# Ectopic cambia in Japanese wisteria (*Wisteria floribunda*) vines are associated with the expression of conserved *KNOX* genes

Israel L. Cunha Neto[1]*, Anthony A. Snead[2], Jacob B. Landis[3], Chelsea D. Specht[3], and Joyce G. Onyenedum[1]*

[1]Department of Environmental Studies, New York University, New York 10012, NY, USA
[2] Department of Biology, New York University, New York 10012, NY, USA
[3]School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca NY, USA

*Corresponding author: israellopescn@gmail.com; jgo5750@nyu.edu

**Abstract**
Secondary growth is a conserved mechanism that gives rise to vascular tissues produced via a single vascular cambium. Molecular mechanisms underlying this process are characterized mostly in model species bearing typical vascular architecture, while the genetics underlying ecologically-important atypical vascular architectures remain unexplored. We use developmental anatomy, comparative transcriptomics, and molecular evolutionary analyses to address this knowledge gap, investigating how multiple ectopic cambia (EC) form in the woody vine Japanese wisteria. Anatomical studies show EC in Japanese wisteria arise from cortical parenchyma, while cambium-specific transcriptome comparisons reveal that genes acting as regulators of typical cambium development in model species are likewise associated with atypical EC development. Gene trees of KNOX proteins indicate duplication events may contribute to EC formation, including a Fabaceae-specific duplication of KNAT6 detected as under positive selection. These findings reveal insights into the genetics of EC formation, advancing our understanding of the development of complex vascular traits.

**Keywords**: cell differentiation, climbing plants, development, gene expression, vascular cambium, vascular variants, woody vines

## Introduction

The vast majority of woody plants undergo a common type of secondary growth that involves the production of a single meristematic niche called the vascular cambium, responsible for generating secondary xylem (wood) and secondary phloem (inner bark) and enabling the plants to expand in diameter[1]. With only a single vascular cambium, plants can live up to 5,000 years (e.g. Bristlecone pines [*Pinus longaeva* D.K.Bailey] or build a sophisticated plumbing system to transport water upwards of 380 ft (e.g. Coastal Redwood [*Sequoia sempervirens* (D.Don) Endl.][2,3]. However, some plants deviate from the canonical modality to create woody structures with unique growth patterns that result in novel anatomical formations, such as those found within the stems of woody vines or “lianas”. Lianas reach the forest canopy by climbing over other plants rather than supporting their weight[4]. As they weave their way through dense forest understory canopy, the stems of lianas are frequently damaged either through the collapse of host trees, or the regular wear and tear of twisting and twining[5]. The woody stems of many lianas present with “ectopic cambia”, a developmental aberration that may be a mechanism for lianas to repair their stems after injury while maintaining key vascular functions of transportation and support[5]. Ectopic cambia (EC) result from multiple vascular cambia developing in atypical locations, secondary to the first-formed and central vascular cambium[6]. In addition to their role in lianas, EC may also facilitate the success of self-supporting herbs, shrubs, and trees living in harsh conditions such as mangroves and arid environments[7,8]. EC have been identified in the stems of the temperate *Wisteria* vines[9,10], potentially facilitating their success as invasive vines in the Eastern United States[11].

To date, all anatomical studies show that EC arise through three steps: (1) de-differentiation of living parenchyma cells, (2) cell division to give rise to a new cambium, and (3) production and differentiation of localized vascular tissues from the new cambium layer[7,9,12]. While the vast majority of the literature covering the natural diversity of EC is purely anatomical and descriptive, transgenic phenotypes in model species suggest that ECs can emerge by simply manipulating the expression of conserved genes responsible for vascular development[6].

Recent investigations demonstrate that *Arabidopsis thaliana* (L.) Heynh and *Populus* spp. share a suite of genes in the regulatory pathway controlling the canonical model of secondary growth by a single vascular cambium[1,13,14]. Notably, transgenic lines that perturb the expression of conserved vascular genes produce EC in stems and/or roots[6]. Various studies indicate that EC can appear as the result of the misexpression of members of the class III homeodomain-leucine zipper (HD-ZIP III) transcription factor genes: for example, EC are developed in *Populus* stems through overexpression of *PRE*, an ortholog of the gene *REVOLUTA (REV)*[15], while EC have been shown to develop in *A. thaliana* through the overexpression of *AtHB7*[16] and repression of *AtHB4*[17]. ECs can also be formed following overexpression of *WOX4*, a member of the WOX gene family[18], and EC were observed in double mutants following the overexpression of *WOX4* and repression of *PETAL LOSS* (*PTL*)[19]. Finally, regenerated cambium can be induced by girdling trees, allowing investigators to characterize how such cambia (here categorized as EC given their de novo establishment) can function to stitch together a wounded vascular cambium[20,21]. KNOTTED1-like HOMEOBOX transcription factors—key regulators of vascular proliferation—often emerge as differentially expressed genes in cambium development studies[16,17,19,22]. Yet, their role is underexplored in the formation of EC development. Our current knowledge of EC formation arises from molecular manipulation or direct perturbation—either mechanical (i.e., girdling) or genetic (i.e., transgenics)—in model species that don't normally form EC. To understand how EC contribute to plant diversity in form and function, we must understand how EC are produced in systems in which they naturally occur, particularly in lianas where they have a key functional role in allowing plants to climb.

In this work, we seek to elucidate the developmental anatomy, transcriptomic profile, and molecular evolution underlying EC formation in a pair of vines in the Fabaceae family, the ectopic cambia-producing Japanese wisteria (*Wisteria floribunda* (Willd.) DC.) and the non-ectopic cambia producing Common bean (*Phaseolus vulgaris* L.). We confirm previous reports that ectopic cambia form within the cortex of Japanese wisteria in a haphazard fashion. To reveal the transcriptomic profile underlying EC formation, we leveraged tangential cryosectioning to isolate tissues of interest for RNA sequencing. Here we report the differential expression of several conserved genes in species and cross-species analyses of cambium-specific samples, including known cambium markers and genes involved in epigenetic regulation. We identified *KNOX* genes as prospective candidates in EC development in Japanese wisteria, as multiple transcript clusters are differentially expressed in typical v. ectopic cambia contrasts. We adopted a phylogenetic approach to investigate the molecular evolution of *KNOX* genes across 45 species spanning gymnosperms and angiosperms. We implemented branch models to investigate the ratio of nonsynonymous to synonymous substitutions and assess the signatures of selection during the evolution and diversification of the *KNOX* protein family. Our analyses uncovered transcript clusters for *KNOX* proteins under positive selection, further corroborating their important role in EC development in Japanese wisteria. Together, these results expand our understanding of both typical and atypical cambium development, while identifying promising contenders underpinning EC formation in woody plants.

## Results

### *Common bean and Japanese wisteria exhibit typical cambium development; Japanese wisteria develops ectopic cambia during the later stages of stem growth*

To characterize the origin of ectopic cambia (EC), we began by detailing the developmental anatomy of the Common bean and Japanese wisteria (Fig. 1a, e). Young stems of both Common bean and Japanese wisteria have the same organization of primary tissues, with vascular bundles arranged in a ring (Fig. 1b,

f). In young stems (3rd to 5th internodes, see Fig. 1a, e), we find the initiation of the vascular cambium from the fascicular cambium within each bundle and the interfascicular cambium connecting the bundles (Fig. 1c, g), which gives rise to secondary xylem (wood) inward and secondary phloem (inner bark) outward (Fig. 1d, h). This latter stage marks the final step in stem development of Common bean (Fig. 1d) giving rise to the mature form; however, Japanese wisteria development continues with further elaborations (Fig. 1i) to arrive at maturity. In stems with at least 20mm diameter, Japanese wisteria starts to produce additional cambia ectopically (Fig. 1j), with older stems displaying multiple layers of EC and its products, i.e., secondary xylem and secondary phloem (Fig. 1i). EC in Japanese wisteria arise from the cortex (Fig. 1j) where existing cortical cells undergo localized cell divisions, creating a mass of cells from which a new meristem will arise to produce secondary xylem and phloem (Fig. 1j). These EC do not arise in a uniform concentric pattern but rather initiate asynchronously across the stem circumference forming discrete strands of secondary xylem and secondary phloem (Fig. 1i).

***Transcriptomes of Common bean and Japanese wisteria***

Having isolated key developmental stages, we sought to identify which genes underlie the formation of EC. Transcriptomic profiles were made from the typical cambium from the Common bean and Japanese wisteria and the EC of Japanese wisteria, with at least five biological replicates for each of the three tissues sampled. Sequencing resulted in an average of 21,420,218 (128,521,310 total) and 23,468,348 (258,151,832 total) paired-end reads per sample for the Common bean and Japanese wisteria, respectively. The transcriptome assemblies vary in BUSCO, mapping rate, and N50, but both assemblies resulted in BUSCO scores and mapping rates over 80% with N50s greater than 1500 base pairs (bp) (Table 1).

**Table 1. The transcriptome assembly statistics (i.e., BUSCO, Mapping Rate, and N50) for the final transcriptome assemblies for *Phaseolus vulgaris* (Common bean) and *Wisteria floribunda* (Japanese wisteria).**

| Species | BUSCO | Mapping Rate | N50 |
|---|---|---|---|
| Bean | C:81.8%[S:77.1%,D:4.7%], F:3.9%, M:14.3% | 82.41% | 1562 |
| Wisteria | C:88.4%[S:79%,D:9.4%], F:1.6%, M:10% | 84.99% | 1624 |

***Differential Cluster Expression and Gene Ontology Enrichment***

**Comparison across species.** Using the transcript clusters for Common bean (42,827) and Japanese wisteria (74,960), 10,498 one-to-one orthologous transcript clusters were identified and passed filtering (Expressed > 0.5 counts per million). The number of differentially expressed clusters varied dramatically between contrasts. The most differentially expressed (DE) transcript clusters were found in the “Japanese wisteria v. Common bean” contrast, representing overall differences between the two species (> 6,000). This was followed by the number of DE transcripts when comparing the “Typical Cambia: Japanese wisteria v. Common bean” (> 5,000). The third highest number of DE transcripts was found in the “Ectopic Cambia v. Typical Cambia of Japanese wisteria and Common bean” contrast (> 2,000) (Table 2). Lastly, we compared the “Japanese wisteria: Ectopic v. Typical cambia” contrast, which surprisingly yielded no DE transcript clusters in this analysis (see further analyses "*Comparisons within Japanese wisteria*” below). In annotating the DE transcript clusters with Trinotate[23] against the UniProt database for Fabales (UniProt Taxon ID 72025[24]), we found transcript clusters related to DNA methylation (e.g., methyltransferases, DNA methylase), auxin (e.g., indoleacetic acid, auxin response factors), cell cycle control (e.g., cyclins, elongation effective 1), cell differentiation (e.g., *KNOX*) and other cambial markers (e.g., *PXY*). The “Ectopic Cambia v. Typical Cambia of Japanese wisteria and Common bean” contrast is the only

comparison that yielded enriched gene ontology terms based on identified DE clusters, and the GO terms are mostly related to RNA binding, splicing, and processing (Table 3).

**Table 2. Significantly differentially expressed transcript clusters based on a false discovery rate of less than 0.05 for each contrast (one-to-one orthologous analysis).**

| Contrast | Direction | Genes |
|---|---|---|
| Japanese wisteria v. Common bean | Up | 3,168 |
| | Down | 3,071 |
| Typical Cambia: Japanese wisteria v. Common bean | Up | 2,832 |
| | Down | 2,594 |
| Ectopic Cambia v. Typical Cambia of Japanese wisteria and Common bean | Up | 1,636 |
| | Down | 730 |

**Table 3. The significantly enriched gene ontology (GO) terms (< 0.05 FDR) for differentially expressed transcript clusters in the "Ectopic Cambia v. Typical Cambia of Japanese wisteria and Common bean" contrast.**

| Term | Gene Ontology | Direction | p-value |
|---|---|---|---|
| mRNA processing | Biological Process | ▲ | 0.0002 |
| mRNA Splicing, via Spliceosome | Biological Process | ▲ | 0.0007 |
| RNA Splicing | Biological Process | ▲ | 0.04 |
| UV Protection | Biological Process | ▲ | 0.04 |
| Spliceosomal Complex | Cellular Component | ▲ | 0.002 |
| Nucleolus | Cellular Component | ▲ | 0.004 |
| Nucleus | Cellular Component | ▲ | 0.006 |
| RNA binding | Molecular Function | ▲ | 0.000005 |
| mRNA binding | Molecular Function | ▲ | 0.003 |
| nucleic acid binding | Molecular Function | ▲ | 0.04 |
| DNA-directed 5'-3' RNA polymerase activity | Molecular Function | ▲ | 0.05 |

Footnote: Each term includes its ontology and corrected p-value rounded to one significant figure. "Ectopic Cambia v. Typical Cambia of Japanese wisteria and Common bean" is the only contrast with enriched GO terms using the differentially expressed transcript clusters. The filled arrows indicate if the terms were significantly overrepresented (▲).

*Comparison within Japanese wisteria* – Given our focus on elucidating the genes related to ectopic cambia, we narrowed our focus on genes differentially expressed in the "Japanese wisteria: Ectopic v. Typical cambia" contrast. This approach enabled us to isolate the most likely candidate genes associated with EC formation in this species. After filtering (Expressed > 0.5 counts per million in at least two samples), 46,225 transcript clusters were retained for differential transcript cluster expression analysis across the two Japanese wisteria cambium types using dream[25], with 14 clusters (2▲ and 12▼) found to be differentially expressed. The differentially expressed clusters include transcripts related to DNA methylation, cell expansion (e.g., expansin-like B1), and cell differentiation (e.g., *KNOX*), all of which were found to be downregulated in the EC of Japanese wisteria (Fig. 2; Supplementary Table 1). Only two transcript clusters are upregulated, one of which includes the epigenetic gene Methyltransf_7 (Table 4). Interestingly, we found five differentially expressed *KNOX* genes across the four contrasts (Table 5), which

171 makes them of particular interest given their defined role in vascular cambium proliferation in model
172 systems.
173
174 **Table 4. Log fold change (logFC) of differentially expressed transcript clusters in the "Japanese**
175 **wisteria: Ectopic v. Typical cambia" using dream**.
176

| Cluster/Protein annotation | logFC |
|---|---|
| KNOX1 | -7.70 |
| EXLB1_ARATH | -7.63 |
| Cluster_W13 | -5.92 |
| Cluster_W7 | -5.76 |
| Glyco_hydro_18 | -5.22 |
| Bowman-Birk_leg | -5.04 |
| DIOX_N | -4.94 |
| Dimerisation | -4.72 |
| Mem_trans | -4.22 |
| glycine-rich cell wall structural protein 1-like* | -3.62 |
| Aminotran_1_2 | -3.33 |
| Thioredoxin | -2.48 |
| Cluster_W10 | 4.07 |
| Methyltransf_7 | 5.21 |

177
178 **Footnote:** (*) this gene was not originally annotated but retrieved through BLAST analysis after it was
179 identified as significant.
180
181 **Table 5. Log fold change (logFC) of transcript clusters annotated as *KNOX* genes and differentially**
182 **expressed across contrasts.**
183

| Cluster | Protein annotation* | Japanese wisteria v. Common bean | Typical Cambia: Japanese wisteria v. Common bean | Ectopic Cambia v. Typical Cambia of Japanese wisteria and Common bean | Japanese wisteria: Ectopic v. Typical cambia |
|---|---|---|---|---|---|
| **Cluster_W2** | KNOX1 (KNAT2/6) | NA | NA | NA | -7.70 |
| **Cluster_9937** | KNOX1 (STM) | -8.24 | -8.17 | -4.22 | NA |
| **Cluster_2082** | KNOX1 (STM) | -3.47 | -3.65 | NA | NA |
| **Cluster_4072** | KNOX1 (KNATM) | -3.33 | -3.40 | NA | NA |
| **Cluster_1332** | KNOX2 (KNAT2/6) | -6.38 | -7.69 | NA | NA |

184
185 Footnote: The asterisk (*) indicates the KNOX protein based on our phylogenetic analysis (see below).
186
187 ***Identification, characterization, and phylogenetic analysis of KNOX genes analysis***
188 To probe the relationship between *KNOX* genes and EC evolution, we began by testing the hypothesis
189 that there has been a diversification of *KNOX* genes leading to EC. Toward this aim, we broadened our
190 analyses to encompass all nine members of the *KNOX* gene family (based on the *A. thaliana* genome)
191 across 45 seed plant species, including gymnosperms, monocots, and eudicots. Our sampling included
192 species with and without EC (Supplementary Table 5-6). Although retrieved through the tBLASTx search,

193 we removed eight species given their poor alignment, thus potentially erroneous phylogenetic placement
194 (Supplementary Fig. 2).
195 The final sequence alignment contained 466 sequences (see Supplementary Dataset 1 and 2 for
196 aligned and unaligned nucleotide alignments, respectively). The maximum likelihood phylogeny grouped
197 *KNOX* genes into three main clades that correspond to defined classes (Fig. 3; Supplementary Fig. 1):
198 KNOX1, KNOX2, and KNATM (Fig. 3). In this phylogeny rooted with a representative of *BEL1* genes,
199 KNOX2 diverges first, with the subsequent divergence of KNOX1 and KNATM as sister lineages. The
200 three KNOX classes can be further divided into major clades named according to their putative *A. thaliana*
201 ortholog (Fig. 3; Supplementary Fig. 1). Class KNOX1 includes three main clades corresponding to STM,
202 KNAT1, and KNAT2/6. Class KNOX2 is subdivided into two clades corresponding to KNAT3/4/5 and
203 KNAT7.
204 Using our phylogenetic approach, we were able to determine the orthology of each of the
205 differentially expressed KNOX transcript clusters (see Table 6). All DE clusters belong to class KNOX1:
206 Two of our transcript-clusters group with *A. thaliana* orthologs of STM, one of which is nested within a
207 clade derived from a Fabaceae-specific gene duplication containing the soybean SBH1 reference; two
208 transcript-clusters group with KNAT2/6*,* one being associated with each of the major clades resulting from
209 gene duplication following the divergence of angiosperms (Supplementary Fig. 1); and one transcript-
210 cluster groups with KNATM (Fig. 3; Supplementary Fig. 1).
211 The number of KNOX gene copies recovered from each taxon varies as a result of duplication
212 events within major clades representing different KNOX proteins. We recovered duplication events in most
213 lineages with EC including at the base of the angiosperms for KNAT2/6 (indicated by sequences of
214 *Amborella trichopada* in Fig. 3), eudicots for STM*,* KNAT2/6, and KNAT3/4/5) (indicated by sequences of
215 *Vitis vinifera* in Fig. 3), as well as within the Fabaceae family for all KNOX1 and KNOX2 genes (indicated
216 by sequences of *Wisteria floribunda* or *Spatholobus suberectus* in Fig. 3). Duplication events are also
217 observed in lineages without EC, including the conifers for KNAT1 and the Poales (monocots) for KNAT1
218 and KNAT3/4/5 (indicated by sequences of *Zea mays or Triticum dicoccoides* in Fig. 3). We did not find an
219 association between the number of *KNOX* genes and the presence/absence of EC across plants (Fig. 4).
220 The *A. thaliana* genome contains nine members of the *KNOX* gene family, but lineages with EC include
221 both increases (e.g., Japanese wisteria with 16 copies and other species of Fabaceae with 17 copies) and
222 decreases (e.g., Amaranthaceae species with five copies), including the possible absence of *KNOX* genes
223 in the EC-producing gymnosperms *Cycas panzhihuaensis* and *Gnetum montanum* (Fig. 4). A Fabaceae-
224 specific duplication in KNATM after the divergence of *Cercis canadensis* is apparent in this dataset (Fig.
225 4).
226
227 ***Selection analyses***
228 To test whether KNOX genes in lineages with EC are under positive selection (indicative of adaptive
229 evolution) we compared PAML codon-based nested models (models 0 = one omega value for the tree, 1 =
230 a unique omega per branch, 2 = foreground and background omega values[26]). We tested these models in
231 individual gene trees for members of KNOX genes that were differentially expressed in the transcriptome
232 analysis (i.e., STM, KNATM, and KNAT2/6). By comparing foreground (EC present) and background
233 (absent/typical growth) branches (model 2), the only gene set that showed a significant difference was
234 KNAT2/6 (p-value = 0.00406). To detect whether some branches along KNAT2/6 lineages were under
235 positive selection, we employed branch model 1 on the KNAT2/6 phylogenetic rooted tree, which showed
236 nine branches with omega values greater than 1 ($\omega > 1$), therefore, indicating positive selection in certain
237 lineages (Supplementary Fig. 3). Out of the nine branches, only two have EC, both of which are
238 sequences differentially expressed from two transcript clusters in the gene expression analysis
239 (*Wisteria_floribunda*_Cluster_W2, $\omega = 999$; *Wisteria_floribunda*_Cluster_1332, $\omega = 958.234$). Taxa under
240 positive selection belonging to species with typical growth include *Dioscorea rotundata* ($\omega = 999$)*, Populus*
241 *trichocarpa* ($\omega = 964.749$ and 999)*,* and *Z. mays* ($\omega = 1.29156$), as well as some Fabaceae such as

242 *Phaseolus vulgaris* (ω = 999) and *Prosopis alba* (1.27446). Despite statistical tests not supporting
243 differences in synonymous v. nonsynonymous changes for the KNATM protein, results indicate only two
244 branches with higher omega values both of which are species with EC from the Fabaceae including
245 *Phaseolus lunatus* (ω = 1.53725) and *W. floribunda* (ω = 1.15616), representing the KNATM sequence
246 from the DGE analysis (Cluster_4072) (Supplementary Fig. 3). Only one branch belonging to *P.*
247 *trichocarpa* (ω = 999) is under positive selection in the rooted STM tree (Supplementary Fig. 3).
248
249 **Discussion**
250 ***KNOX* genes are likely candidates for modulating EC formation in Japanese wisteria**
251 In Japanese wisteria, ectopic cambia (EC) originate from cortical parenchyma, forming one to several
252 vascular increments that may or may not be continuous[9]. This ontogenetic origin from the cortex is unlike
253 most reported species with EC, where cambia typically form from vascular tissues, e.g., pericycle
254 parenchyma, phloem parenchyma[6]. Until recently, information on how this process occurs at the molecular
255 level was derived solely from experimental studies in *Populus* and *A. thaliana*, whereby mis-expression of
256 conserved genes related to vascular development yielded the formation of EC from the cortex or vascular
257 tissues[6]. Given these data and considering the conserved nature of vascular development across seed
258 plants, we had originally hypothesized only a few developmental pathways likely leading to the formation of
259 EC, i.e., HD-ZIP III and *WOX4* overexpression, similar to the observations from experimental mutants[6].
260 Although candidate genes were annotated in the Common bean and Japanese wisteria transcriptome,
261 they were not present in the final analysis of differentially expressed genes due to either not being
262 assigned as one-to-one orthologs across species, not being annotated in one or the other species (e.g.,
263 *ANT, AtHB8, WOX4*), not being annotated in both species (e.g., *CLE41/42/44, WOX14*), or being present
264 in both species and not differentially expressed (e.g., *REV*). However, other genes known to be associated
265 with EC formation were not annotated (e.g., *AtHB4*). These results are a combination of differences
266 associated with both biological variation and methodological limitations. Here we found evidence that
267 members of the *KNOTTED1-like homeobox (KNOX)* gene family are among the top candidates for genes
268 related to the formation of EC. Our KNOX phylogeny recovered three KNOX classes with relationships
269 among the classes similar to those recovered in previous studies[27], while the phylogenetic position of
270 subclasses (e.g., KNOX1) in some lineages (e.g., conifers) differ across different studies, including our
271 tree[28,29]. Placing our five differentially expressed KNOX transcript clusters into a phylogenetic context, we
272 show that all copies fall into either KNOX class I or KNATM. We will focus our discussion on these two
273 classes.
274 KNOX1 genes have been reported to act as major regulators of organ initiation, tissue
275 proliferation, and meristem maintenance, including the formation of vascular cambium[19,30–32]. In *Populus*,
276 the overexpression of ARK1 (ortholog of *SHOOTMERISTEMLESS* (*STM*) in *A. thaliana*) generated
277 dispersed/unequal differentiation of cambium derivatives during the secondary growth[30], while the
278 overexpression of ARK2 (ortholog of KNAT1 in *A. thaliana*) promoted a wider cambial zone[31]. In *A.*
279 *thaliana*, KNAT1 mutants exhibited reduced radial growth although meristematic activity was maintained,
280 which indicated functional redundancy with other genes (e.g., KNAT6) in the transcriptional regulation of
281 cambial activity and differentiation of vascular tissues[19,33,34]. In Japanese wisteria, we saw downregulation
282 of a transcript-cluster identified as KNAT2/6 in the "Japanese wisteria: Ectopic v. Typical cambia" contrast
283 (Table 3). Since KNAT6 is known to be expressed in the vascular cambium and vascular tissues of most
284 species[33–35], although absent in the cambial zone of some Cactaceae[36], the misexpression of these genes
285 in the "Japanese wisteria: Ectopic v. Typical cambia" contrast may suggest their involvement in EC
286 development. Other evidence of the involvement of KNAT6 in vascular oddities may be noted from the
287 higher transcript levels of KNAT6 in *knat1* and *pny* (*PENNYWISE* also known as *BEL1-LIKE*
288 *HOMEODOMAIN 9*) *A. thaliana* mutants. These plants show vascular defects ranging from increased
289 cambial activity between the phloem and xylem cells[37] to extravascular bundles that enlarge to reduce
290 interfascicular space[38]. Curiously, StPOTH1, an ortholog of KNAT6 in *Solanum tuberosum* L., may also be

291 involved in the development of potato tubers[39], which have been described as having atypical vascular
292 tissue formation[40]. Nevertheless, just as the mechanism by which KNAT6 functions in vascular
293 development remains largely unclear in model species[30,33,34,38], additional studies will be required to clarify
294 how KNAT2/6 plays a role in EC development, as well as the involvement of other KNOX proteins (e.g.,
295 STM), which is also downregulated in the contrast comparing "Ectopic Cambia v. Typical Cambia of
296 Japanese wisteria and Common bean". The functional validation of KNOX proteins in relation to vascular
297 development should address gene expression dynamics and potential interactions, which suggest that
298 they may share similar functions[41] as well as act in antagonistic relationships (e.g., KNAT1 and
299 KNAT6)[34,42], including in other systems such as leaf development[43].
300 In the cambium regulatory network of woody plants, KNOX proteins (e.g., KNAT1 and STM) are
301 downstream of the *TDIF-PXY-WOX4* signaling pathway, a conserved module responsible for cambium
302 development in woody plants[1,13,14]. These observations suggest that EC development in Japanese wisteria
303 probably derives from modifications to the *TDIF-PXY-WOX4*, perhaps leading to changes in the regulation
304 of downstream *KNOX* genes. That is also evidenced by the findings that KNAT1 was downregulated in
305 *wox4 A. thaliana* mutants with cambia formed ectopically[19] and by the fact that the overexpression of
306 *WOX4* alone resulted in EC in *Populus*[18]. In addition, *knat1 wox4* mutants exhibited severe defects in
307 cambium proliferation in *A. thaliana,* however, cambium initiation and activity were still detected, indicating
308 redundancy among cambium regulators[19]. The manipulation of these transcription factors (e.g., *WOX4*)
309 that are putative upstream regulators of KNOX1 genes (e.g., KNAT1) provides evidence of their
310 involvement in vascular development and potentially EC formation. Future studies investigating the
311 expression of these KNOX proteins in EC formation will shed more light on their role in vascular
312 development and the development and evolution of EC in natural systems.
313
314 ***Both constitutively and injury-induced ectopic cambia converge on similar molecular processes***
315 In this study, we found downregulation of KNOX proteins within the constitutively formed EC of
316 Japanese wisteria. *KNOX* genes also emerge as important regulators underpinning the ability of perturbed
317 *Populus* trees to form de novo cambia post-girdling[20] (see Supplementary Table 2 and Supplementary
318 Methods for a comparison of cambial markers across papers on vascular development). In girdled
319 *Populus*, KNOX genes (i.e., KNAT1, KNAT3, KNAT6) are up-regulated in most stages (e.g.,
320 dedifferentiating xylem parenchyma, differentiating cambium). However, they become down-regulated in
321 the final stage when a regenerated cambium is finally observed 12 days after girdling (Supplementary
322 Table 2). KNOX genes are also down-regulated in the cambial zone of unperturbed *Populus*[22] (i.e.,
323 KNAT1, KNAT2, KNAT6; Supplementary Table 2), corroborating their reduced expression profiles in
324 mature stages of cambium development. Conversely, *KNOX* genes are up-regulated in mutants in which
325 ectopically-formed cambia are observed, such as *Populus HB4*-repressed (i.e., *KNAT3*[17]) and the *A.*
326 *thaliana* w*ox4* (i.e., KNAT1[19]) (Supplementary Table 2). Thus, *KNOX* genes appear to have been
327 independently co-opted in the molecular regulation of de novo cambia in both constitutively formed
328 (*Wisteria* vines) and injury-induced (girdled stems of *Populus*) conditions.
329 In addition to *KNOX* genes, other major cambium regulators such as phytohormones are
330 differentially expressed in constitutively and injury-induced EC (see Supplementary Table 3 for gene
331 families differentially expressed across different studies). Previous investigations suggest that *ETHYLENE*
332 *RESPONSE FACTORS* (*AP2/ERF*) may be upregulated in the absence of *PXY*, defining an alternative
333 pathway to the *PXY/CLE41* module that regulates cell division in vascular tissue development[44]. Notably,
334 *PXY/CLE41* (as well as *TDIF* and *WOX4* which may also be part of this signaling pathway) are not
335 retained as differentially expressed genes in the comparisons among cambium types in our study, while
336 *AP2/ERF* genes are differentially expressed in both Japanese wisteria and girdled *Populus*, along with
337 other genes from the ethylene signaling cascade (e.g., *EIN2* and *EIN3*) (Supplementary Table 3). Auxin-
338 related genes are considered master regulators of vascular development because they induce the
339 expression of other conserved transcription factors (e.g., *WOX4*, *PXY,* and *AtHB8*)[45,46]. In line with those

340 findings, changes in gene expression patterns in auxin-related genes in both injury-[21,47] and non-injury-
341 induced systems (this study) add evidence to shared mechanisms in de novo cambium formation (e.g.,
342 *ARF, AUX, PIN6, TIR;* Supplementary Table 3). The re-establishment of typical auxin pathways in
343 cambium regeneration in injury-induced cambia is consistent with typical cambium development, and auxin
344 is also sufficient to induce cambium regeneration *in vitro*[47]. Investigations into the role of auxin in EC
345 initiation are a required and exciting venue for future research on vascular biology.
346 Lastly, constitutively and injury-induced EC share signatures of epigenetic regulation. We found
347 that numerous methyltransferases were differentially expressed across cambium types, being upregulated
348 in Japanese wisteria EC (Supplementary Tables 1, 3). EC in lianas such as Japanese wisteria are
349 potentially triggered by external environmental factors such as mechanical and hydraulic constraints linked
350 to their climbing habit. Findings from *Bignonia magnifica* W.Bull show that adding external support for
351 these vines to climb triggers changes in the xylem anatomy from self-supporting to liana-like with
352 associated changes in the transcription profiling of the cambium[48]; thus external cues relate to anatomical
353 changes in these vines (see Supplementary Table 4 for comparison of differentially expressed genes
354 among woody vines; see also Supplementary Methods for further explanation). Although the epigenetic
355 regulation of the vascular cambium remains poorly understood, recent evidence suggests a mechanism
356 involving histone acetylation at the *WOX4* promoter[49]. The association between external factors, gene
357 regulation, and atypical vascular architectures has also been reported in swallow underground storage
358 organs, which also thicken through parenchyma proliferation, additional vascular bundles, or EC itself
359 (e.g., beets)[40]. A notable similarity across underground storage organs and EC in Japanese wisteria is the
360 role of expansin encoding genes (associated with cell wall loosening), in cell division and proliferation. We
361 found expansin downregulated in the Japanese wisteria EC, which was also reflected in the underground
362 storage organs of *Bomarea multiflora* (L.f.) Mirb.[50]. However, their specific role is yet to be further
363 elucidated given that these genes are up-regulated or down-regulated in different vascular systems (e.g.,
364 types of underground storage organs or EC)[40].
365 The enriched GO classes of genes that were differentially expressed in the "Ectopic Cambia v.
366 Typical Cambia of Japanese wisteria and Common bean" contrast (Table 3) include those related to
367 mRNA processing, splicing, and binding, as well as nucleus, nucleolus, and polymerase activity. These
368 terms may be associated with cambium formation as chromatin alterations, increased transcription, and
369 histone modification are necessary for cell dedifferentiation and cell cycle re-entry during plant
370 regeneration[51–53], as also indicated for the injury-induced EC development in *Populus*[20]. This molecular
371 program may also explain how cortical and pericyclic cells are induced to undergo cell division
372 independently of stem cell lineage in regeneration systems[51,54]. In addition, previous research shows that
373 different levels of mRNA processing factors involved in pre-mRNA splicing are critical for plant
374 development, including the differentiation process before initiation of cell division[55]. Taken together, these
375 observations indicate that similar molecular events may be involved in cambial cell competency during EC
376 development.
377 According to our KNOX phylogeny, gene copy number is variable across lineages, but there is no
378 association with EC evolution. There are instances of higher copy number of *KNOX* genes in lineages with
379 EC (e.g., Fabaceae species), while in other lineages both an increase and a reduction are observed (e.g.,
380 Amaranthaceae) in comparison to the nine copies of *A. thaliana*. Copy number variation is related to
381 duplication events identified at various taxonomic levels. Fabaceae-specific duplications were observed for
382 most genes set within the KNOX1 and KNOX2 classes, as well as within KNATM—which has been
383 reported only in a few eudicot lineages[56]. Our study suggests that the expansion of *KNOX* genes is derived
384 from either whole genome duplications (WGD), segmental, or tandem duplications but the specific
385 underlying cause for each has not been identified in this study. For instance, both WGD or segmental
386 duplications in multiple genes could explain an increase in copy number in eudicots or the Fabaceae.
387 Polyploidy events are well-known in the evolution of the genus *Glycine*[57] and the studied grasses[58,59],
388 which is more likely to explain higher copy numbers, although segmental and tandem duplications have

389 also been reported as the mechanisms in the expansion of *KNOX* genes in grasses (e.g., *Triticum*
390 *aestivum*[60]). In addition, paleopolyploidization events have been reported for all sampled angiosperm
391 genera/species, however, most have gone through diploidization events (see Supplementary Table 7 for a
392 list of polyploidization events) indicating that associating copy number variation with WGD should be taken
393 with caution. Differences in copy number within some lineages may also be a result of sampling
394 transcriptomes rather than genomes for some species (e.g., *Maianthemum canadense, Ipomoea*
395 *purpurea*).
396 Our selection analyses indicate that KNOX genes (e.g., KNAT2/6) are under positive selection,
397 including two sequences of Japanese wisteria from the transcript clusters obtained in the differential
398 expression analysis. In addition to Japanese wisteria, however, contigs from other species with or without
399 any vascular oddity are also under positive selection. Cases such as *Dioscorea,* which forms underground
400 storage organs and potentially atypical vascular tissue formation, may corroborate selective pressures in
401 the evolution of copies of KNAT2/6 associated with vascular oddities. However, we found a positive
402 selection of KNAT2/6 in species without any vascular aberration (e.g., *Populus*), thus further analyses are
403 necessary to investigate the gene-phenotype relationship.
404 In conclusion, our work provides anatomical and molecular insights into the unusual emergence of
405 EC formation from cortical parenchyma in Japanese wisteria vines. Amongst the top most differentially
406 expressed genes are numerous conserved genes involved in hormone signaling, cell division and
407 differentiation, and epigenetic regulation. Our findings reveal *KNOX* genes as important factors modulating
408 EC development in Japanese wisteria. Although none of the HD-ZIP III genes (i.e., *REV, AtHB7, AtHB4*)
409 generating EC in experimental mutants were retained as a differentially expressed gene in Japanese
410 wisteria, we cannot rule out that this pathway represents an alternative route for the emergence of de novo
411 cambia, since EC have arisen through the misexpression of HD-ZIP III genes in experimental mutants.
412 Future investigations leveraging spatial, and time-specific transcriptomes will further elucidate the genetic
413 mechanisms controlling EC and its similarities with de novo cambia in other natural and induced vascular
414 phenomena in plants (e.g., cambium regeneration, tuber formation). This knowledge will inform the
415 contribution of conserved genetic pathways to the repeated evolution of the climbing habit in plants.
416
417 **Methods**
418 ***Plants and stem sample collection***
419 *Wisteria floribunda* (Japanese wisteria) plants were collected in two populations growing in Ithaca, NY,
420 United States. We performed DNA barcoding using the chloroplast intergenic spacer gene "ndhJ-trnF" to
421 confirm the identification of plants belonging to the species *W. floribunda*[10]*. Phaseolus vulgaris* (genotype
422 L88-57; Common bean) plants were grown at Cornell University Agricultural Experiment Station Guterman
423 greenhouses (Ithaca, NY) under the following conditions: 14h light, 75 F daytime temperature, 65 F
424 nighttime temperature, and ambient humidity. Samples from these plants were used for developmental
425 anatomy and RNAseq analyses (see methods below).
426
427 ***Histology and microscopy***
428 Stem samples were hand-sectioned (with the aid of a razor blade) and processed to generate temporary
429 slides or processed using standardized methods in wood anatomy (i.e., embedded in polyethylene glycol,
430 sectioned, and stained) to generate permanent slides[61]. Slides were imaged using a light microscope
431 (Olympus BH2) with an attached digital camera (AmScope MU1000).
432
433 ***RNA sequencing and data processing***
434 *Sampling and RNA isolation* – Developed stem samples were cut into small blocks containing cambium,
435 secondary xylem, and secondary phloem, which were sectioned (~12 mm thick) using a cryostat at -20°C
436 (tangential cryosectioning approach[22]) to obtain the tissue for total RNA isolation with RNeasy Mini Kit
437 (QIAGEN). For RNA extraction, about 100 sections of each sample containing cambium, secondary xylem,

438 and secondary phloem were used for each replicate. Six biological replicates were used for the typical
439 cambium of Common bean and ectopic cambia (EC) of Japanese wisteria, while five replicates were used
440 for the typical cambium of Common bean. The same adult plants used for the comparative transcriptomes
441 were used for anatomical studies (as indicated above).
442 *RNA integrity and quantification* – Total RNA was quantified using RNA Assay Kit in Qubit® 2.0
443 Fluorometer (Life Technologies, CA, USA), and RNA integrity was assessed using an AATI Fragment
444 Analyzer (Agilent Technologies, CA, USA). Purification, library preparation, and Illumina sequencing were
445 performed by Novogene Co., Ltd (Davis, California). Quality control was performed from the RNA samples
446 to the final data using Qubit® 2.0, and RNA integrity was assessed using the RNA Nano 6000 Assay Kit of
447 the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).
448 *Library preparation for Transcriptome Sequencing* – RNA library for poly(A)-dependent RNA-seq
449 was prepared using NEBNext Ultra II RNA Library Prep Kit for Illumina (New England BioLabs). The
450 libraries were subjected to 150 bp paired-end sequencing using NovaSeq 6000 with the v1.5 S4 reagent
451 kit. The library was checked with Qubit and real-time PCR for quantification and bioanalyzer for size
452 distribution detection.
453 *Transcriptome Assembly and Annotation* – Raw paired-end reads were trimmed for adapter
454 sequences using an overlap stringency of three, a Phred score cut off of twenty, and a minimum sequence
455 length of 50 base pairs (bp) using Trim Galore! version 0.6.6[62]. De novo transcriptome assemblies were
456 developed for *P. vulgaris* and *W. floribunda*. For each trimmed dataset, the range of potential kmer values
457 was estimated with Kmergenie version 1.7016[63]. De novo assemblies were generated following part of the
458 TransPi pipeline that uses a multi-assembler approach followed by merging assemblies, reducing
459 complexity, and annotation[36] (see Supplementary methods section "TransPi pipeline" for additional
460 details). Within the TransPi pipeline, the UniProt database for Fabales (UniProt Taxon ID 72025[24])was
461 used for annotation (*see Supplementary methods section* "*Limitations of the one-to-one orthologs*
462 *approach"* for additional details).
463 *Differential Expression and Gene Ontology Enrichment* – Trimmed reads were subset to 20 million
464 reads, mapped to their respective transcriptomes with Hisat2 version 2.2.1[64] including all multiple
465 alignments, and the mate-pair information of the resulting SAM files were fixed using Picard Tools version
466 2.27.5 FixMateInformation before being converted to BAM files with SAMtools version 1.14[65]. Corset
467 version 1.07[66] was used to group the mapped transcripts into transcript clusters and estimate read counts
468 for each cluster for both transcriptomes separately. Each cluster was annotated using the Trinotate
469 annotation of the longest sequence within the cluster. Clusters were retained if at least two samples in
470 each species had 0.5 counts per million. For the cross-species comparison, Orthofinder2 version 2.5.2[67]
471 was used with our protein annotations from TransDecoder, and the proteomes of *Vigna unguiculata*
472 (UniProt Taxon ID: 3917), *Cajanus cajan* (UniProt Taxon ID: 3821), *Lotus japonicus*[68], *Medicago*
473 *truncatula*[69] and *Mucuna pruriens* (UniProt Taxon ID: 157652) from UniProt[24] and Phytozome[70] to identify
474 orthologs between *P. vulgaris* and *W. floribunda*. We identified one-to-one orthologs at the transcript
475 cluster level by removing any clusters with more than one ortholog group and any ortholog groups that
476 contained sequences from more than one cluster. Differential cluster expression were tested using limma[71]
477 and voom[72] comparing between species regardless of cambium type, between cambium types regardless
478 of species, between the typical cambium of each species, and between the ectopic and typical cambium of
479 Japanese wisteria (*see Supplementary methods section* "*Limitations of the one-to-one orthologs approach"*
480 for additional details). Gene ontology enrichment was tested using the differentially expression transcript
481 clusters for each using GOseq[73]. To better leverage the repeated sampling within an individual, Japanese
482 wisteria clusters were tested separately from the cross-species comparison using differential expression
483 for repeated measures (dream[25]) with voom[72] and variancePartition[74] to compare across cambium types
484 with individual as a random effect. Gene ontology enrichment for differentially expressed clusters was
485 implemented in GOseq. For comparison with the cross-species analysis, ortholog groups were
486 subsequently mapped back to the differentially expressed clusters identified by dream.

487
488 **Phylogenetic and Selection analyses**
489 To further investigate the role of *KNOX* genes in EC formation across seed plants, we performed a
490 phylogenetic analysis and selection analysis for this gene family. Candidate genes were identified from
491 available genomes and transcriptomes using a BLAST pipeline that was previously used for investigating
492 gene evolution across the angiosperms[75]. Coding sequences (cds) were downloaded from Genbank or
493 TreeGenes for 39 species and raw RNA-seq files were downloaded from the SRA for six species
494 (Supplementary Table 5). In total, 45 species were included in the analysis, 11 containing EC. Raw
495 sequencing files were cleaned with fastp version 0.23.3[76] and assembled with Trinity version 2.14.0[77] with
496 the following flags: --max_memory 200G --no_normalize_reads --min_contig_length 250 --trimmomatic.
497 Reference sequences of candidate genes from *A. thaliana* were downloaded from Genbank and a
498 tBLASTx search was performed using blast version 2.13[78]. A maximum of 30 hits per taxon were retained
499 if the matching length was 10 bp and an e-value of 0.0001. A fasta file containing all retrieved sequences
500 was generated and checked for duplicates using the sRNAtoolbox[79]. Twenty-one sequences from different
501 *KNOX* genes and taxa were downloaded from Genbank and included as additional references to identify
502 specific clades (Supplementary Table 6). A species tree to depict this sampling (Fig. 4) was built using the
503 R package V.PhyloMaker2 using scenario 3 for taxa binding[80]. Finally, the sequences of five transcript
504 clusters identified as KNOX genes were included in the phylogenetic analysis. The resulting fasta file for
505 phylogenetic reconstruction contained 466 sequences. These contigs were first aligned with MAFFT
506 version 7.52[81] with the following options: --auto --adjustdirectionaccurately --op 3 --leavegappyregion. The
507 sequences were then realigned with the codon-aware aligner MACSE version 2.07[82] and the
508 alignSequences command. The resulting alignment was cleaned up with trimAL version 1.4.1[83] with the -
509 gappyout option before converting premature stop codons to NNN while retaining the final stop codons
510 with the exportAlignment command in MACSE, which were cleaned up with trimAL version 1.4.1[83]. Tree
511 building was performed reiteratively to exclude poorly aligned and gappy contigs (see Supplementary
512 Methods). Trees were generated with RaxML version 8.2.12 with the GTRGAMMA model of molecular
513 evolution and 100 bootstrap replicates[84].
514 For further investigation of specific candidate genes, a fasta file containing the nucleotide
515 sequences of specific *KNOX* genes identified in the previous tree was aligned with MAFFT version 7.52,
516 realigned with the codon-aware aligner MACSE version 2.07, and then cleaned with trimAL version 1.4.1.
517 The inferred phylogeny and cleaned alignment were inputs into PAML version 10.6[26] for codeml analysis to
518 explore signatures of selections. Both model = 0 (one omega value for the entire alignment), model = 1
519 (omega value for each branch), and model = 2 (foreground and background omega values) were
520 conducted on both the rooted and unrooted trees. Branch-specific omega values were plotted on the
521 phylogeny with the R package ggtree version 3.6.2[85].
522
523 **Reporting summary**
524 Further information on research design is available in the Nature Portfolio Reporting Summary linked to
525 this article.
526
527 **Data availability**
528 The datasets generated and/or analyzed during the current study are publicly available. The RNA
529 sequencing data is available in SRA (SRR28389658, SRR29802342-SRR29802358). Other datasets
530 generated in this study have been deposited in the GitHub database
531 (https://github.com/anthonysnead/Fabaceae-Ectopic_Cambia_Transcriptomics).
532
533 **Code availability**
534 A custom R script was written for gene expression analyses and is available on GitHub
535 (https://github.com/anthonysnead/Fabaceae-Ectopic_Cambia_Transcriptomics).

## References


1. Hunziker, P. & Greb, T. Stem Cells and Differentiation in Vascular Tissues. *Annual Review of Plant Biology* **75**, 11.1-11.27 (2024).
2. Lanner, R. M. Why do trees live so long? *Ageing Research Reviews* **1**, 653–671 (2002).
3. Sillett, S. C. *et al.* Increasing wood production through old age in tall trees. *Forest Ecology and Management* **259**, 976–994 (2010).
4. Rowe, N. P. & Speck, T. The evolution of angiosperm lianescence: a perspective from xylem structure-function. in *Ecology of lianas.* 221–250 (JohnWiley & Sons, Chinchester, 2015).
5. Fischer, J. B. & Ewers, F. W. Structural responses to stem injury in vines. in *The Biology of Vines* 99–124 (Cambridge University Press, Cambridge, 1991).
6. Cunha Neto, I. L. & Onyenedum, J. G. Ectopic cambia: Connections between natural and experimental vascular mutants. *American J of Botany* **110**, e16246 (2023).
7. Terrazas, T., Aguilar-Rodríguez, S. & Ojanguren, C. T. Development of successive cambia, cambial activity, and their relationship to physiological traits in *Ipomoea arborescens* (Convolvulaceae) seedlings. *American J of Botany* **98**, 765–774 (2011).
8. Robert, E. M. R. *et al.* Successive Cambia: A Developmental Oddity or an Adaptive Structure? *PLoS ONE* **6**, (2011).
9. Nejapa, R., Cabanillas, P. A. & Pace, M. R. Cortical origin of the successive cambia in the stems of the charismatic temperate lianescent genus *Wisteria* (Fabaceae) and its systematic importance. *Botanical Journal of the Linnean Society* **199**, 667–677 (2022).
10. Cunha Neto, I. L. & Onyenedum, J. G. How to Climb and Grow Stronger: Lessons from Ornamental Wisterias. *Arnoldia* **80**, 38–47 (2023).
11. Trusty, J. L., Lockaby, B. G., Zipperer, W. C. & Goertzen, L. R. Horticulture, hybrid cultivars and exotic plant invasion: a case study of *Wisteria* (Fabaceae). *Botanical Journal of the Linnean Society* **158**, 593–601 (2008).
12. Tamaio, N., Vieira, R. C. & Angyalossy, V. Origin of successive cambia on stem in three species of Menispermaceae. *Rev. bras. Bot.* **32**, 839–848 (2009).
13. Turley, E. K. & Etchells, J. P. Laying it on thick: a study in secondary growth. *Journal of Experimental Botany* **73**, 665–679 (2022).
14. Groover, A. The vascular cambium revisited. *IAWA Journal* **44**, 531–538 (2023).
15. Robischon, M., Du, J., Miura, E. & Groover, A. The *Populus* Class III HD ZIP, *popREVOLUTA* , Influences Cambium Initiation and Patterning of Woody Stems. *Plant Physiology* **155**, 1214–1225 (2011).
16. Zhu, Y., Song, D., Sun, J., Wang, X. & Li, L. PtrHB7, a class III HD-Zip Gene, Plays a Critical Role in Regulation of Vascular Cambium Differentiation in Populus. *Molecular Plant* **6**, 1331–1343 (2013).
17. Zhu, Y., Song, D., Xu, P., Sun, J. & Li, L. A *HD-ZIP III* gene, *PtrHB4,* is required for interfascicular cambium development in *Populus*. *Plant Biotechnology Journal* **16**, 808–817 (2018).
18. Kucukoglu, M., Nilsson, J., Zheng, B., Chaabouni, S. & Nilsson, O. *WUSCHEL - RELATED HOMEOBOX 4 ( WOX 4)* -like genes regulate cambial cell division activity and secondary growth in *Populus* trees. *New Phytologist* **215**, 642–657 (2017).
19. Zhang, J. *et al.* Transcriptional regulatory framework for vascular cambium development in Arabidopsis roots. *Nat. Plants* **5**, 1033–1042 (2019).
20. Zhang, J. *et al.* Molecular features of secondary vascular tissue regeneration after bark girdling in *Populus*. *New Phytologist* **192**, 869–884 (2011).
21. Zhang, Y., Wang, X., Zhang, J. & He, X.-Q. Plant *in situ* tissue regeneration: dynamics, mechanisms and implications for forestry research. *f* **3**, 0–0 (2023).
22. Schrader, J. *et al.* A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell* **16**, 2278–2292 (2004).
23. Bryant, D. M. *et al.* A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep* **18**, 762–776 (2017).
24. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523–D531 (2023).
25. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).
26. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**, 1586–1591 (2007).

27. Jia, P. *et al.* KNOTTED1-like homeobox (KNOX) transcription factors - Hubs in a plethora of networks: A review. *International Journal of Biological Macromolecules* **253**, 126878 (2023).
28. Guillet-Claude, C., Isabel, N., Pelgas, B. & Bousquet, J. The Evolutionary Implications of knox-I Gene Duplications in Conifers: Correlated Evidence from Phylogeny, Gene Mapping, and Analysis of Functional Divergence. *Molecular Biology and Evolution* **21**, 2232–2245 (2004).
29. Bueno, N., Alvarez, J. M. & Ordás, R. J. Characterization of the KNOTTED1-LIKE HOMEOBOX (KNOX) gene family in Pinus pinaster Ait. *Plant Science* **301**, 110691 (2020).
30. Groover, A. T. *et al.* The Populus homeobox gene ARBORKNOX1 reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Mol Biol* **61**, 917–932 (2006).
31. Du, J., Mansfield, S. D. & Groover, A. T. The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. *The Plant Journal* **60**, 1000–1014 (2009).
32. Liebsch, D. *et al.* Class I KNOX transcription factors promote differentiation of cambial derivatives into xylem fibers in the *Arabidopsis* hypocotyl. *Development* **141**, 4311–4319 (2014).
33. Zhao, Y. *et al.* *KNAT 2/6b* , a class I KNOX gene, impedes xylem differentiation by regulating NAC domain transcription factors in poplar. *New Phytologist* **225**, 1531–1544 (2020).
34. Woerlen, N. *et al.* Repression of BLADE-ON-PETIOLE genes by KNOX homeodomain protein BREVIPEDICELLUS is essential for differentiation of secondary xylem in Arabidopsis root. *Planta* **245**, 1079–1090 (2017).
35. Kim, M.-H. *et al.* Wood Transcriptome Profiling Identifies Critical Pathway Genes of Secondary Wall Biosynthesis and Novel Regulators for Vascular Cambium Development in Populus. *Genes* **10**, 690 (2019).
36. Reyes-Rivera, J. *et al.* Expression of the KNOTTED HOMEOBOX Genes in the Cactaceae Cambial Zone Suggests Their Involvement in Wood Development. *Front. Plant Sci.* **8**, (2017).
37. Smith, H. M. S. & Hake, S. The Interaction of Two Homeobox Genes, *BREVIPEDICELLUS* and *PENNYWISE* , Regulates Internode Patterning in the Arabidopsis Inflorescence. *Plant Cell* **15**, 1717–1727 (2003).
38. Khan, M. *et al.* Antagonistic Interaction of BLADE-ON-PETIOLE1 and 2 with BREVIPEDICELLUS and PENNYWISE Regulates Arabidopsis Inflorescence Architecture. *Plant Physiology* **158**, 946–960 (2012).
39. Hoang, N. V., Park, C., Kamran, M. & Lee, J.-Y. Gene Regulatory Network Guided Investigations and Engineering of Storage Root Development in Root Crops. *Front. Plant Sci.* **11**, 762 (2020).
40. Plunkert, M. L., Martínez-Gómez, J., Madrigal, Y., Hernández, A. I. & Tribble, C. M. Tuber, or not tuber: Molecular and morphological basis of underground storage organ development. *Current Opinion in Plant Biology* **80**, 102544 (2024).
41. Staudt, A. & Wenkel, S. Regulation of protein function by 'microProteins'. *EMBO Reports* **12**, 35–42 (2011).
42. Ragni, L., Belles-Boix, E., Günl, M. & Pautot, V. Interaction of *KNAT6* and *KNAT2* with *BREVIPEDICELLUS* and *PENNYWISE* in *Arabidopsis* Inflorescences. *The Plant Cell* **20**, 888–900 (2008).
43. Champagne, C. E. M. *et al.* Compound Leaf Development and Evolution in the Legumes. *The Plant Cell* **19**, 3369–3378 (2007).
44. Etchells, J. P., Provost, C. M. & Turner, S. R. Plant Vascular Cell Division Is Maintained by an Interaction between PXY and Ethylene Signalling. *PLoS Genet* **8**, e1002997 (2012).
45. Johnson, L. A. & Douglas, C. J. *Populus trichocarpa MONOPTEROS/AUXIN RESPONSE FACTOR5* ( *ARF5* ) genes: comparative structure, sub-functionalization, and *Populus – Arabidopsis* microsyntenyThis article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. *Can. J. Bot.* **85**, 1058–1070 (2007).
46. Smetana, O. *et al.* High levels of auxin signalling define the stem-cell organizer of the vascular cambium. *Nature* **565**, 485–489 (2019).
47. Chen, J. *et al.* Differential regulation of auxin and cytokinin during the secondary vascular tissue regeneration in *Populus* trees. *New Phytologist* **224**, 188–201 (2019).
48. Lima, A. C. *et al. Liana Attachment to Supports Leads to Profound Changes in Xylem Anatomy and Cambium and Differentiating Xylem Transcriptional Profile*. https://www.authorea.com/users/657685/articles/662372-liana-attachment-to-supports-leads-to-profound-changes-in-xylem-anatomy-and-cambium-and-differentiating-xylem-transcriptional-

647 profile?commit=0c8718ff94a4550f3d9a1db3557aee9ac5fe4a5e (2023)
648 doi:10.22541/au.169332183.38358782/v1.
649 49. Dai, X. *et al.* Cell-type-specific PtrWOX4a and PtrVCS2 form a regulatory nexus with a histone
650 modification system for stem cambium development in Populus trichocarpa. *Nat. Plants* **9**, 96–111
651 (2023).
652 50. Tribble, C. M., Martínez-Gómez, J., Alzate-Guarín, F., Rothfels, C. J. & Specht, C. D. Comparative
653 transcriptomics of a monocotyledonous geophyte reveals shared molecular mechanisms of
654 underground storage organ formation. *Evolution & Development* **23**, 155–173 (2021).
655 51. Koukalova, B. *et al.* Dedifferentiation of Tobacco Cells Is Associated with Ribosomal RNA Gene
656 Hypomethylation, Increased Transcription, and Chromatin Alterations. *Plant Physiol* **139**, 275–286
657 (2005).
658 52. Birnbaum, K. D. & Alvarado, A. S. Slicing across Kingdoms: Regeneration in Plants and Animals. *Cell*
659 **132**, 697–710 (2008).
660 53. Sugimoto, K., Gordon, S. P. & Meyerowitz, E. M. Regeneration in plants and animals:
661 dedifferentiation, transdifferentiation, or just differentiation? *Trends in Cell Biology* **21**, 212–218 (2011).
662 54. Mironova, V. & Xu, J. A single-cell view of tissue regeneration in plants. *Current Opinion in Plant*
663 *Biology* **52**, 149–154 (2019).
664 55. Ohtani, M., Demura, T. & Sugiyama, M. Arabidopsis ROOT INITIATION DEFECTIVE1, a DEAH-Box
665 RNA Helicase Involved in Pre-mRNA Splicing, Is Essential for Plant Development. *The Plant Cell* **25**,
666 2056–2069 (2013).
667 56. Gao, J., Yang, X., Zhao, W., Lang, T. & Samuelsson, T. Evolution, diversification, and expression of
668 KNOX proteins in plants. *Front. Plant Sci.* **6**, (2015).
669 57. Zhuang, Y. *et al.* Phylogenomics of the genus Glycine sheds light on polyploid evolution and life-
670 strategy transition. *Nat. Plants* **8**, 233–244 (2022).
671 58. Landis, J. B. *et al.* Impact of whole-genome duplication events on diversification rates in angiosperms.
672 *American J of Botany* **105**, 348–363 (2018).
673 59. Zhang, T. *et al.* Phylogenomic profiles of whole-genome duplications in Poaceae and landscape of
674 differential duplicate retention and losses among major Poaceae lineages. *Nat Commun* **15**, 3305
675 (2024).
676 60. Han, Y. *et al.* Genome-wide analysis of TALE superfamily in Triticum aestivum reveals TaKNOX11-A
677 is involved in abiotic stress response. *BMC Genomics* **23**, 89 (2022).
678 61. Barbosa, A. C. F., Pace, M. R., Witovisk, L. & Angyalossy, V. A new method to obtain good
679 anatomical slides of heterogeneous plant parts. *IAWA J* **31**, 373–383 (2010).
680 62. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
681 *EMBnet.journal* **17**, 10–12 (2011).
682 63. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly.
683 *Bioinformatics* **30**, 31–37 (2014).
684 64. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements.
685 *Nat Methods* **12**, 357–360 (2015).
686 65. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
687 (2009).
688 66. Davidson, N. M. & Oshlack, A. Corset: enabling differential gene expression analysis for de
689 novoassembled transcriptomes. *Genome Biology* **15**, 410 (2014).
690 67. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics.
691 *Genome Biol* **20**, 238 (2019).
692 68. Li, H., Jiang, F., Wu, P., Wang, K. & Cao, Y. A High-Quality Genome Sequence of Model Legume
693 Lotus japonicus (MG-20) Provides Insights into the Evolution of Root Nodule Symbiosis. *Genes*
694 *(Basel)* **11**, 483 (2020).
695 69. Tang, H. *et al.* An improved genome release (version Mt4.0) for the model legume Medicago
696 truncatula. *BMC Genomics* **15**, 312 (2014).
697 70. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids*
698 *Res* **40**, D1178-1186 (2012).
699 71. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray
700 studies. *Nucleic Acids Research* **43**, e47 (2015).
701 72. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis
702 tools for RNA-seq read counts. *Genome Biology* **15**, R29 (2014).

73. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**, R14 (2010).
74. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).
75. Phillips, H. R., Landis, J. B. & Specht, C. D. Revisiting floral fusion: the evolution and molecular basis of a developmental innovation. *Journal of Experimental Botany* **71**, 3390–3404 (2020).
76. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
77. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* **29**, 644–652 (2011).
78. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
79. Aparicio-Puerta, E. *et al.* sRNAbench and sRNAtoolbox 2022 update: accurate miRNA and sncRNA profiling for model and non-model organisms. *Nucleic Acids Research* **50**, W710–W717 (2022).
80. Jin, Y. & Qian, H. V.PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Diversity* **44**, 335–339 (2022).
81. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
82. Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E. J. P. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLOS ONE* **6**, e22594 (2011).
83. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
84. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
85. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36 (2017).

**Acknowledgments**

We thank undergraduate research technician Danielle C. Sonnenleiter (Cornell University) for assistance in data collection, past and current members of the Onyenedum lab for continuous feedback, especially Angelique A. Acevedo for assistance in growing the bean plants, and Mariane S. S. Baena for insightful discussions. We also thank Jocelyn Rose's lab at Cornell University for allowing access to the cryostat. We would like to acknowledge the Arnold Arboretum of Harvard University for providing access to the living collections and financial support through a Sargent Award for Visiting Scholars (I.L.C.N.). This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise as well as the Boyce Thompson Institute's Computational Biology Center and Cornell's BioHPC. This work was funded by startup laboratory funds from Cornell University's College of Agriculture and Life Sciences and New York University, and NSF 2237046 to J.G.O.

**Author contributions**

Conceptualization: I.L.C.N., J.G.O. Methodology: I.L.C.N., J.B.L., C.D.S, A.A.S., J.G.O. Investigation: I.L.C.N., J.B.L., A.A.S. Data curation: I.L.C.N., A.A.S Writing – Original Draft: I.L.C.N., J.B.L., A.A.S., J.G.O. Writing – Review & Editing: I.L.C.N., J.B.L., A.A.S., C.D.S., J.G.O.

**Competing interests**

The authors declare no competing interests.

**Additional information (Supplementary information)**

**The online version contains supplementary material available at:**

https://github.com/anthonysnead/Fabaceae-Ectopic_Cambia_Transcriptomics

**Supplementary Table 1**: All significantly differentially expressed transcript clusters across contrasts.
**Supplementary Table 2**: Comparison between differentially expressed cambial markers across seven studies.
**Supplementary Table 3**: Comparison between differentially expressed genes in this study and induced cambium generation in *Populus*.
**Supplementary Table 4**: Comparison between differentially expressed genes from transcriptomes of woody vines.
**Supplementary Table 5**: List of taxa and source of data used for molecular phylogenetic analysis.
**Supplementary Table 6**: List of taxa downloaded from GenBank used as additional references for molecular phylogenetic analysis.
**Supplementary Table 7**: Number of paleopolyploid events in each species in this study.
**Supplementary Figure 1**: Schematic phylogram of the *KNOX* gene family indicating the three classes, respective genes, and duplication events.
**Supplementary Figure 2**: Schematic phylogram of the *KNOX* gene family including deleted species.
**Supplementary Figure 3**: Rooted tree fitted for branch models for the three genes investigated under selection analysis.
**Supplementary Dataset 1**: Nucleotide alignment for 466 KNOX sequences used in this study.
**Supplementary Dataset 2:** Unaligned fasta file for 466 KNOX sequences used in this study.
**Supplementary Methods**: Description of how Supplementary Tables 2-4 were developed, steps for the KNOX phylogeny, and details on the transPi pipeline.

**Figure 1**

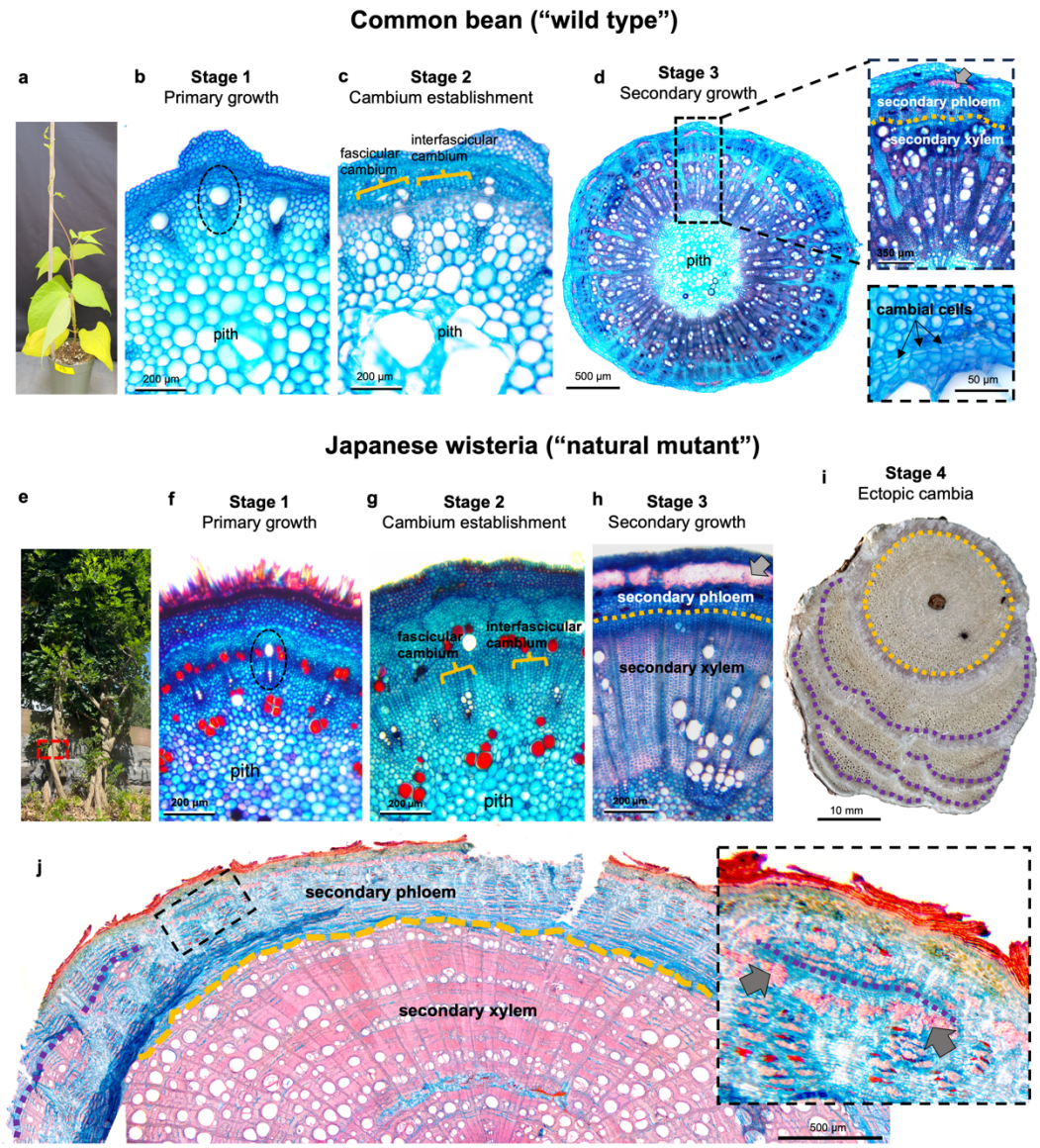
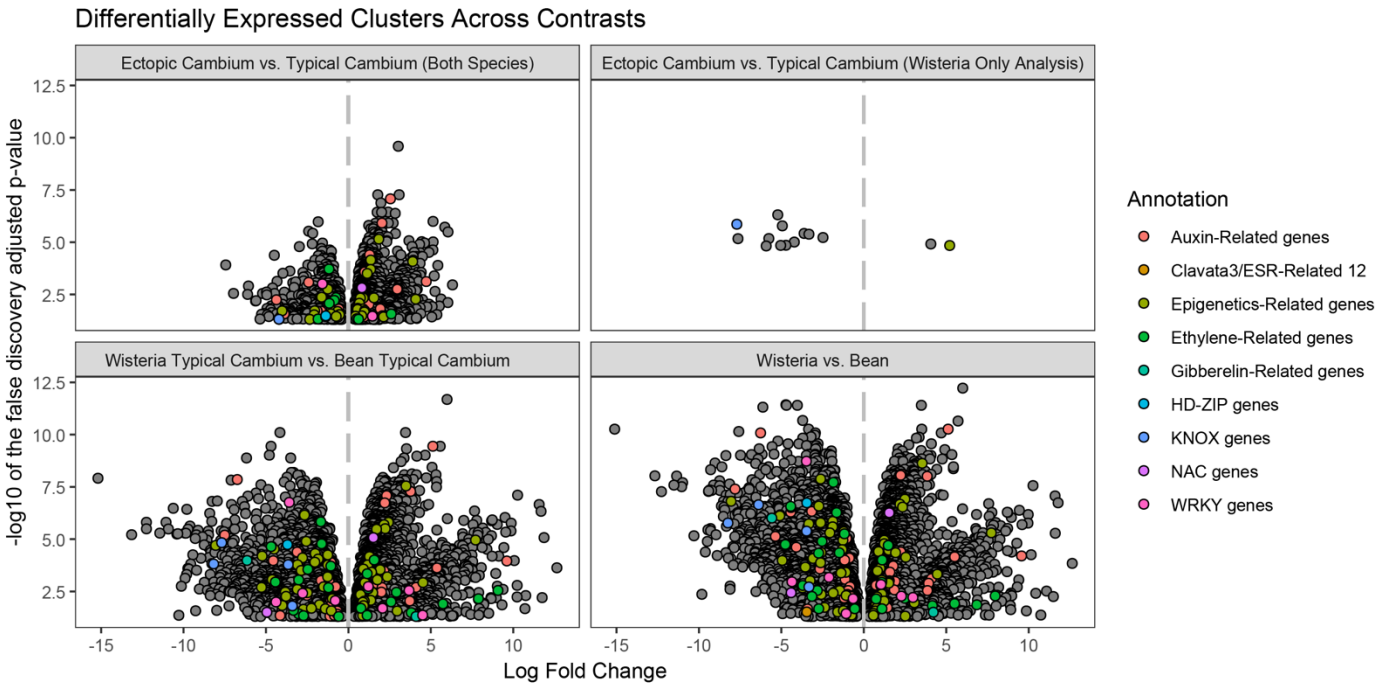


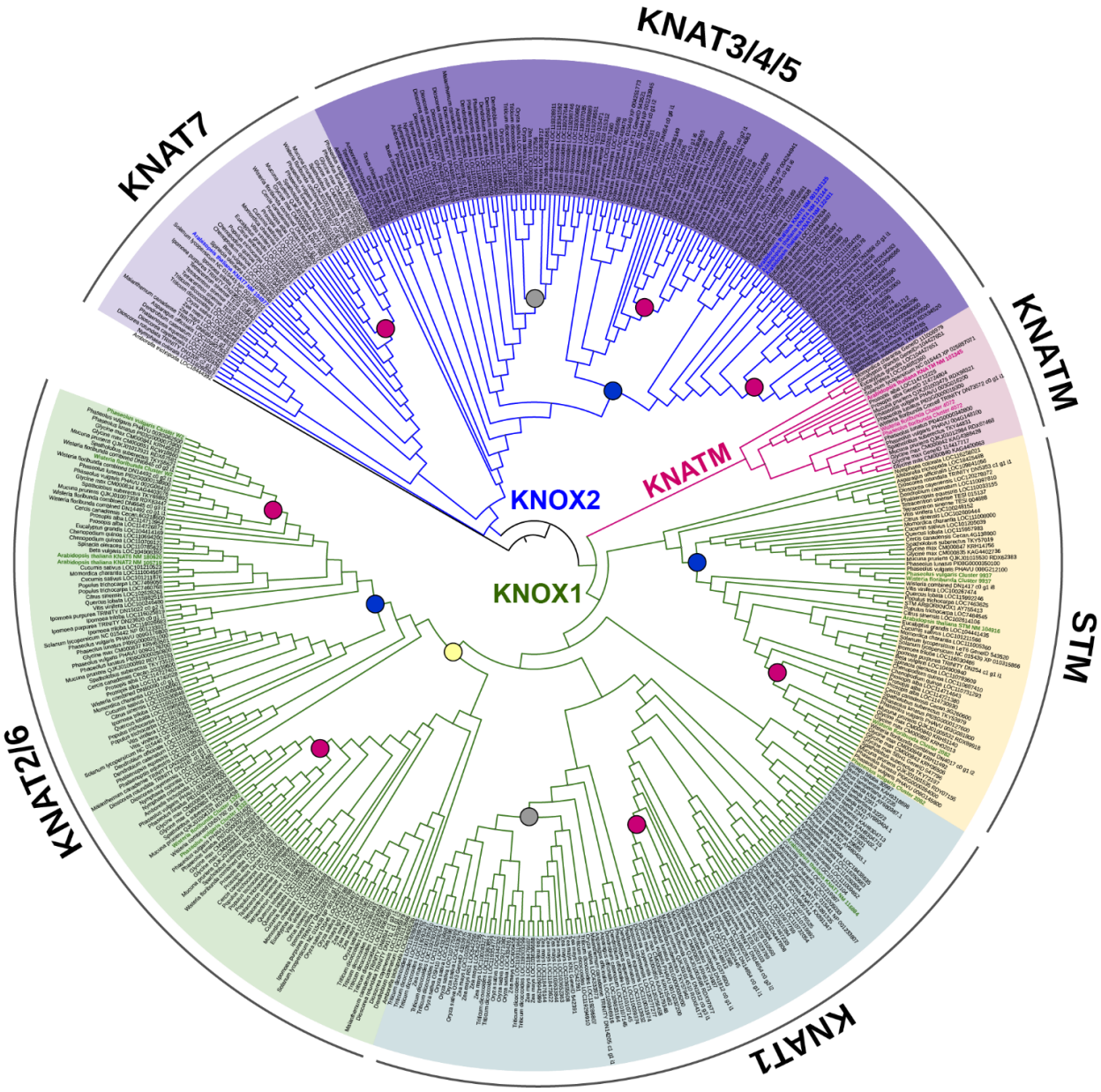

**Fig 1. Vascular development in Common bean (*Phaseolus vulgaris*) and Japanese wisteria (*Wisteria floribunda*) share conserved developmental stages, while Japanese wisteria later forms ectopic cambia (EC) from cortical parenchyma. a-d** Common bean. **e-i** Japanese wisteria. **a, e** Common bean and Japanese wisteria plants sampled for developmental anatomy. The dashed rectangle in "**e**" indicates the position of the image displayed in stage 4 (**i**). **b, f** Primary growth showing vascular bundles (dashed circle), formed by xylem inwards and phloem outwards, arranged in a single ring delimiting the pith. **c, g** Initiation of typical cambium from the fascicular cambium located within the bundles, and the interfascicular cambium between the bundles. **d, h** Secondary growth generated by the typical cambium (dashed yellow line) producing secondary xylem and secondary phloem. In **d,** the bottom right inset shows details of the typical cambium extracted from the top right inset. **i,** Mature stem of Japanese wisteria with EC (dashed purple line) forming multiple vascular increments in addition to the vascular cylinder produced by the typical cambium (dashed yellow line). **j** Detail of stem in advanced typical secondary growth derived from the typical cambium (dashed yellow line) and the emergence of EC (dashed purple line) in Japanese wisteria; the dashed box indicates the position of the inset (on the right). Grey arrow = pericyclic fibers. All figures are displayed in cross-section and indicate magnification.

825 **Figure 2**
826




827
828 **Fig. 2. Volcano plot of differentially expressed transcript clusters by contrast.** The y-axis is the
829 negative log10 of the false discovery adjusted p-value, while the x-axis is the log fold change. The
830 transcript clusters classified with colors are related to cambium development in other plants and the group
831 annotation is provided in the legend.
832
833

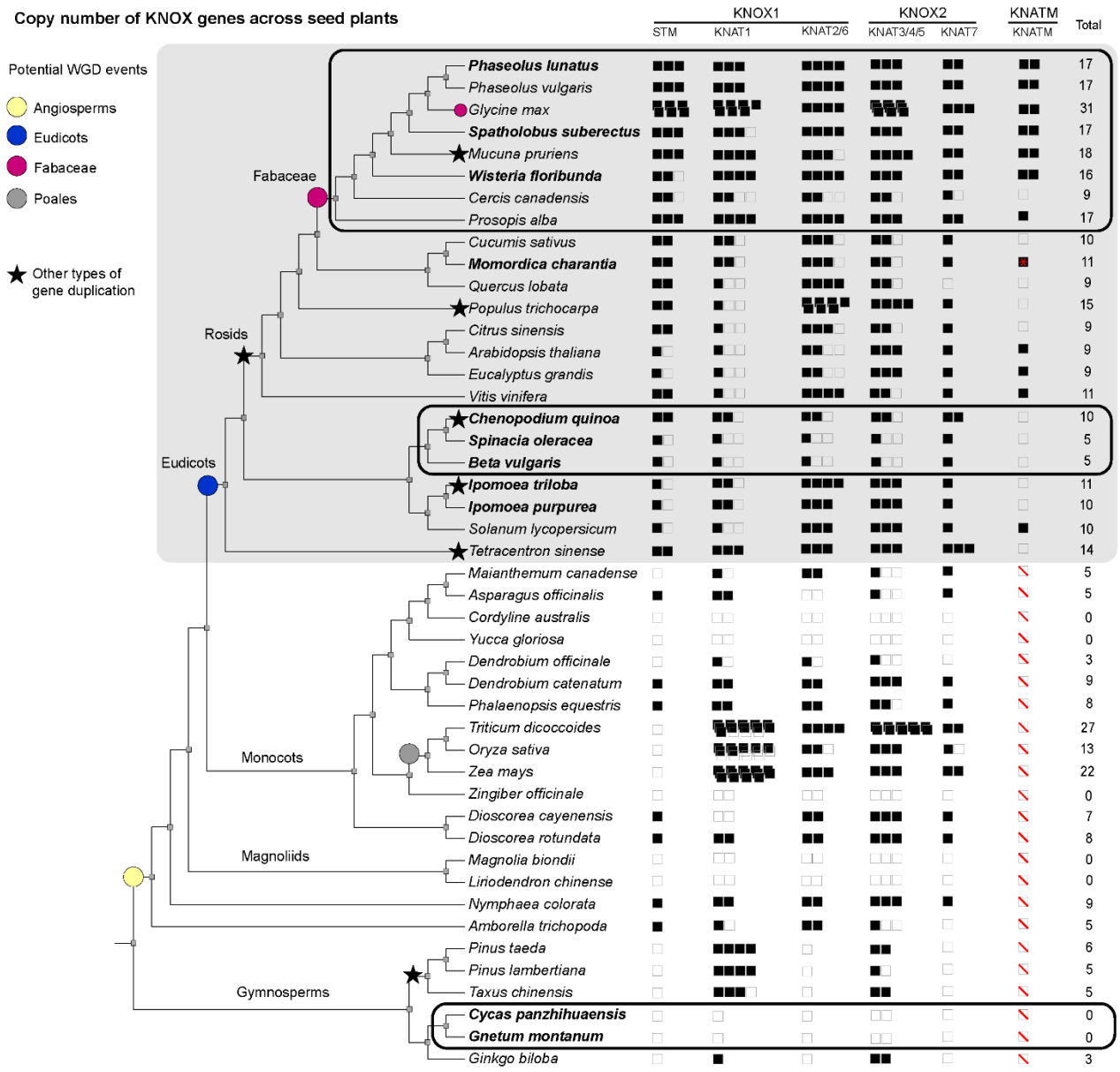
834 **Figure 3**




835

836 **Fig 3. Maximum likelihood phylogenetic tree of KNOX proteins derived from amino acid alignment.**
837 The phylogenetic tree includes members of *KNOX* genes obtained from a tBLASTx search, including
838 members of seed plants from 20 families, 33 genera, and 38 species, including 11 species that develop
839 ectopic cambia. KNOX proteins are divided into three classes; KNOX1 (branches in green), KNOX2
840 (branches in blue), and KNATM (branches colored in magenta). These classes are further divided into
841 clades corresponding to individual genes named after *Arabidopsis thaliana* orthologs, as indicated in bold
842 colors (green, blue, magenta). All recovered transcript clusters annotated for KNOX genes are likewise
843 highlighted in bold colors corresponding to their identified clades (KNOX1, KNOX2, or KNATM). The tree is
844 rooted in a member of the *BEL1* gene family. Circles indicate angiosperm (yellow), eudicots (blue), Poales
845 (grey), and Fabaceae-specific (magenta) duplications.

846 **Figure 4**
847




848
849 **Fig. 4. Copy number variation of *KNOX* genes plotted in the species tree from the sampled taxa**
850 **used in the phylogenetic analyses (Fig. 3).** Copy numbers are inferred from the phylogeny in Fig. 3 and
851 associated with duplication events. Squares indicate copy number, with filled squares indicating gene
852 presence, empty squares inferring gene losses, and white crossed squares indicating gene absence.
853 Whole genome duplications (WGD) inferred in Fig. 3 are indicated by colored circles; stars indicate
854 inferred taxon and/or copy-specific duplication events. Expansion of some *KNOX* genes is evident after
855 WGD events in angiosperms (yellow circle), Eudicots (blue circle; grey box), and Fabales (magenta circle).
856 Names of species with ectopic cambia (EC) are in bold. Black ellipses indicate variation in copy number in
857 lineages with EC. All copies were recovered from the blast pipeline to build the phylogeny (Fig. 3), except
858 for *Momordica charantia* KNATM (red asterisk) which was downloaded from Genbank.

**Figure 5**

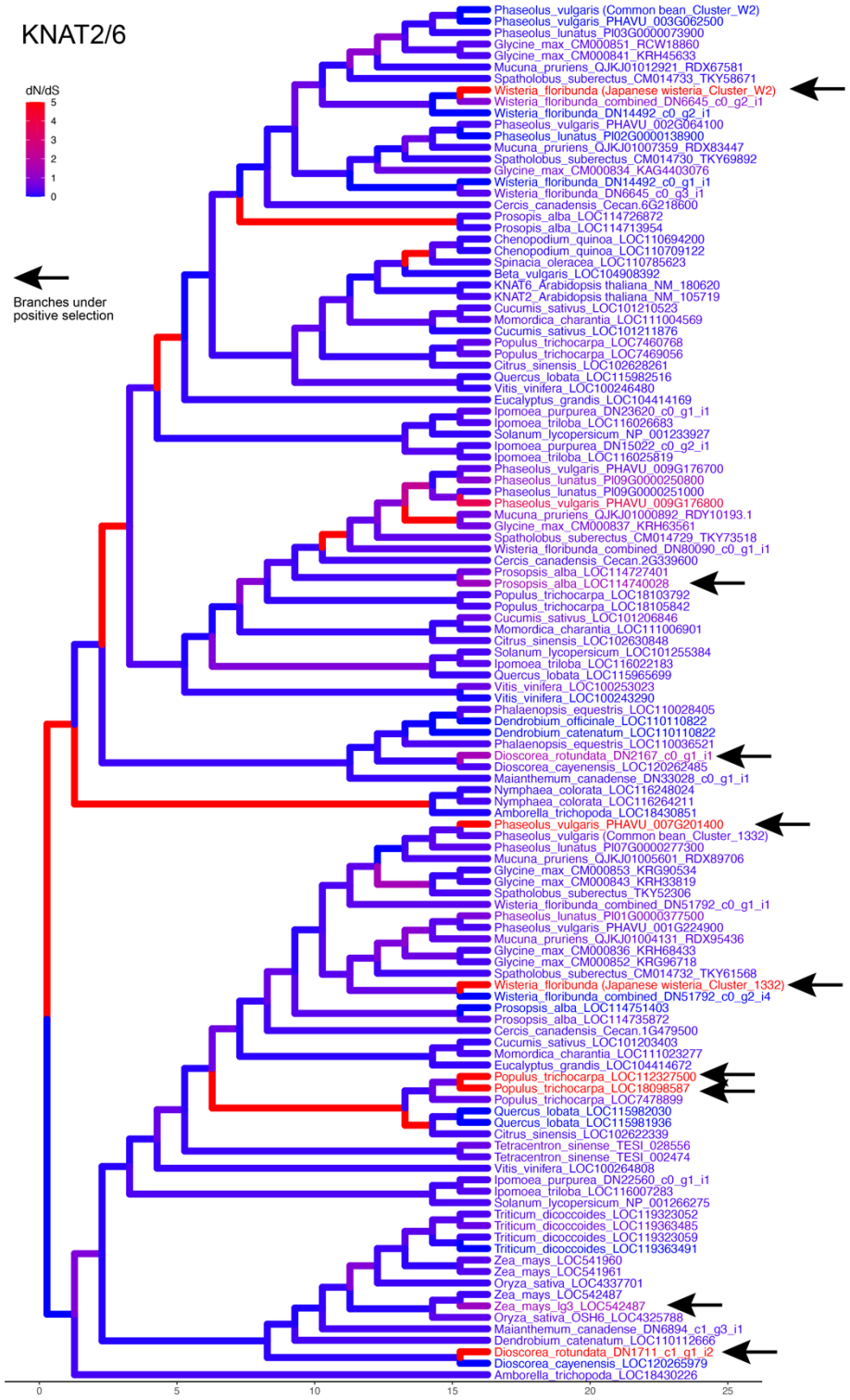



**Fig. 5. Branch-specific selection on KNAT2/6 gene family.** Rooted tree fitted for branch models for KNAT2/6 gene investigated under selection analysis. A unique omega value (dN/dS ratio, ω) is assigned per branch (model 1). Branches in red indicate ω > 1 (positive selection; indicated by arrows), and blue branches indicate ω < 1 (purifying selection).

**Ectopic cambia in Japanese wisteria (*Wisteria floribunda*) vines are associated with the expression of conserved *KNOX* genes**

Israel L. Cunha Neto[1]*, Anthony A. Snead[2], Jacob B. Landis[3], Chelsea D. Specht[3], and Joyce G. Onyenedum[1]*

[1]Department of Environmental Studies, New York University, New York 10012, NY, USA
[2] Department of Biology, New York University, New York 10012, NY, USA
[3]School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca NY, USA

*Corresponding author: israellopescn@gmail.com; jgo5750@nyu.edu

## Supplementary information

**Supplementary Table 1**: All significantly differentially expressed transcript clusters across contrasts.
**Supplementary Table 2**: Comparison between differentially expressed cambial markers across seven studies.
**Supplementary Table 3**: Comparison between differentially expressed genes in this study and induced cambium generation in *Populus*.
**Supplementary Table 4**: Comparison between differentially expressed genes from transcriptomes of woody vines.
**Supplementary Table 5**: List of taxa and source of data used for molecular phylogenetic analysis.
**Supplementary Table 6**: List of taxa downloaded from GenBank used as additional references for molecular phylogenetic analysis.
**Supplementary Table 7**: Number of paleopolyploid events in each species in this study.
**Supplementary Figure 1**: Schematic phylogram of the *KNOX* gene family indicating the three classes, respective genes, and duplication events.
**Supplementary Figure 2**: Schematic phylogram of the *KNOX* gene family including deleted species.
**Supplementary Figure 3**: Rooted tree fitted for branch models for the three genes investigated under selection analysis.
**Supplementary Dataset 1**: Nucleotide alignment for 466 KNOX sequences used in this study.
**Supplementary Dataset 2:** Unaligned fasta file for 466 KNOX sequences used in this study.
**Supplementary Methods**: Description of how Supplementary Tables 2-4 were developed, steps for the KNOX phylogeny, and details on the transPi pipeline.

## Supplementary Figures (legends)

**Fig. S1. Schematic phylogeny of the *KNOX* gene family indicating the three classes with respective genes.** Numbers indicate bootstrap value. Branches of major clades are color-coded as follows: eudicots (green), monocots (dark red), *Amborella* +*Nymphaea* (black), and gymnosperms (yellow). Black arrows indicate the position of *Arabidopsis thaliana* references and names of additional references included in the phylogeny are shown in light green (see also Table S4 and S5). The sequences differentially expressed in our study are shown in pink (indicated with arrows). Circles indicated potential whole genome duplications according to Fig. 3 and Fig. 4.

**Fig. S2.** Schematic phylogram of the *KNOX* gene family including deleted species. The nine species indicated in the purple clade were not retained for the final KNOX phylogeny.

**Fig. S3**: Rooted tree fitted for branch models for the three genes investigated under selection analysis. The arrows indicate the branches under positive selection, *i.e.*, omega values greater than 1 (ω >1). **a** KNATM. **b** STM. Arrows indicate branches under positive selection.

## Supplementary Methods S1

### Building Supplementary Tables S2-S4

To construct **Supplementary Table S2**, we utilized two gene sets as queries across referenced papers. The first set comprises nine *KNOX* genes identified in *Arabidopsis thaliana*, also used in phylogenetic analysis (see Methods). The second set includes 20 genes enriched in single-cell transcriptome analysis of vascular meristems in *Populus*[1]. Gene names (acronyms in column C of Supplementary Table 2) were obtained by querying the NCBI database for gene names (column D) and loci (column E). Column F information was sourced from the Gene Ontology database (https://amigo.geneontology.org/) using *A. thaliana* locus IDs (column E), emphasizing ontology descriptions linked to ectopic cambia (EC) development. Original annotations are in Supplementary Table 1 and referenced studies.

Selected genes (by locus number or acronyms in column C) were cross-referenced with differential gene expression databases from our study and seven others focusing on cambium in *A. thaliana* and *Populus*. These papers fall into two groups: (1) studies without mutant lines such as the wild *Populus* cambium transcriptome[2] and cambium regeneration after girdling[3], detailed further in **Supplementary Table S3** due to relevance to EC. (2) Studies with mutant lines including KNAT1 (=ARK2) in *Populus* and *A. thaliana* mutants reported with EC[4–8]. For studies with multiple mutants, each line's gene expression is shown in separate columns. Studies without differential gene expression data were not included[9].

In **Supplementary Table S2**, cells are labeled "down-regulated" or "up-regulated" based on log fold change (logFc) from original studies, except for one study which is represented by its "rC/diC" comparisons[3]. Other studies indicated log-fold change through plots[2] and those were interpreted accordingly.

**Supplementary Table S4** was created using differentially expressed genes listed across the sole studies on liana stem transcriptomes[10,11], and cross-checked with our study.

994 **Summary of steps to generate KNOX phylogeny**
995 (Files are available on https://github.com/anthonysnead/Fabaceae-
996 Ectopic_Cambia_Transcriptomics).
997

| Step | Description | Notes | # of sequences |
|---|---|---|---|
| Step_1 | Original output from BLAST pipepline | Includes additional sequences from GenBank (see Table S6), and the transcript clusters from our differential expression analysis. | 5571 |
| Step_2 | Deleted duplicated sequences using sRNAtoolbox | 4660 sequences were removed. | 911 |
| Step_3 | Deleted *BEL1* genes, duplicated isoforms, and small sequences. | Small sequences (<450 bp) were deleted from all taxa, except the Fabaceae. In total, 255 sequences were removed. | 656 |
| Step_4 | Deleted an additional 39 duplicated contigs. | | 617 |
| Step_5 | Running MACSE | We aligned the previous fasta file using MAFFT and ran MACSE and then trimAL + raxML to generate a tree. | 617 |
| Step_6 | Deleted remaining *BEL1* contigs. | BELL contigs were deleted based on their phylogenetic position. In total, 92 sequences were deleted. One BELL sequence was maintained to root the tree. | 525 |
| Step_7 | Detected contigs that poorly align and that likely does not belong to *KNOX* genes. | A tree is shown to illustrate this clade. | 525 |
| Step_8 | Deleted the "non-KNOX" group mentioned above and ran a new MACSE alignment. | | 497 |
| Step_9 | Deleted 31 gappy sequences (with ~1/3 of missing data). | See the list of deleted sequences in the file "Deleted_sequences", which include also the sequences listed under Step_3 and Step_4. | 466 |

998
999
1000 **TransPi Pipeline**
1001
1002 All de novo transcriptomes were constructed with the TransPi pipeline[12]. Briefly, assemblies were generated
1003 for each dataset using rnaSPAdes version 3.14[13], SOAPdenovo-Trans version 1.03[14], TRANS-ABYSS
1004 version 2.0.1[15], Velvet version 1.2.12/Oasis version 0.2.09[16,17] for kmer values 25, 35, 45, 55, and 65 along
1005 with Trinity version 2.9.1 [18] before being merged and reduced with the EvidentialGene pipeline version
1006 2019.05.14[19]. The EvidentialGene pipeline reduces the complexity of the transcriptomes by collapsing
1007 duplicates, clustering protein sequences, and performing similarity searches between the transcripts with a
1008 combination of CD-HIT[20] and BLAST version 2.2.31[21]. Quality assessment was done using TransRate
1009 version 1.0.3 (Smith-Unna et al. 2016) for assembly statistics, Hisat2 version 2.2.1[22] to assess mapping
1010 efficiency and Benchmarking Universal Single-Copy Orthologs (BUSCO) version 5.5.0cv1 to evaluate the
1011 completeness of expected universal single copy genes with the fables database[23]. For annotation,
1012 TransDecoder version 5.5.0[24] was employed to denote open reading frames (ORFs) and complete homology
1013 searches of the ORFs against proteins via BLAST to preserve ORFS with potential functional significance).
1014 HMMER version 3.3[25] was used to complete protein domain searches against the PFAM database. Trinotate

version 3.2.0[26] was used for functional annotation and DIAMOND version 0.9.30[27] was used for the similarity searches between the transcripts and the UniProt database for fables (UniProt Taxon ID 72025[28]) The final annotation report from Trinotate included the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO) as well as protein alignments.

**Limitations of the one-to-one orthologs approach**

It is important to note that a limitation of working with transcriptomes instead of genomes is that our clusters may not accurately represent genes, and we have been careful to refer to these as transcript clusters to reflect this uncertainty. Genes may have orthologs across species, but the corset clustering might split a real gene into two clusters, affecting our ability to identify one-to-one orthologs. It is also important to note that not all genes are likely annotated in the transcriptomes. Because RNA-seq investigates relative gene expression, reads were subset to a standard 20 million reads before quantification which could reduce the counts of rare genes. Additionally, when filtering for lowly expressed genes by biological treatment, a threshold (i.e., 10 counts) is set for the number of reads that need to be aligned to a given transcript cluster for at least two replicates for it to be retained for downstream analysis. Therefore, if at least two samples in the species do not have at least 10 reads aligned to that transcript cluster, it is discarded. Generally, these are considered genes with extremely low expression and do not contain enough information to be statistically valid. Therefore, three main scenarios may lead to conserved genes not being represented in the final analysis: (1) not assigning one-to-one orthologs due to gene splitting across clusters, (2) unsuccessful protein annotations, and (3) no differential expression in the current analysis, which may be affected by filtering.

**References**


1. Du, J. *et al.* High-resolution anatomical and spatial transcriptome analyses reveal two types of meristematic cell pools within the secondary vascular tissue of poplar stem. *Molecular Plant* **16**, 809–828 (2023).
2. Schrader, J. *et al.* A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell* **16**, 2278–2292 (2004).
3. Zhang, J. *et al.* Molecular features of secondary vascular tissue regeneration after bark girdling in *Populus*. *New Phytologist* **192**, 869–884 (2011).
4. Du, J., Mansfield, S. D. & Groover, A. T. The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. *The Plant Journal* **60**, 1000–1014 (2009).
5. Robischon, M., Du, J., Miura, E. & Groover, A. The *Populus* Class III HD ZIP, *popREVOLUTA* , Influences Cambium Initiation and Patterning of Woody Stems. *Plant Physiology* **155**, 1214–1225 (2011).
6. Zhu, Y., Song, D., Sun, J., Wang, X. & Li, L. PtrHB7, a class III HD-Zip Gene, Plays a Critical Role in Regulation of Vascular Cambium Differentiation in Populus. *Molecular Plant* **6**, 1331–1343 (2013).
7. Zhu, Y., Song, D., Xu, P., Sun, J. & Li, L. A *HD-ZIP III* gene, *PtrHB4,* is required for interfascicular cambium development in *Populus*. *Plant Biotechnology Journal* **16**, 808–817 (2018).
8. Zhang, J. *et al.* Transcriptional regulatory framework for vascular cambium development in Arabidopsis roots. *Nat. Plants* **5**, 1033–1042 (2019).
9. Kucukoglu, M., Nilsson, J., Zheng, B., Chaabouni, S. & Nilsson, O. *WUSCHEL - RELATED HOMEOBOX 4 ( WOX 4)* -like genes regulate cambial cell division activity and secondary growth in *Populus* trees. *New Phytologist* **215**, 642–657 (2017).
10. Hou, C. *et al.* Comparative Analyses of Full-Length Transcriptomes Reveal Gnetum luofuense Stem Developmental Dynamics. *Front. Genet.* **12**, 615284 (2021).
11. Lima, A. C. *et al.* Liana Attachment to Supports Leads to Profound Changes in Xylem Anatomy and Cambium and Differentiating Xylem Transcriptional Profile. Preprint at https://doi.org/10.22541/au.169332183.38358782/v1 (2023).

12. Rivera-Vicéns, R. E., Garcia-Escudero, C. A., Conci, N., Eitel, M. & Wörheide, G. TransPi-a comprehensive TRanscriptome ANalysiS PIpeline for de novo transcriptome assembly. *Mol Ecol Resour* **22**, 2070–2086 (2022).
13. Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A. D. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* **8**, giz100 (2019).
14. Xie, Y. *et al.* SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **30**, 1660–1666 (2014).
15. Robertson, G. *et al.* De novo assembly and analysis of RNA-seq data. *Nat Methods* **7**, 909–912 (2010).
16. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829 (2008).
17. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
18. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* **29**, 644–652 (2011).
19. Gilbert, D. G. Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? 829184 Preprint at https://doi.org/10.1101/829184 (2019).
20. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
21. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).
22. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357–360 (2015).
23. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
24. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).
25. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29–W37 (2011).
26. Bryant, D. M. *et al.* A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep* **18**, 762–776 (2017).
27. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60 (2015).
28. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523–D531 (2023).