1*), involved in cellulose biosynthesis and tension wood formation [42]. Interestingly, *VND4* appears to have been lost in monocots, whereas both *IRX10* and *CESA6/9* have been lost in the seagrass *Zostera marina* [43]. Gene families tied to secondary growth and cell wall processes were also retracted in herbaceous plants, and included *S*-adenosyl-L-methionine dependent *O*-methyltransferases proteins with xylem-specific expression patterns [44,45], expansins facilitating cell wall loosening [46], and DUF247s involved in cell wall biosynthesis [47]. Because many of these gene families lacked detailed functional annotations, we confirmed that several, including the *O*-methyltransferases, were expressed in wood-forming tissues on the basis of *P. tremula* expression data [48] (Figure 3D; Supplemental materials S2).

Lignin-related genes, important for secondary cell wall structure and mechanical support in woody plants [49,50], also showed relaxed selection and gene copy number reductions. Both *Caffeate O-methyltransferase 1* (*COMT1*) and *4-coumarate:CoA ligase 2* (*4CL2*) were under relaxed selection in herbaceous species, whereas a family of lignin-polymerizing peroxidases related to *PEROXIDASE 52* (*PRX52*) [31] were retracted. Additionally, a family of lignin-polymerizing laccases (such as *LAC4*, *LAC11*, *LAC17*) [51] showed variable expansions and retractions in herbaceous species across the phylogeny, potentially affecting lignin composition. Lignin biosynthesis is a critical innovation that allowed plants to conquer terrestrial habitats [52,53], which underscores the evolutionary significance of these genes.

### Genes linked to abiotic stress response and disease resistance

The single gene family found to be significantly retracted in herbaceous plants in all phylogeny-independent and informed analyses included *FLAGELLIN-SENSITIVE 2* (*FLS2*) [54], *EF-TU RECEPTOR* (*EFR*) [55], and *GASSHO2* (*GSO2*) [56], which all play roles in disease resistance (Table 1). Genes in this family were also expressed in wood-forming tissues on the basis of expression data from *P. tremula* (Figure 3C) [48]. Phylogenetic analysis also revealed retraction of disease resistance genes in the ETI pathway [57] and a family of cytochrome P450s (~family 82, subfamily C; Table 1). These contractions may reflect the shorter lifespans and lower defense demands of herbs compared with long-lived trees, which require more diverse mechanisms against a wider range of threats [58]. Additionally, several gene families linked to abiotic stress responses, including families of basic helix-loop-helix (bHLH) transcription factors, were retraced in herbs (Table 1). These included orthologs of *INDUCER OF CBF EXPRESSION 1* (*ICE1*) [59], involved in cold response, and *INCREASED LEAF INCLINATION1 BINDING bHLH1*, *IBH1*, whose upregulation improves drought and salt tolerance [60]. Interestingly, the bHLH-transcription factor *ICE2* is highly expressed in developing spruce stems and upregulated in the cambium and radial expansion zones, supporting a role in tracheid expansion and wood formation [61]. Together, these findings suggest that trees face greater biotic and abiotic stress diversity across their extended lifespans.

### Concluding remarks and future perspectives

In light of the twentieth anniversary of Andrew Groover's landmark paper on what genes make a tree a tree, we performed comparative analyses with new genomic data to provide an updated perspective on the genetic differences between trees and herbs. We identified many genes with potentially large phenotypic effects related to lifespan, flowering time, and secondary growth. Several were transcription factors or chromatin-remodeling genes, suggesting parallel changes in the regulatory landscape of annual herbs. Mutations in transcription factors often affect multiple target genes and therefore have more severe pleiotropic effects than *cis*-regulatory changes

### Outstanding questions

Are regulatory or coding changes more important for growth form evolution in angiosperms?

What genetic differences define intermediate growth forms between herbaceous annuals and woody perennials, and what can these forms tell us about how plants evolve different life history strategies?

Why do angiosperms display such a wide variation of growth forms compared with gymnosperms, and is this variation linked to their evolutionary success?

What genetic changes lead to insular woodiness, and are there cases of convergent evolution?

To what degree do the developmental networks orchestrating secondary growth overlap in gymnosperms and angiosperms?

Are loss-of-function mutations important in the evolution of growth forms?

that typically affect single gene targets [62]. Candidate genes involved in gene regulation support Andrew Groover's insight that the main differences between trees and herbs are found in regulatory patterns. Unfortunately, a more detailed comparative analysis of regulatory regions was not feasible in this study because of the difficulty of reliably aligning noncoding sequences, which is often complicated by rapid sequence divergence, structural variation, and inconsistencies in genome assembly quality [63]. However, methods for aligning regulatory regions are improving [64,65], and comparative regulatory studies can also be performed using other types of genomic data, such as RNA-sequencing, chromatin accessibility data, and transcription factor binding maps [48,66,67]. *Cis*-regulatory regions may play a key role in the evolution of plant development because they allow flexible changes in when, where, and how much a gene is expressed without altering the gene's protein product [68–70]. Comparative regulomics in trees and herbs is therefore a promising research avenue for the future.

In addition to regulatory divergence, our results indicate that gene loss processes, such as observed through relaxed selection and gene family retractions, may also underlie differences between woody and herbaceous plants. In particular, transitions to herbaceous growth are associated with relaxed selection or loss of genes involved in secondary growth and secondary cell wall formation, abiotic stress response, and disease resistance. One possibility is that life cycle acceleration, facilitated by loss-of-function mutations in genes such as *PIE1* and *ADAP*, reduces the need for sustained structural investment and long-term defense. In this context, genes essential to wood formation and stress tolerance in trees may become functionally redundant and accumulate mutations without fitness costs. Furthermore, gene loss may lead to simplification of gene regulatory networks, streamlining the genome for a shorter, faster life cycle. Although we cannot be certain whether our highlighted genes are causal for growth form evolution or have simply experienced a relaxation of selective pressures as a secondary consequence of herbaceousness and annuality, we argue that (i) the scale of the potential phenotypic effects of these genes, combined with (ii) the frequency with which these genes have experienced relaxed selection during herbaceous transitions, suggests a causal role in growth form evolution. We note, however, several limiting factors in our approach. (i) We did not investigate patterns of positive selection in herbaceous species (Figure 2C), because we assumed woodiness to be the ancestral form and we expected this type of analysis to be less informative in regard to important 'tree genes.' This leaves open the possibility that some 'tree genes' may have evolved new functions in herbaceous plants, which may be critical to understanding herbaceousness and annuality (Figure 2B). (ii) Gene duplications may have allowed one copy to experience relaxed selection, whereas another retained its original function (Figure 2B). Nevertheless, our focus on orthologs with the smallest genetic distance should make our estimates conservative. (iii) Our interpretations rely heavily on gene annotations from the herbaceous model plant *Arabidopsis thaliana*, meaning some tree-specific genes with unknown functions may have been overlooked, although these are included in our supplemental information. Functional validation of our highlighted genes would further illuminate their roles in growth form evolution.

Our study provides a starting point for comparative genomic studies of trees and herbs. The results also highlight potential hotspots for growth form evolution, which could inform future research in crop genetics and breeding. For instance, *PIE1* and *ADAP* may act as evolutionary switches promoting a faster life cycle, relevant to many herbaceous crops such as cucurbits and legumes, whereas *LBD1*, *KNAT1*, and *WOX4* could offer targets for modifying wood development in perennial or woody species. Future research on the genetic basis of tree evolution could explore additional types of genetic change (Figure 2), examine intermediate growth forms between herbaceous annuals and woody perennials, compare gene expression patterns across gymnosperms and angiosperms, and delve into the genetics of re-evolved versus ancestral

woodiness. The latter was not possible in our present study because of the limited availability of genome assemblies from species that have re-evolved woodiness. A promising approach would be to investigate multiple recent transitions from herbaceousness to woodiness in island ecosystems. What is certain is that in the next 20 years, new genomic data and techniques will complement experimental studies and move us closer to understanding what makes a tree a tree in a new era of tree evolutionary genomics (see also [Outstanding questions](#)).

### Acknowledgments

We extend our gratitude to Andrew T. Groover for his inspiring perspective, which motivated us to revisit this topic 20 years later. We also thank Giancarlo Pasquali for his assistance. This article was greatly improved by the feedback of two anonymous reviewers. Computational analyses were performed on the Saga Cluster owned by Uninet/Sigma2. Some vector graphics were created using resources designed by Freepik ([www.freepik.com](http://www.freepik.com)). This work was supported by the Research Council of Norway through the EVOTREE project (RCN 287465), awarded to T.R.H. and by FORMAS through project FR-2017/0009 awarded to M.R.G-G.

### Declaration of interests

The authors have no interests to declare.

### Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve readability and language but never to generate text *de novo*. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

### Supplemental information

Supplemental information associated with this article can be found at <https://doi.org/10.1016/j.tplants.2025.09.006>.

### References

- Groover, A.T. (2005) What genes make a tree a tree? *Trends Plant Sci.* 10, 210–214
- Friedman, J. (2020) The evolution of annual and perennial plant life histories: ecological correlates and genetic mechanisms. *Annu. Rev. Ecol. Syst.* 51, 461–481
- Groover, A. and Cronk, O. (2017) *Comparative and Evolutionary Genomics of Angiosperm Trees*, Springer
- Stebbins, G.L. (1950) *Variation and Evolution in Plants*, Columbia University Press
- Soltis, D.E. *et al.* (2013) Phylogenetic relationships and character evolution analysis of *Saxifragales* using a supermatrix approach. *Am. J. Bot.* 100, 916–929
- Smith, S.A. *et al.* (2010) An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 107, 5897–5902
- Luo, A. *et al.* (2023) Spatio-temporal patterns in the woodiness of flowering plants. *Glob. Ecol. Biogeogr.* 32, 384–396
- Klimeš, A. *et al.* (2022) The ecological drivers of growth form evolution in flowering plants. *J. Ecol.* 110, 1525–1536
- Zizka, A. *et al.* (2022) The evolution of insular woodiness. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2208629119
- Tuskan, G.A. *et al.* (2006) The genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604
- Salguero-Gómez, R. *et al.* (2016) Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proc. Natl. Acad. Sci. U. S. A.* 113, 230–235
- Li, Z. *et al.* (2022) Towards understanding the biological foundations of perennality. *Trends Plant Sci.* 27, 56–68
- Bergonzi, S. and Albani, M.C. (2011) Reproductive competence from an annual and a perennial perspective. *J. Exp. Bot.* 62, 4415–4422
- Melzer, S. *et al.* (2008) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nat. Genet.* 40, 1489–1492
- Singh, R.K. *et al.* (2017) Photoperiod- and temperature-mediated control of phenology in trees – a molecular perspective. *New Phytol.* 213, 511–524
- Osakabe, Y. *et al.* (2012) Responses to environmental stresses in woody plants: key to survive and longevity. *J. Plant Res.* 125, 1–10
- Judd, W.S. (2008) *Plant Systematics: A Phylogenetic Approach*, Sinauer Associates
- Zhang, J. *et al.* (2014) The formation of wood and its control. *Curr. Opin. Plant Biol.* 17, 56–63
- Zhong, R. *et al.* (2019) Secondary cell wall biosynthesis. *New Phytol.* 221, 1703–1723
- Kucukoglu, M. *et al.* (2017) WUSCHEL-RELATED HOMEBOX4 (WOX4)-like genes regulate cambial cell division activity and secondary growth in *Populus* trees. *New Phytol.* 215, 642–657
- Yordanov, Y.S. *et al.* (2010) Members of the LATERAL ORGAN BOUNDARIES DOMAIN transcription factor family are involved in the regulation of secondary growth in *Populus*. *Plant Cell* 22, 3662–3677
- Turley, E.K. and Etchells, J.P. (2022) Laying it on thick: a study in secondary growth. *J. Exp. Bot.* 73, 665–679
- Ohtani, M. *et al.* (2011) A NAC domain protein family contributing to the regulation of wood formation in poplar. *Plant J.* 67, 499–512
- Bennett, M.D. and Leitch, I.J. (2005) Genome size evolution in plants. In *The Evolution of the Genome* (Gregory, T.R., ed.), pp. 89–162, Elsevier Academic Press
- Bennett, M.D. (1972) Nuclear DNA content and minimum generation time in herbaceous plants. *Proc. R. Soc. Lond. B Biol. Sci.* 181, 109–135
- Beaulieu, J.M. *et al.* (2010) On the tempo of genome size evolution in angiosperms. *J. Bot.* 2010, 1–8
- Morse, A.M. *et al.* (2009) Evolution of genome size and complexity in *Pinus*. *PLoS One* 4, e4332
- Wang, H. *et al.* (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U. S. A.* 106, 3853–3858
- Wertheim, J.O. *et al.* (2015) RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32, 820–832
- Mendes, F.K. *et al.* (2021) CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518

31. TAIR (2024) The Arabidopsis Information Resource. <https://www.arabidopsis.org/>, Accessed date: 8 October 2024
32. Choi, K. *et al.* (2007) Arabidopsis homologs of components of the SWR1 complex regulate flowering and plant development. *Development* 134, 1931–1941
33. Veerappan, V. *et al.* (2012) A novel HSI2 mutation in Arabidopsis affects the PHD-like domain and leads to derepression of seed-specific gene expression. *Planta* 236, 1–17
34. Jia, H. *et al.* (2014) Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. *Wiley Interdiscip. Rev. Dev. Biol.* 3, 135–145
35. Yuan, W. *et al.* (2016) A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in Arabidopsis. *Nat. Genet.* 48, 1527–1534
36. Lee, S.-J. *et al.* (2009) An ARIA-interacting AP2 domain protein is a novel component of ABA signaling. *Mol. Cells* 27, 409–416
37. Lei, Q. *et al.* (2015) The FOUR LIPS and MYB88 transcription factor genes are widely expressed in *Arabidopsis thaliana* during development. *Am. J. Bot.* 102, 1521–1528
38. Xie, Z. *et al.* (2010) Role of the stomatal development regulators FLP/MYB88 in abiotic stress responses. *Plant J.* 64, 731–739
39. Liebsch, D. *et al.* (2014) Class I KNOX transcription factors promote differentiation of cambial derivatives into xylem fibers in the Arabidopsis hypocotyl. *Development* 141, 4311–4319
40. Zhou, J. *et al.* (2014) Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. *PLoS One* 9, e105726
41. Jensen, J.K. *et al.* (2014) Arabidopsis thaliana IRX10 and two related proteins from psyllium and *Physcomitrella patens* are xylan xylosyltransferases. *Plant J.* 80, 207–215
42. Andersson-Gunnerås, S. *et al.* (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J.* 45, 144–165
43. Roodt, D. *et al.* (2019) Loss of wood formation genes in monocot genomes. *Genome Biol. Evol.* 11, 1986–1996
44. Barakat, A. *et al.* (2011) Comparative genomics and evolutionary analyses of the O-methyltransferase gene family in *Populus*. *Gene* 479, 37–46
45. Sundell, D. *et al.* (2015) The plant genome integrative explorer resource: PlantGenIE.Org. *New Phytol.* 208, 1149–1156
46. Cosgrove, D.J. (2015) Plant expansins: diversity and interactions with plant cell walls. *Curr. Opin. Plant Biol.* 25, 162–172
47. Wannitukul, P. *et al.* (2023) Disruption of a DUF247 containing protein alters cell wall polysaccharides and reduces growth in Arabidopsis. *Plants* 12, 1977
48. Sundell, D. *et al.* (2017) AspWood: high-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in *Populus tremula*. *Plant Cell* 29, 1585–1604
49. Campbell, M.M. and Sederoff, R.R. (1996) Variation in lignin content and composition (mechanisms of control and implications for the genetic improvement of plants). *Plant Physiol.* 110, 3–13
50. Sarkanen, K.V. and Ludwig, C.H. (1971) *Lignins: Occurrence, Formation, Structure and Reactions*, Wiley-Interscience
51. Zhao, Q. *et al.* (2013) LACCASE is necessary and nonredundant with PEROXIDASE for lignin polymerization during vascular development in Arabidopsis. *Plant Cell* 25, 3976–3987
52. Weng, J.-K. and Chapple, C. (2010) The origin and evolution of lignin biosynthesis. *New Phytol.* 187, 273–285
53. Niklas, K.J. *et al.* (2017) The evolution of hydrophobic cell wall biopolymers: from algae to angiosperms. *J. Exp. Bot.* 68, 5261–5269
54. Peck, S.C. *et al.* (2001) Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell* 13, 1467–1475
55. Zipfel, C. *et al.* (2006) Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell* 125, 749–760
56. Okuda, S. *et al.* (2020) Molecular mechanism for the recognition of sequence-divergent CIF peptides by the plant receptor kinases GSO1/SGN3 and GSO2. *Proc. Natl. Acad. Sci. U. S. A.* 117, 2693–2703
57. Nepal, M.P. *et al.* (2017) Comparative genomics of non-TNL disease resistance genes from six plant species. *Genes* 8, 249
58. Carlsson-Granér, U. and Thrall, P.H. (2006) The impact of host longevity on disease transmission: host–pathogen dynamics and the evolution of resistance. *Evol. Ecol. Res.* 8, 659–675
59. Chinnusamy, V. *et al.* (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes Dev.* 17, 1043–1054
60. Moreno, J.E. *et al.* (2018) The antagonistic basic helix-loop-helix partners BEE and IBH1 contribute to control plant tolerance to abiotic stress. *Plant Sci.* 271, 143–150
61. Baison, J. *et al.* (2020) Genetic control of tracheid properties in Norway spruce wood. *Sci. Rep.* 10, 18089
62. Jones, D.M. and Vandepoele, K. (2020) Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Curr. Opin. Plant Biol.* 54, 42–48
63. Reineke, A.R. *et al.* (2011) Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes. *Nucleic Acids Res.* 39, 6029–6043
64. Song, B. *et al.* (2021) Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize. *Genome Res.* 31, 1245–1257
65. Xu, Y. *et al.* (2023) Evolutionary analysis of conserved non-coding elements subsequent to whole-genome duplication in opium poppy. *Plant J.* 116, 1804–1824
66. Yocca, A.E. *et al.* (2021) Evolution of conserved noncoding sequences in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 38, 2692–2703
67. Galli, M. *et al.* (2025) Transcription factor binding divergence drives transcriptional and phenotypic variation in maize. *Nat. Plants* 11, 1205–1219
68. Carroll, S.B. (2000) Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101, 577–580
69. Schmitz, R.J. *et al.* (2022) Cis-regulatory sequences in plants: their importance, discovery, and future challenges. *Plant Cell* 34, 718–741
70. Marand, A.P. *et al.* (2023) Cis-regulatory elements in plant development, adaptation, and evolution. *Annu. Rev. Plant Biol.* 74, 111–137
71. Paradis, E. *et al.* (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289–290
72. Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528