the plant journal
SEB Society for Experimental Biology

The Plant Journal (2016) 86, 376–390 doi: 10.1111/tpj.13157

Functional network analysis of genes differentially expressed during xylogenesis in *soc1ful* woody Arabidopsis plants

Nicolas Davin[1,*], Patrick P. Edger[2], Charles A. Hefer[3,4], Eshchar Mizrachi[5,6], Mathias Schuetz[3,7], Erik Smets[1,8], Alexander A. Myburg[5,6], Carl J. Douglas[3], Michael E. Schranz[9] and Frederic Lens[1]

[1]*Naturalis Biodiversity Center, Leiden University, PO Box 9517, 2300 RA Leiden, The Netherlands,*
[2]*Department of Horticulture, Michigan State University, East Lansing, MI 48823, USA,*
[3]*Department of Botany, University of British Columbia, Department of Botany, 6270 University Blvd, Vancouver BC V6T 1Z4, Canada,*
[4]*Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa,*
[5]*Department of Genetics, University of Pretoria, PO Box X20, Pretoria, 0028, South Africa,*
[6]*Genomics Research Institute (GRI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa,*
[7]*Michael Smith Laboratories, University of British Columbia, 6270 University boulevard, V6T 1Z4, Vancouver, BC, Canada,*
[8]*Ecology, Evolution and Biodiversity Conservation Section, Katholieke Universiteit Leuven, Kasteelpark Arenberg 31 box 2435, 3001 Leuven, Belgium, and*
[9]*Biosystematics Group, Wageningen University, PO Box 16, 6700AP Wageningen, The Netherlands*

Received 26 November 2015; revised 29 February 2016; accepted 3 March 2016; published online 8 March 2016.
*For correspondence (e-mail nicolas.davin@outlook.com).
Accession numbers for sequence data: GSM1694242–GSM1694262.

SUMMARY

Many plant genes are known to be involved in the development of cambium and wood, but how the expression and functional interaction of these genes determine the unique biology of wood remains largely unknown. We used the *soc1ful* loss of function mutant – the woodiest genotype known in the otherwise herbaceous model plant Arabidopsis – to investigate the expression and interactions of genes involved in secondary growth (wood formation). Detailed anatomical observations of the stem in combination with mRNA sequencing were used to assess transcriptome remodeling during xylogenesis in wild-type and woody *soc1ful* plants. To interpret the transcriptome changes, we constructed functional gene association networks of differentially expressed genes using the STRING database. This analysis revealed functionally enriched gene association hubs that are differentially expressed in herbaceous and woody tissues. In particular, we observed the differential expression of genes related to mechanical stress and jasmonate biosynthesis/signaling during wood formation in *soc1ful* plants that may be an effect of greater tension within woody tissues. Our results suggest that habit shifts from herbaceous to woody life forms observed in many angiosperm lineages could have evolved convergently by genetic changes that modulate the gene expression and interaction network, and thereby redeploy the conserved wood developmental program.

Keywords: *Arabidopsis thaliana*, network analysis, secondary woodiness, transcriptome remodeling, wood formation.

INTRODUCTION

Primary growth in plants is established during seedling development by the root and shoot apical meristems (RAM/SAM), enabling apical growth, and by the procambium initiating the vascular bundles. At later stages of development, the formation and activation of lateral meristems (vascular cambium and cork cambium) in many plants results in secondary growth, a crucial adaptation that has allowed the evolution of large terrestrial woody plants. By far the largest proportion of plant biomass in mature trees and shrubs is secondary xylem (or wood), which is produced by the vascular cambium towards the inside of the stem where it serves to transport water and nutrients upwards and to provide mechanical strength (Esau, 1977; Larson, 1994).

In contrast to gymnosperms where the woody growth form is predominant, angiosperms show numerous habit forms and shifts, ranging from the woody ancestral state ('primary' woodiness) to herbaceousness and back to derived woodiness (Rowe and Paul-Victor, 2012). This re-establishment of woodiness in otherwise herbaceous angiosperm groups reflects an evolutionary phenomenon known as insular woodiness or 'secondary' woodiness (Carlquist, 1974; Lens *et al.*, 2013a). The fact that derived woodiness has evolved hundreds of times suggests that the underlying genetic mechanism is relatively simple. Currently, the classical hypothesis explaining these convergent shifts is that modification in expression of a small number of genes, acting as 'master switches', are responsible for these habit shifts in many herbaceous plant groups (Groover, 2005; Spicer and Groover, 2010). The loss-of-function double mutant *soc1ful* in Arabidopsis may be an example supporting such a master switch hypothesis. In this two-gene model, knock-out of genes encoding MADS-box flowering time transcription factors *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *FRUITFULL* (*FUL*), results in the development of a shrubby phenotype and production of a much larger extent of woody tissue than any Arabidopsis genotype or treatment described to date (Melzer *et al.*, 2008; Lens *et al.*, 2012). However, it is not known whether the flowering time genes *SOC1* and *FUL*, or specific up/downstream genes, or perhaps genes that do not directly interact with *SOC1* and *FUL*, serve as the master switches that directly suppress secondary growth in wild-type plants. It is also possible that the convergent habit shifts can be explained by the occurrence of a global gene interaction network, where any genetic change(s) modulating the gene expression and interaction network could potentially act as 'master switches' and thereby redeploy the conserved wood developmental pathway (Kaufmann *et al.*, 2010).

Transcriptional regulation is a primary control point for wood development characterized by the tight spatial and temporal regulation of genes expressed during specific stages of wood development, starting from cambial cell division, cell differentiation, cell expansion, secondary cell wall deposition and ending in programmed cell death (Hertzberg *et al.*, 2001; Israelsson *et al.*, 2003; Schrader *et al.*, 2004; Zhou *et al.*, 2009; Du and Groover, 2010; Agusti *et al.*, 2011b; Hussey *et al.*, 2013). Phytohormone signaling is another important component that regulates secondary growth. Initiation and maintenance of the vascular cambium is influenced by auxin (Aloni, 1987; Uggla *et al.*, 1996, 1998; Little *et al.*, 2002; Nilsson *et al.*, 2008; Agusti *et al.*, 2011b) and cytokinin (Matsumoto-Kitano *et al.*, 2008; Hejatko *et al.*, 2009), but other hormones such as ethylene (Love *et al.*, 2009; Chang *et al.*, 2013), gibberellins (Björklund *et al.*, 2007; Ragni *et al.*, 2011), jasmonates (Sehr *et al.*, 2010) and strigolactones (Agusti *et al.*, 2011a) have also been implicated in secondary growth. Interestingly, gene expression studies have revealed overlapping mechanisms between primary and secondary growth, since homologs of at least some genes expressed in the cambial zone during secondary growth of trees are also involved in establishing/function of the RAM/SAM and the procambium (Schrader *et al.*, 2004; Groover *et al.*, 2006; Baucher *et al.*, 2007; Du and Groover, 2010; Aichinger *et al.*, 2012; Sanchez *et al.*, 2012; Jouannet *et al.*, 2015). This may explain the conservation of secondary growth gene networks in herbaceous plants and the apparent ease of shifting back to the woody growth habit.

We compared gene expression profiles in stem developmental stages in the *soc1ful* woody Arabidopsis mutant with that of wild-type (WT) plants to identify genes involved in various stages of secondary growth, and linked the changes in gene expression with detailed anatomical observations. We also highlighted the expression profiles of genes and associated gene regulatory mechanisms already identified in previous studies (Aichinger *et al.*, 2012; Sanchez *et al.*, 2012; Schuetz *et al.*, 2013; Zhang *et al.*, 2014), and report differentially expressed (DE) genes found during cambium and wood formation as well as their position in the STRING v.10 global Arabidopsis gene association network (Szklarczyk *et al.*, 2015) to identify clusters of co-expressed and interacting genes acting as functional units.

RESULTS

Histology of the developmental stages in *soc1ful* and WT inflorescence stems

Identification of the developmental stages of secondary growth in the Arabidopsis inflorescence shoot was a necessary step before characterizing transcriptome remodeling during interfascicular cambium development and subsequent wood formation. We performed histological analyses of the inflorescence stem and observed that young (63 days after germination) *soc1ful* individuals already produced a large wood cylinder at the base of the main stem (about 20 cell rows), including mainly vessels and fibers, but no rays (Figure 1e) (see Lens *et al.*, 2012; for more details). Also, a closed cambium ring was formed at the base of side shoots (Figure 1d). In contrast, secondary growth was limited to only a few cells within the vascular bundle region at the base of the main inflorescence stem in WT plants (42 days after germination; Figure 1g) (see also Sankar *et al.*, 2014). For RNA sequencing, we selected herbaceous stages where the interfascicular cambium was lacking (Figure 1c for *soc1ful* and Figure 1f for WT), cambium initiation stages showing actively dividing interfascicular cambium cells (Figure 1d for *soc1ful* and Figure 1g for WT), and a woody growth stage including active cambium producing a pronounced wood cylinder (Figure 1e, in *soc1ful* only).

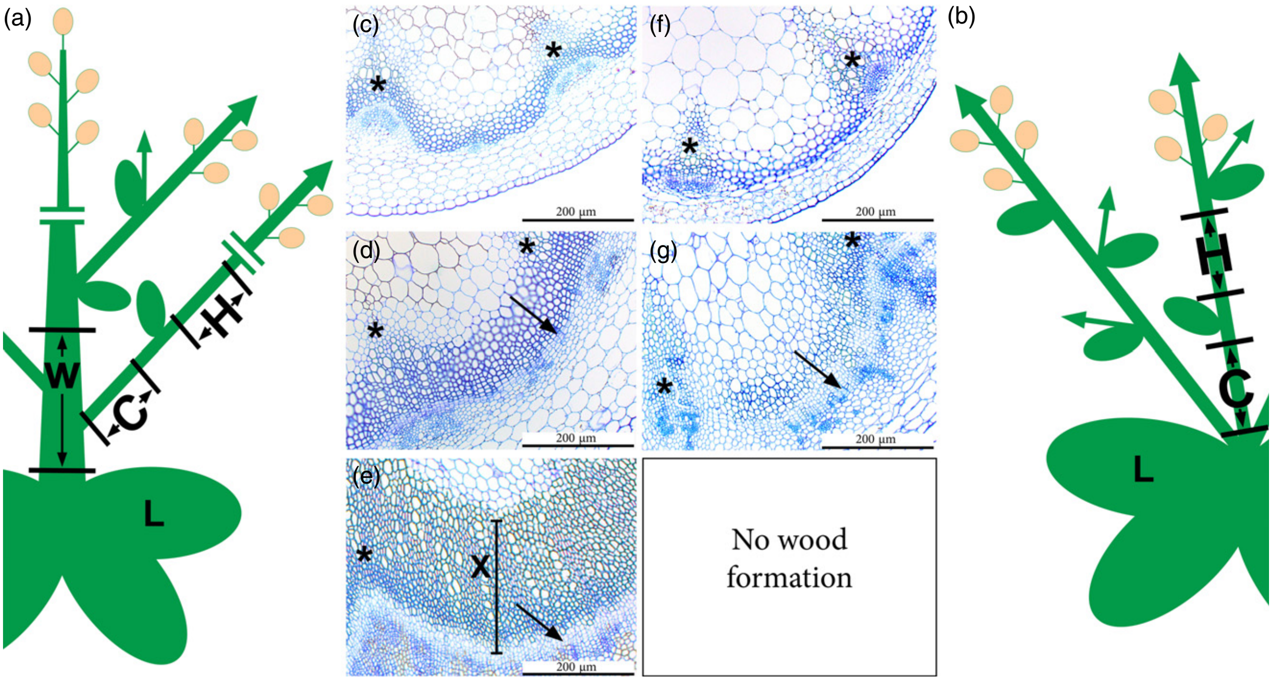


Figure 1. Sampling strategy and cross-sections of *Arabidopsis thaliana soc1-6 ful-7* double mutant and wild-type (WT) stem samples selected for RNA sequencing.
(a) Schematic view of a *soc1ful* plant. Wood stage 'W' refers to the most basal 1.5 cm of the main inflorescence stem, cambium stage 'C' corresponds to the first centimeter of a basal side shoot and herbaceous stage 'H' to the 4–5 cm region from the same shoot. (b) Schematic view of a WT plant. Cambium stage 'C' corresponds to the first centimeter of an inflorescence stem, and herbaceous stage 'H' corresponds to upper stem parts starting from 4 to 5 cm from the base of an inflorescence stem. 'L' corresponds to pooled rosette leaves. (c)–(e) basal side of the 'H', 'C' and 'W' stages in *soc1ful*, respectively. (f), (g) Basal side of the 'H' and 'C' stages in WT, respectively. Arrows indicate a complete vascular cambium ring in (d), (e) and (g). Asterisks indicate vascular bundle regions. 'X' indicates secondary xylem.

Transcript abundance and differential gene expression in *soc1ful* and WT

In order to gain insight into signaling pathways and regulators involved in cambium initiation, cambium activity and wood formation, we identified genes DE during secondary growth. We performed transcriptome profiling using RNA sequencing focusing on two (WT) or three (woody mutant *soc1ful*) stem developmental stages of anatomically identical stem parts, and we also sequenced RNA of pooled rosette leaves from both genotypes (Figure 1).

Of the 34 134 annotated gene models in TAIR10 we obtained detectable expression for 19 033 genes [fragments per kilobase of transcript per million mapped reads (FPKM) > 1 in at least one of the sampled tissues]. Transcript abundance profiles of leaf and stem stages, within each genotype, differed the most (between 3578 and 5627 DE genes; *q*-value <0.05), while stem stage comparisons within genotype were most similar (between 557 and 1389 DE genes; *q*-value <0.05). The numbers of DE genes in all pairwise comparisons of the seven stem and leaf stages are presented in Figure 2(a).

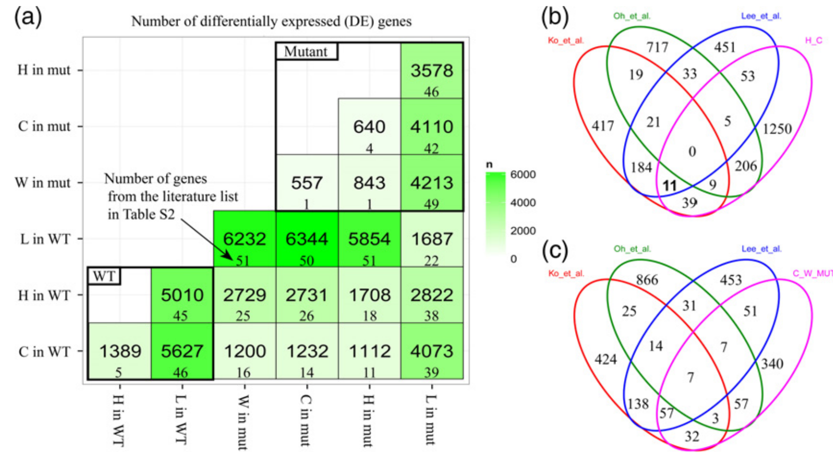
Transcriptome comparisons between stems and leaves

MapMan analysis of metabolism-related DE genes in WT (cambium versus leaf) and *soc1ful* (cambium versus leaf) *A. thaliana*, and in two woody species, *Populus trichocarpa* and *Eucalyptus grandis* (leaf versus stem xylem) are presented in Figure S3 in the Supporting Information (Hefer *et al*., 2015). Similar changes in metabolic activities can be seen in all four plant datasets, with genes preferentially expressed in leaves being involved mainly in photosynthesis (e.g. tetrapyrrole synthesis, light reactions, Calvin cycle and photorespiration), and genes preferentially expressed in xylem/stem representing cell wall polysaccharide and lignin synthesis (Figure S3). Several additional distinctions could be made, for example the xylem-specific investment in synthesis of nitrogen- and sulfur-containing glucosinolates in the stem tissue of Arabidopsis (Figure S3a,b), the expression of leaf-specific family members involved in phenylpropanoid synthesis in *Populus* and *Eucalyptus* (Figure S3c,d) and the expansion and redundancy in xylem-specific galactinol synthase (GolS) genes in *Eucalyptus* (Figure S3d) and laccase genes in *Populus* (Figure S3d).

Transcriptome remodeling during stem development

In order to characterize transcriptome remodeling during secondary growth, we first selected the DE genes between the herbaceous (H) and the cambium (C) stage in either the WT and/or *soc1ful* inflorescence stem to identify candidate

Figure 2. Number and overlap of differentially expressed (DE) genes in *Arabidopsis thaliana soc1-6 ful-7* double mutant (mut) and wild-type (WT) tissues.
(a) Matrix showing number of DE genes in the seven *A. thaliana* pairwise sample comparisons; including, in subscript, reviewed genes involved in meristem activity and xylem development (Aichinger *et al.*, 2012; Sanchez *et al.*, 2012; Schuetz *et al.*, 2013; Zhang *et al.*, 2014) (Table S2, Figure S4). H, herbaceous; C, cambium; W, wood; L, leaf.
(b), (c) The overlap of DE genes during cambium formation (H versus C) in both genotypes (b) and in wood formation (C versus W) (c), compared with published transcriptome datasets from plants responding to mechanical stress (Oh *et al.*, 2003; Ko *et al.*, 2004; Lee *et al.*, 2005).



genes potentially involved in cambium initiation and formation. Secondly, we selected the DE genes between the cambium (C) and wood (W) stages in *soc1ful* to reveal candidate genes related to wood formation in Arabidopsis. Only DE genes with two-fold change were used in further analyses (Table S4) to reduce the number of false positives in our analyses. We then screened for potential regulators of secondary growth such as transcription factors, genes involved in hormone signaling and biosynthesis as well as lignification, cell wall formation and flowering time (Table S4).

We compared genes that were DE during cambium (C versus H) and wood formation (W versus C) with previously published stem microarray expression profiling studies that applied stress by removing flower buds (Oh *et al.*, 2003), or bending the stem (Ko *et al.*, 2004) or the leaves (Lee *et al.*, 2005) (Figure 2b,c, Table S4). During cambium initiation, we found a higher proportion of genes shared with the Oh *et al.* (2003) study compared with the Ko *et al.* (2004) and Lee *et al.* (2005) mechanical stress gene lists (21.9% versus 8.6% and 8.2%, respectively). However, during wood formation, there was a higher overlap between our dataset and the Ko *et al.* and Lee *et al.* gene lists compared with the Oh *et al.* study (18.0% and 22.2% versus 13.5%, respectively) (Figure 2b,c).

Transcriptome remodeling during cambium formation

The DE genes between the C and H stages in both WT and *soc1ful* are listed in Table S4. Based on candidate genes already mentioned in the literature and specific expression profiles related to transcription factors, and regulation and synthesis of secondary cell walls and hormones, we further refined our selection of genes putatively involved in cambium formation and activity (Table 1). We found changes in multiple hormone-related genes, especially genes involved in cytokinin, jasmonate (JA) and ethylene signaling. The DE cytokinin genes *CKX1* and *Cyp79B2* (Werner *et al.*, 2003) are upregulated in the cambium samples. The same applies to the three JA regulators *JAZ5*, *ATST2A* and *IAR3* (Kazan and Manners, 2012), the JA-responsive genes *NATA1* (Adio *et al.*, 2011) and *WRKY51* (Gao *et al.*, 2011) and the ethylene related genes *ETR2*, *ADH*, *JAL30* and *At4G27450* (Liu and Wen, 2012). Likewise, the expression of the JA–ethylene integrator *ORA59* (Pre *et al.*, 2008) and the JA–ethylene responsive gene *AthCHIB* (Zander *et al.*, 2010) were upregulated. The brassinosteroid-induced gene *ARL* (Hu *et al.*, 2006) was also upregulated. Moreover, there appears to be a link between cambium formation in stems and meristematic activity in roots, as evidenced by a number of upregulated genes such as *AtMYB38*, *-68* (Dubos *et al.*, 2010) and *LRP1* (Krichevsky *et al.*, 2009) that are associated with root meristem activity, and *BRC2* (Aguilar-Martinez *et al.*, 2007), associated with meristematic activity in stems. One flowering time gene, *GATA21*, was downregulated during cambium formation in both genotypes. In addition, genes encoding two key enzymes involved in providing UDP-glucose for cellulose and xylan biosynthesis, *SUS3* and *SUS4*, were preferentially expressed during cambium formation (Baroja-Fernandez *et al.*, 2012). *LTP1* and *LTP2*, which are attributed to cell wall loosening, are upregulated (Chae *et al.*, 2010). Finally, three hydroxyproline-rich glycoproteins (HRGP), *PELPK1* (Rashid and Deyholos, 2011), AT5G09480 and AT4G38080, were preferentially expressed in the cambium samples compared with herbaceous developmental stages, especially in the mutant (>5 $\log_2$-fold in *soc1ful*; Table 1).

There were also a number of DE genes associated with cambium formation and activity in only one of the genotypes, especially in WT (Table S4). Based on candidate genes already mentioned in the literature and specific expression profiles related to transcription factors, and regulation and synthesis of secondary cell walls and hormones, we further refined our selection (Table 2). Again, there was a strong differential gene expression from

Table 1 Selected genes differentially expressed during cambium formation in *Arabidopsis thaliana soc1ful* mutant (Mut) and wild-type (WT) plants (full list available in Table S4)

		Transcript abundance (FPKM)							Differentially expressed - Log2-fold							
gene ID	symbol TAIR10	L_WT	H_WT	C_WT	L_Mut	H_Mut	C_Mut	W_Mut	C vs L in WT	C vs L in Mut	C vs H in WT	C vs H in Mut	W vs C in Mut	Ko *et al.* (2004)	Oh *et al.* (2003)	Lee *et al.*, (2005)
Hormone - Cytokinin																
AT2G41510	CKX1	2.75	2.80	10.87	7.11	3.40	21.41	8.54	1.98	1.59	1.96	2.66	−1.33			
AT4G39950	CYP79B2	13.51	19.87	64.68	13.00	23.64	54.27	59.85	2.26	2.06	1.70	1.20	ns			
Hormone - Jasmonate																
AT5G07010	ST2A	0.94	3.38	63.85	2.02	2.77	13.95	10.96	6.09	2.78	4.24	2.33	ns			
AT1G17380	JAZ5	15.85	25.32	63.12	17.98	19.73	50.72	188.71	1.99	1.50	1.32	1.36	1.90	7.2		4.72
AT1G51760	IAR3	30.71	40.17	89.34	22.25	24.44	49.23	49.17	1.54	1.15	1.15	1.01	ns			
AT2G39030	NATA1	30.60	20.05	462.80	36.55	8.04	101.47	88.31	3.92	1.47	4.53	3.66	ns			
AT5G64810	WRKY51	9.78	1.52	8.06	7.71	3.14	16.30	14.54	ns	ns	2.41	2.38	ns			
Hormone - Jasmonate/Ethylene																
AT3G12500	HCHIB	3.92	2.86	208.28	18.98	29.47	639.03	275.09	5.73	5.07	6.19	4.44	ns			
AT1G06160	ORA59	0.63	2.72	9.81	3.19	4.26	13.87	9.80	3.97	2.12	1.85	1.70	ns			
Hormone - Ethylene																
AT3G23150	ETR2	4.44	6.77	27.35	6.01	7.55	24.83	29.65	2.62	2.05	2.01	1.72	ns			
AT1G77120	ADH	14.55	28.39	112.15	23.09	44.58	129.56	155.78	2.95	2.49	1.98	1.54	ns			
AT3G16420	JAL30	8.01	45.94	150.49	22.25	51.62	208.06	109.91	4.23	3.23	1.71	2.01	ns			
AT4G27450	NA	35.45	26.50	68.50	25.56	6.84	37.64	47.76	ns	ns	1.37	2.46	ns			
Hormone - Brassinosteroid																
AT2G44080	ARL	21.47	12.71	37.96	32.06	7.32	17.91	42.76	ns	ns	1.58	1.29	1.26			
Meristem (cell division, expansion, meristem identity)																
AT5G65790	MYB68	0.00	0.31	3.94	0.70	3.08	8.44	3.71	Inf	3.59	3.66	1.45	ns			
AT2G36890	MYB38 (RAX2)	0.31	0.76	3.77	2.87	3.17	15.76	7.29	3.61	2.46	2.31	2.31	ns			
AT5G12330	LRP1	0.05	0.30	2.63	0.45	1.95	7.52	6.06	5.71	4.08	3.11	1.95	ns			
AT1G68800	BRC2 (TCP12)	0.07	0.43	2.52	0.57	0.78	3.42	1.78	ns	2.60	2.55	2.14	ns			
Flowering time																
AT5G56860	GATA21	40.99	10.88	1.51	22.44	1.85	0.27	1.87	−4.76	−6.37	−2.85	−2.77	2.78			
Cell wall assembly																
AT4G02280	SUS3	16.1	11.0	26.7	26.9	26.1	67.5	51.1	ns	1.33	1.27	1.37	ns			
AT3G43190	SUS4	0.0	0.8	19.8	1.7	1.9	14.3	41.1	ns	3.05	4.67	2.90	1.52			
AT2G38540	LTP1	1335.14	50.10	320.60	1167.25	57.30	173.67	182.39	−2.06	−2.75	2.68	1.60	ns			
AT2G38530	cdf3 (LTP2)	12.81	14.74	88.90	62.14	83.19	309.37	36.33	2.79	2.32	2.59	1.89	−3.09			
Miscellaneous																
AT5G09480	NA	0.1	0.0	0.9	1.2	0.2	14.5	5.6	ns	3.65	Inf	5.86	ns			
AT5G09530	PELPK1	0.6	0.2	1.5	1.6	0.4	30.5	8.0	ns	4.24	3.17	6.12	−1.94			
AT4G38080	NA	0.7	1.4	7.1	7.9	1.7	163.5	36.4	3.33	4.37	2.32	6.57	−2.17			

L, leaf; H, herbaceous; C, cambium; W, wood; vs, versus; FPKM, fragments per kilobase of transcript per million mapped reads; NA, not available; Inf, infinite.
'ns' means non-significant at the threshold we applied (q-value < 0.05 and greater than two-fold change).
For each gene, relative expression per sample (FPKM) is given; colors for the FPKM values range from pale yellow (low expression) to dark red (high expression). Log_2-fold changes are given for the genes also differentially expressed in previous studies (Oh *et al.*, 2003; Ko *et al.*, 2004; Lee *et al.*, 2005).

specific hormone pathways. The DE genes included those involved in cytokinin (upregulation of *CHS*, *ARGOS*, *LOG5* and *DRM2*), JA (upregulation of *AOC3*, *JAZ4*, *AtLOX2*, *TT8*, *ERF1*, *-13*, *-15*, *-104* and *ERF-1*, downregulation of *FAD7*), and ethylene signaling (upregulation of *ERF11*, *RAP2.2*, *GAPCP-1*, *PCK1*). Genes related to gibberellin action were also DE (upregulation of *MYB62*, *ABC33*, *RGL3*, *RGL*; downregulation of *2301* and *MGP*), and *RGL* and *MGP* also play a role in cell division/proliferation (Yoshida and Ueguchi-Tanaka, 2014). In addition, six genes involved in SAM function/formation (upregulation of *CLE42*, *RTFL13*, *PDF1*, *ATHB-2*; downregulation of *TCP2* and *HAT1*) were DE in the cambium stage, predominantly in WT. Other genes which are reported to be involved in meristematic activity or cell division in roots (upregulation of *AGL12*, *AtHSFB4*, *AtHB53*), leaves (upregulation of *AN3*) or stems (upregulation of *BRC1*) (Aguilar-Martinez *et al.*, 2007) were also DE. In addition to the flowering time genes *RTFL13* and *AN3*, there was an upregulation of *DVL3* and a downregulation of *FT*, *SPL9*, *-10* and *-11*, and *TRY* in WT. Finally, genes related to cell wall assembly and modification (upregulation of *TBL36*, *AtTEXT3*, downregulation of *TBL38*, *-40*, *GXMT1*, *TEXP15*, *FLA12*) were DE in either the cambium of WT or *soc1ful*.

Transcriptome remodeling during wood formation in *soc1ful*

Comparing the transcriptome profiles in the cambium and wood stage in *soc1ful* allowed us to identify DE genes during wood formation in the mutant (Table S4). Based on candidate genes already mentioned in the literature and specific expression profiles related to transcription factors, and regulation and synthesis of secondary cell walls and hormones, we further refined our selection of genes to those DE during wood formation (Table 3). In this analysis we again observed strong differential gene expression relating to phytohormone biosynthesis and signaling pathways. For instance, we identified DE genes involved in the cytokinin signaling pathway (downregulation of *CKX1*,

Table 2 Selected genes differentially expressed during cambium formation in either *A. thaliana soc1ful* mutant (Mut) or wild-type (WT) plants (full list available in Table S4)

		FPKM							Differential expression - Log-2-fold change							
gene_ID	symbol TAIR10	L_WT	H_WT	C_WT	L_mut	H_mut	C_mut	W_mut	C vs L in WT	Cvs L in Mut	C vs H in WT	C vs H in Mut	W vs C in mut	Ko *et al.* (2004)	Oh *et al.* (2003)	Lee *et al.* (2005)
Hormone - Cytokinin																
AT5G13930	CHS (tt4)	60.13	15.60	582.57	75.92	117.35	163.13	211.80	3.28	1.10	5.22	ns	ns		−3.67	
AT3G59900	ARGOS	2.70	1.25	5.18	0.83	1.37	2.28	1.45	ns	ns	2.06	ns	ns			
AT4G35190	LOG5	2.26	1.76	5.80	8.66	17.23	31.73	12.65	ns	1.87	1.72	ns	−1.33			
AT2G33830	DRM2	97.46	177.69	396.60	115.86	135.60	275.83	194.11	2.02	1.25	1.16	ns	ns		4.03	
Hormone - Jasmonate																
AT3G25780	AOC3	5.20	16.11	36.79	10.53	9.93	18.41	139.82	2.82	ns	1.19	ns	2.92	3.1		
AT1G48500	JAZ4	1.47	3.76	10.69	2.77	7.78	11.70	15.30	2.86	2.08	1.51	ns	ns			
AT3G11170	FAD7	112.24	47.05	18.26	68.41	12.03	10.20	14.65	−2.62	−2.75	−1.37	ns	ns		−1.23	
AT3G45140	LOX2	658.69	14.75	17.02	596.89	3.56	11.83	77.93	−5.27	−5.66	ns	1.73	2.72			
AT4G09820	TT8	0.03	0.29	8.81	1.04	1.54	2.23	1.59	ns	ns	4.92	ns	ns			
AT2G44840	ERF13	0.42	0.98	4.99	5.06	0.40	1.84	76.72	3.57	ns	2.34	ns	5.38	4.8		
AT3G23240	ERF1	4.00	2.23	10.02	2.30	2.12	3.95	6.38	ns	ns	2.17	ns	ns			
AT2G31230	ERF15	4.28	4.32	13.73	24.06	19.92	23.51	24.29	1.68	ns	1.67	ns	ns		1.09	
AT4G17500	ERF-1	27.50	19.92	49.14	19.99	5.97	13.21	84.60	ns	ns	1.30	ns	2.68	5.3		
AT5G61600	ERF104	39.43	39.56	84.51	36.27	10.95	13.85	215.93	1.10	−1.39	1.10	ns	3.96	8.5		2.26
Hormone - Ethylene																
AT1G28370	ERF11	21.17	14.34	42.65	14.18	8.09	13.56	62.55	ns	ns	1.57	ns	2.21			
AT3G14230	RAP2.2	37.70	178.83	443.70	68.99	203.27	273.58	281.07	3.56	1.99	1.31	ns	ns		1.79	
AT1G79530	GAPCP-1	2.77	13.62	47.39	6.32	18.13	33.53	43.29	4.09	2.41	1.80	ns	ns			
AT4G37870	PCK1	55.95	202.73	655.81	166.85	486.33	932.40	1017.70	3.55	2.48	1.69	ns	ns		3.64	
Hormone - Gibberelin																
AT4G25420	2301 (GA20ox1)	0.91	11.32	4.52	4.24	15.69	14.00	23.89	2.32	1.72	−1.32	ns	ns			
AT4G02780	ABC33	0.12	0.60	1.08	0.36	0.54	2.06	2.16	3.12	2.53	ns	1.92	ns			
AT5G17490	RGL3	2.43	1.54	3.09	2.58	1.43	4.23	2.02	ns	ns	ns	1.57	ns			
AT1G68320	MYB62	0.27	0.70	3.40	1.20	3.36	8.57	2.30	3.67	2.84	2.27	ns	−1.90			
Meristem (cell division, expansion, meristem identity)																
AT1G03840	MGP	0.90	0.30	0.14	0.96	2.17	0.50	0.43	−2.73	ns	ns	−2.13	ns			
AT1G66350	RGL	6.58	6.38	19.38	11.76	15.85	26.42	37.26	1.56	1.17	1.60	ns	ns			
AT5G66700	ATHB53	2.41	0.64	4.89	3.40	4.79	9.26	0.71	ns	ns	2.93	ns	−3.71			
AT4G16780	ATHB-2	31.01	23.80	50.72	63.15	62.51	83.82	81.02	ns	ns	1.09	ns	ns		2.32	
AT4G17460	HAT1	7.27	9.54	2.52	10.77	5.65	5.05	5.55	−1.53	ns	−1.92	ns	ns			
AT1G46264	HSFB4	0.96	4.05	13.17	4.34	8.90	16.39	25.04	3.77	1.92	1.70	ns	ns			
AT1G71692	AGL12 (XAL1)	0.15	3.24	9.17	0.56	2.52	3.67	2.91	5.92	2.72	1.50	ns	ns			
AT2G34925	CLE42	0.00	0.00	4.30	2.65	9.17	7.62	9.26	Inf	ns	Inf	ns	ns			
AT2G42840	PDF1	9.26	0.31	2.96	19.39	0.88	9.66	2.43	ns	ns	ns	3.46	−1.99			
AT3G18550	BRC1 (TCP18)	1.98	0.46	2.42	1.01	2.59	4.44	0.18	ns	2.14	2.39	ns	−4.61			
AT4G18390	TCP2	28.61	3.45	0.99	30.62	2.22	1.43	3.11	−4.86	−4.42	−1.81	ns	ns			
Flowering time																
AT5G28640	AN3	1.50	0.30	1.50	1.40	0.20	3.60	0.30	ns	ns	ns	3.89	−3.41			
AT1G27360	SPL11	19.00	16.49	7.17	15.80	10.64	8.16	8.16	−1.41	ns	−1.20	ns	ns			
AT2G42200	SPL9	16.75	11.12	4.06	11.20	8.06	3.96	2.82	−2.05	−1.50	−1.45	ns	ns		−1.30	
AT1G27370	SPL10	10.80	19.49	6.51	12.16	14.50	9.72	8.82	ns	ns	−1.58	ns	ns			
AT5G53200	TRY	35.52	18.58	0.59	17.31	0.89	0.29	1.99	−5.92	−5.90	−4.98	ns	ns			
AT1G65480	FT	142.94	23.48	7.12	13.77	1.51	1.68	3.32	−4.33	−3.04	−1.72	ns	ns			
AT3G23635	RTFL13	0.00	0.00	65.31	0.00	0.00	91.52	22.31	ns	Inf	ns	Inf	ns			
Cell wall assembly																
AT5G60490	FLA12	0.68	162.88	119.70	23.69	170.12	64.90	91.75	7.47	1.45	ns	−1.39	ns			
AT1G33800	GXMT1	1.00	117.70	70.90	13.20	88.10	33.60	35.00	6.08	1.35	ns	−1.39	ns			
AT3G54260	TBL36	1.60	37.40	143.50	16.40	105.70	100.40	142.00	6.52	2.62	1.94	ns	ns			
AT2G31110	TBL40	17.10	71.90	35.00	54.80	69.70	59.30	56.60	1.04	ns	−1.04	ns	ns			
AT1G29050	TBL38	173.90	54.70	15.60	37.30	25.80	7.40	5.90	−3.48	−2.34	−1.81	ns	ns			
AT1G21310	ATEXT3	66.03	150.87	2073.59	205.37	736.76	5256.97	1814.94	4.97	ns	3.78	ns	ns			
AT2G03090	ATEXP15	4.65	27.82	15.90	6.41	68.38	9.88	15.32	1.78	ns	ns	−2.79	ns			

L, leaf; H, herbaceous, C, cambium, W, wood; vs, versus; FPKM, fragments per kilobase of transcript per million mapped reads.
'ns' means non-significant at the threshold we applied (*q*-value < 0.05 and greater than two-fold change); Inf, infinite.
For each gene, relative expression per sample (FPKM) is given; colors for the FPKM values range from pale yellow (low expression) to dark red (high expression). Log_2-fold changes are given for the genes also differentially expressed in previous studies (Oh *et al.*, 2003; Ko *et al.*, 2004; Lee *et al.*, 2005).

LOG5, *LEA4*-5), the auxin pathway (upregulation of *PBP1*, *SAUR72*, *AXR5*, downregulation of *LAX3*), the JA pathway (upregulation in *AOC1*, *-2*, *-3*, *AOS*, *LOX2*, *-3*, *-4*, *MYC2*, *ORA47*, *ERF-1*, *-5*, *-6*, *-13*, *ERF104*, *JAZ1*, *-2*, *-5*, *-6*, *-7*, *-9*, *-10*, *WRKY70*, downregulation in *LOX1* and *AIB*; Wasternack, 2007; Wasternack and Hause, 2013), the ethylene pathway (upregulation in *ERF11*, *MKK9*), and the gibberellin pathway (downregulation of *GA2OX2*; Israelsson *et al.*, 2005). In addition, *JKD* is linked to cell division (downregulated; Yoshida and Ueguchi-Tanaka, 2014), *EXO* is linked to cell expansion (upregulated; Schröder *et al.*, 2009), and *REM1* is normally expressed in the SAM (upregulated; Franco-Zorrilla *et al.*, 2002). Flowering time genes DE in the wood stage were *AGL3*, *DVL10*, *ZPR1*, *RAV2* and *EDF1* (all upregulated) and *AGL72* (downregulated). We also identified several DE genes involved in cell wall

Table 3 Selected genes differentially expressed during wood formation in *A thaliana soc1ful* plants (full list available in Table S4)

		FPKM							Differentially expressed - Log-2-fold							
gene ID	symbol TAIR10	L_WT	H_WT	C_WT	L_Mut	H_Mut	C_Mut	W_Mut	C vs L in WT	C vs L in Mut	C vs H in WT	C vs H in Mut	W vs C in Mut	Ko *et al.* (2004)	Oh *et al.* (2003)	Lee *et al.* (2005)
Hormone - Auxin																
AT5G54490	PBP1	23.10	12.88	25.93	13.66	3.15	5.79	79.23	ns	ns	ns	ns	3.78			1.90
AT3G12830	SAUR72	14.09	6.52	14.16	6.80	8.71	5.13	14.69	ns	ns	ns	ns	1.52			
AT1G77690	LAX3	0.72	24.71	21.36	4.03	32.09	21.38	8.06	4.89	2.41	ns	ns	–1.41			
AT4G14560	AXR5 (IAA1)	89.43	33.86	20.55	49.93	26.41	30.18	60.58	–2.12	ns	ns	ns	1.01			
Hormone - Cytokinin																
AT2G41510	CKX1	2.75	2.80	10.87	7.11	3.40	21.41	8.54	1.98	1.59	1.96	2.66	–1.33			
AT4G35190	LOG5	2.26	1.76	5.80	8.66	17.23	31.73	12.65	ns	1.87	1.72	ns	–1.33			
AT5G06760	LEA4-5	98.6	35.6	120.9	40.0	28.7	63.0	16.8	ns	ns	1.76	1.13	–1.91			
Hormone - Jasmonate																
AT3G25780	AOC3	5.20	16.11	36.79	10.53	9.93	18.41	139.82	2.82	ns	1.19	ns	2.92	3.1		
AT5G42650	AOS	114.99	30.88	54.25	257.11	59.76	47.31	129.98	–1.08	–2.44	ns	ns	1.46			1.34
AT3G25770	AOC2	134.62	25.35	19.29	232.70	11.25	9.05	27.62	–2.80	–4.68	ns	ns	1.61			
AT3G25760	AOC1	31.99	19.81	11.03	160.81	11.37	6.39	18.93	–1.54	–4.65	ns	ns	1.57			
AT1G19180	JAZ1	69.30	31.48	57.57	25.31	14.66	20.90	133.88	ns	ns	ns	ns	2.68			3.99
AT5G13220	JAS1 (JAZ10)	1.45	6.62	17.03	0.87	1.13	3.53	18.31	3.56	ns	ns	ns	2.37			3.27
AT1G17380	JAZ5	15.85	25.32	63.12	17.98	19.73	50.72	188.71	1.99	1.50	1.32	1.36	1.90	7.2		4.72
AT2G34600	JAZ7	3.05	1.99	6.30	1.08	0.40	0.77	8.04	ns	ns	ns	ns	3.38	9.6		
AT1G74950	JAZ2	21.31	28.10	48.59	22.63	20.43	30.97	71.25	1.19	ns	ns	ns	1.20			2.32
AT3G45140	LOX2	658.69	14.75	17.02	596.89	3.56	11.83	77.93	–5.27	–5.66	ns	1.73	2.72			
AT1G72520	LOX4	7.90	3.04	3.29	3.53	2.73	2.23	6.34	–1.26	ns	ns	ns	1.51	9.8		
AT1G17420	LOX3	8.05	2.12	3.06	11.20	4.42	3.41	9.03	–1.39	–1.71	ns	ns	1.40			5.53
AT1G70700	JAZ9	342.69	308.70	222.07	148.36	54.60	83.73	210.52	ns	ns	ns	ns	1.33			1.67
AT1G72450	JAZ6	85.18	59.85	94.51	89.03	38.75	60.44	149.73	ns	ns	ns	ns	1.31	2		2.18
AT1G55020	LOX1	18.26	21.00	21.47	11.30	8.88	16.94	7.98	ns	ns	ns	ns	–1.09			
AT1G32640	MYC2	63.61	89.76	153.42	73.82	58.54	67.75	240.07	1.27	ns	ns	ns	1.83	4.5		2.75
AT3G56400	WRKY70	106.73	51.38	82.25	68.52	22.85	47.19	132.23	ns	ns	ns	1.05	1.49	2.9		
AT1G74930	ORA47	6.29	11.96	29.79	10.57	0.93	0.95	143.73	2.24	–3.48	1.32	ns	7.24			
AT2G44840	ERF13	0.42	0.98	4.99	5.06	0.40	1.84	76.72	3.57	ns	2.34	ns	5.38	4.8		
AT4G17490	ERF6	2.60	1.06	3.11	3.13	0.60	0.96	15.83	ns	ns	ns	ns	4.05			
AT5G61600	ERF104	39.43	39.56	84.51	36.27	10.95	13.85	215.93	1.10	–1.39	1.10	ns	3.96	8.5		2.26
AT5G47230	ERF-5	6.18	8.14	15.96	6.62	3.68	4.53	35.50	1.37	ns	ns	ns	2.97	7.2		
AT4G17500	ERF-1	27.50	19.92	49.14	19.99	5.97	13.21	84.60	ns	ns	1.30	ns	2.68	5.3		
AT2G46510	AIB	16.25	13.24	17.92	11.50	8.04	7.90	18.92	ns	ns	ns	ns	1.26			2.21
Hormone - Ethylene																
AT1G28370	ERF11	21.17	14.34	42.65	14.18	8.09	13.56	62.55	ns	ns	1.57	ns	2.21			
AT1G73500	MKK9	50.74	47.61	82.65	44.39	22.42	27.74	95.21	ns	ns	ns	ns	1.78			1.34
Hormone - Gibberellin																
AT1G30040	GA2OX2	0.98	6.95	6.13	4.47	4.48	7.36	2.23	2.65	ns	ns	ns	–1.72	3.3		
Meristem (cell division, expansion, meristem identity)																
AT4G31610	REM1	0.02	2.09	2.92	0.52	2.44	3.02	9.96	ns	2.54	ns	ns	1.72			
AT4G08950	EXO	0.84	0.31	1.10	0.90	0.49	0.73	4.06	ns	ns	ns	ns	2.47	4.7		2.86
Flowering time																
AT2G03710	AGL3 (SEP4)	81.57	32.09	26.27	68.17	24.30	14.87	67.10	–1.63	–2.20	ns	ns	2.17			
AT5G51860	AGL72	0.04	0.00	0.00	1.49	2.69	6.50	0.78	ns	2.12	ns	ns	–3.05			
AT2G45450	ZPR1	4.68	29.16	51.20	18.81	45.19	58.46	134.19	3.45	1.64	ns	ns	1.20			
AT1G68840	RAV2 (TEM2)	10.81	0.67	3.03	59.11	4.43	3.90	10.38	–1.83	–3.92	2.19	ns	1.41			1.82
AT1G25560	EDF1	25.30	7.86	11.40	37.30	5.66	4.20	10.33	–1.15	–3.15	ns	ns	1.30			
AT4G13395	DVL10	17.4	16.2	32.6	18.5	4.3	4.7	182.0	ns	ns	ns	ns	5.27			
Cell wall assembly																
AT2G04160	AIR3	33.15	3.77	1.85	19.20	0.59	0.16	1.54	–4.16	–6.90	ns	ns	3.26			
AT3G28340	GATL10	1.5	0.8	1.5	0.8	0.3	0.3	8.2	ns	ns	ns	ns	4.83			
AT1G70090	GATL9	2.7	3.8	5.3	3.1	3.1	2.5	13.7	ns	ns	ns	ns	2.45	1.5		1.58
AT3G43190	SUS4	0.0	0.8	19.8	1.7	1.9	14.3	41.1	ns	3.05	4.67	2.90	1.52			
AT5G57560	TCH4	29.4	46.1	54.8	6.9	2.7	2.0	71.9	ns	–1.77	ns	ns	5.15	34.6		4.40
AT4G37980	CAD7	140.7	3.2	0.6	202.2	1.8	0.2	14.8	–7.99	–9.76	–2.53	–2.94	5.99			
Plant/inflorescence architecture																
AT1G73870	BBX16	191.87	58.22	21.50	245.25	28.31	6.41	22.88	–3.16	–5.26	–1.44	–2.14	1.84			
AT1G16410	BUS1	283.55	1095.66	1062.66	335.99	803.02	350.95	99.32	1.91	ns	ns	ns	–1.82		1.47	
AT5G03840	TFL-1	1.2	3.3	10.1	1.3	3.3	5.0	0.4	3.06	ns	ns	ns	–3.50			

L, leaf; H, herbaceous, C, cambium, W, wood.; WT, wild type; Mut, mutant; FPKM, fragments per kilobase of transcript per million mapped reads.

'ns' means non-significant at the threshold we applied (q-value <0.05 and greater than two-fold change).

For each gene, relative expression per sample (FPKM) is given; colors for the FPKM values range from pale yellow (low expression) to dark red (high expression). Log_2-fold changes are given for the genes also differentially expressed in previous studies (Oh *et al.*, 2003; Ko *et al.*, 2004; Lee *et al.*, 2005).

assembly including *TCH4*, *AIR3*, *SUS4*, *GATL9*, *-10* and *CAD7* (all upregulated; Eudes *et al.*, 2006; Kong *et al.*, 2011), and DE genes that may play roles in the altered inflorescence architecture of *soc1ful*, such as *BBX16* (upregulated) and *BUS1* and *TFL-1* (both downregulated).

Gene community structure in cambium and wood formation networks

To investigate the properties of the gene regulatory network during secondary growth we mapped the lists of DE genes during cambium and wood formation to the global Arabidopsis protein association network (STRING v.10). During both cambium and wood formation gene association networks of DE genes are significantly more clustered than expected by chance in similar sized STRING networks: we found significantly more connected nodes and edges, higher network connectivity and average degree (edges per node), but lower average path length and number of isolated networks than expected by chance, especially during wood formation (Figure 3a,c; *P*-value $<1 \times 10^{-6}$ for each statistic for both networks). The cambium formation network consists of four isolated networks with more than three nodes; three networks have only three nodes each and the larger one has 1280 nodes (each node representing a gene product). The networks are composed of groups of nodes more connected with each other than the rest (i.e. communities). The main cambium formation network is composed of 110 communities, of which the 15 largest ones are visualized in Figure 3(b). The wood formation network includes three isolated networks with more than three nodes; two small networks with three and five nodes and a main network with 392 nodes, including 43 communities of which the 10 largest are indicated in Figure 3(d). Community analyses in both networks reveal enriched functionally related components. The five largest communities have identical annotations, namely photosynthesis, kinase, transcription factor activity/regulation, defense response and lignin/phenylpropanoid biosynthesis (Figure 3b,d).

DISCUSSION

Expanding our ideas to explain the repeated genetic switch towards secondary woodiness in Brassicaceae: a global network hypothesis

Several hundred Brassicaceae species have evolved into small trees (sub-)shrubs or woody lianas, representing multiple convergent evolutionary shifts from herbaceousness towards derived woodiness (Al-Shehbaz, 1984). This high number of transitions in Brassicaceae – and angiosperms as a whole – in combination with the *soc1ful* two-gene model in Arabidopsis suggests that the molecular change during the transition from primary to secondary growth requires relatively few genetic changes. The classical hypothesis for the genetic basis behind the repeated shifts towards woodiness involves only a few genes regulating the entire hierarchical wood developmental pathway, i.e. master switches (Groover, 2005; Spicer and Groover, 2010). Several genes have been found to play a role during secondary growth (Tables S2 and S3), but the search for master switches is still ongoing. In our search for potential master switches that are DE during secondary growth, we encountered gene association networks containing functionally enriched hubs. Our analyses revealed that gene association networks of DE genes involved in cambium and wood formation are modular and highly interconnected and consist of many types of dosage-sensitive genes, including transcription factors and components of signal transduction pathways (Figure 3). Consequently, a modification in any number of these dosage-sensitive genes could result in a stoichiometric imbalance in the network and could affect a number of developmental processes such as wood formation (see reviews on gene and genomic balance; Birchler and Veitia, 2007, 2012). In other words, plasticity in habit shifts could also be explained by any genetic changes modulating a complex global network, rather than just changes in a master gene regulating a hierarchical network. To elaborate on this network hypothesis, more experimental work is needed to investigate the nature and direction of each interaction within the gene network (Taylor-Teeples *et al.*, 2015).

Similarities in molecular pathways between primary and secondary growth

Why have herbaceous species conserved the gene functions necessary for secondary growth? Firstly, most herbaceous species do produce a limited amount of wood at the base of their stems, largely for mechanical reasons (Lens *et al.*, 2012). Secondly, genetic networks in primary growth (primary meristems such as procambium, SAM, RAM) and secondary growth (vascular cambium) are highly similar (Jouannet *et al.*, 2015). For example, it has been shown that the class-I knotted homeobox (KNOX) genes *BP* and *STM*, the two class III homeodomain-leucine zipper (HD-ZIP) genes *PHAVOLUTA/PHABULOSA* and *ATHB-15* (*CORONA*) and *KANADI1* and *MP/ARF5*, are known to function in both primary meristems and the vascular cambium. Moreover, related genes such as *CLV3-CLE41/44* and *WUS-WOX4/14* are either active in the SAM or the vascular cambium (Schrader *et al.*, 2004; Groover, 2005; Groover *et al.*, 2006; Spicer and Groover, 2010; Aichinger *et al.*, 2012).

The gene expression data presented in this study suggest that additional shared candidate genes are involved in both primary and secondary meristems. Two genes upregulated during cambium development in both WT and *soc1-ful* are *MYB68* – expressed throughout the root pericycle and especially in cells that giving rise to lateral root meristems (Feng *et al.*, 2004) – and *PDF1*, which is expressed in both SAM and lateral root primordia (Abe *et al.*, 2001) (Table 1). In addition, Table 2 shows three genes, *CLE42*, *ATHB-2* and *HAT1*, all of which preferentially expressed during cambium formation in the WT inflorescence stems. *CLE42* is known to be expressed in SAM and axillary shoot meristems, and potentially regulates apical dominance and organ size (Strabala *et al.*, 2006; Yaginuma *et al.*, 2011). The other two genes are members of the class II HD-ZIP family: *ATHB2* is expressed in the procambium cells of embryos and during vascular development in the seedling, but also in RAM and SAM of mature embryos (Turchi

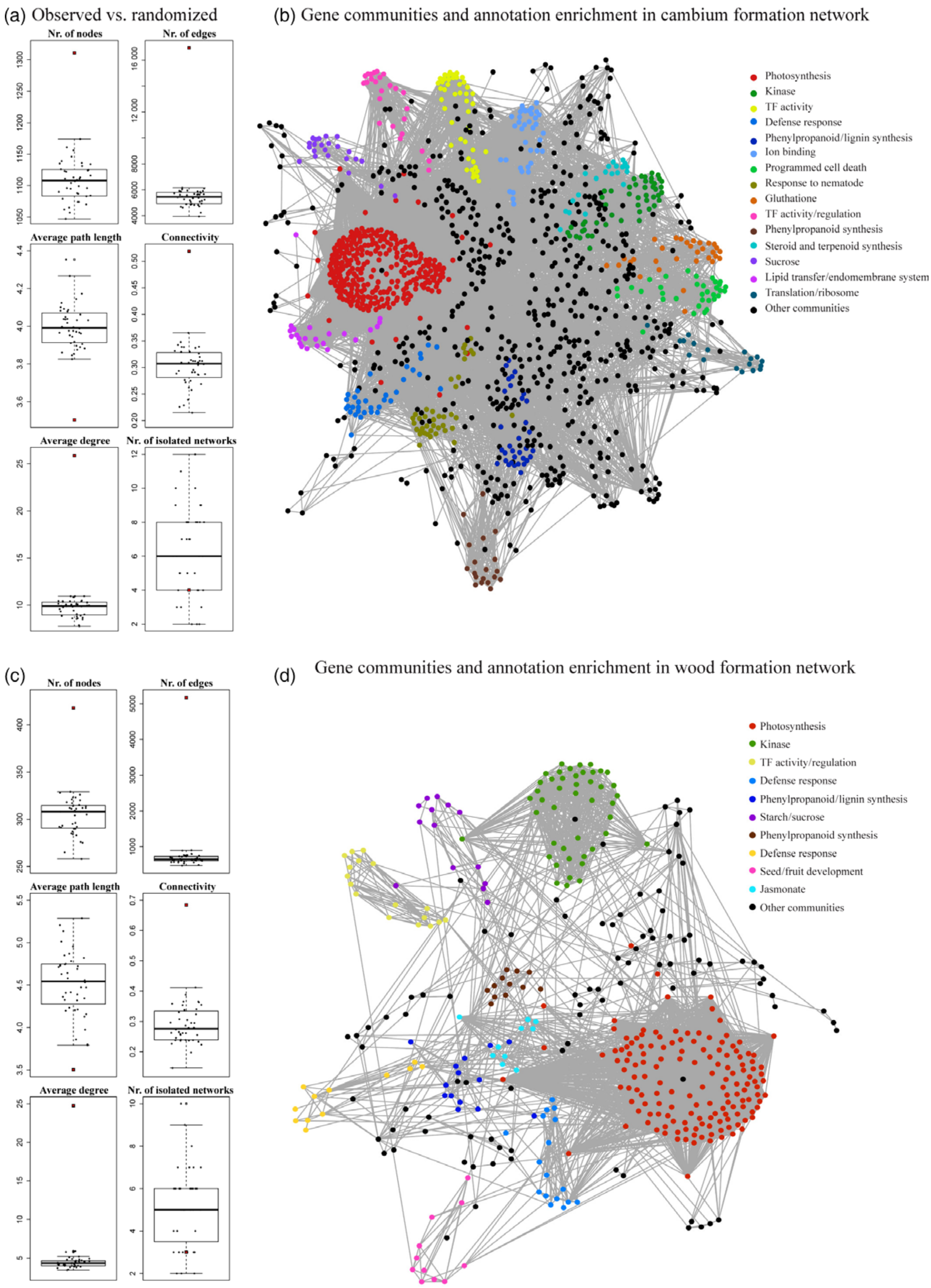


Figure 3. Association networks of genes differentially expressed during cambium (a, b) and wood formation (c, d).
Network parameters were compared with 40 randomized networks; the squares represent the parameters of the observed networks (a, c; *P*-value $<1 \times 10^{-6}$ for each statistic for both networks). The main communities are indicated using colored nodes and the corresponding main enriched annotation theme is indicated in the legend. Directed forced layout was applied to improve readability.

et al., 2013), while *HAT1* interacts with *WUS* in SAM and activates *PLT* that is required for root meristem development (Servet *et al.*, 2010).

Interestingly, most of the genes known from the literature to be involved in various aspects of secondary growth are already expressed in early stages of stem development (Figures S4 and S5, Table S2). For most of these genes, although they are preferentially expressed in stems compared with leaves, changes in expression during secondary growth are not significant and/or do not reach the two-fold threshold compared with primary growth. This emphasizes the genetic resemblance between both stem developmental processes, and may help explain the apparent simple genetic mechanism that angiosperms have utilized to convergently shift from one life form to another.

Secondary growth in *A. thaliana* overlaps with stress-related genes

Several studies of transcriptome remodeling during inflorescence stem formation in Arabidopsis have emphasized the link between mechanical stress and initiation of secondary growth. For instance, Ko *et al.* (2004) observed cambium activity at the base of WT inflorescence stems after applying weights, while Sehr *et al.* (2010) found a significant proportion of touch-inducible genes (with data-derived from Lee *et al.*, 2005) that were upregulated during initiation of vascular cambium. Furthermore, manual clipping of the developing inflorescences resulted in the formation of a complete vascular cambium ring at the base of the WT stem (Oh *et al.*, 2003).

Sehr *et al.* (2010) hypothesized that expression of touch-inducible genes reflects intra-tissue tension resulting from periclinal divisions in the vascular cambium at the onset of secondary growth. While we also found DE of mechanical stress-related genes during cambium formation in our study, the proportion of these genes is higher during wood formation (Table 3, Figure 3). This can be interpreted as a logical consequence of greater intra-tissue tension during more extensive wood formation in *soc1ful*. Likewise, the remarkably high frequency of upregulated genes involved in the JA pathway further highlights stress responses during wood formation. Many key genes involved in JA biosynthesis (*AtLOX2*, *-3*, *-4*, *AOS*, *AOC1*, *-2*, *-3*) as well as JA signaling genes are DE during wood formation in *soc1ful*, including some of the JA master switches (*ORA47*, *AtMYC2*, *JAZ1*, *-2*, *-5*, *-6*, *-7*, *-9*, *-10*, *AtERF-1*, *-5*, *-6*, *-13*, *ERF104*, *ATWRKY70*) (Table 3; Wasternack, 2007; Wasternack and Hause, 2013).

Despite the correlation between stress-related genes and secondary growth, we still do not fully understand the causal relationship. Although the stress responses we observed during the formation of cambium and wood could be the result of increased internal mechanical stress in the stem due to lateral growth, a whole array of abiotic stresses have been suggested to influence the evolutionary habit shift towards secondary woodiness throughout angiosperms. It seems that drought stress during evolutionary time-scales may be involved here, since most of the secondarily woody species occur in continental regions with at least a few consecutive dry months, such as savannas and (semi-)deserts (Lens *et al.*, 2013a). This is in agreement with xylem hydraulic experiments estimating differences in drought stress using stems of Arabidopsis (secondarily woody *soc1ful* and herbaceous WT; Lens *et al.*, 2013b). However, some secondarily woody species grow in very wet climates, suggesting that other abiotic triggers for wood formation that can also vary among groups. For example, the woody Southeast Asian *Begonia* species grow in high altitude, ultramafic soils that probably cause stress to the plants (Kidner *et al.*, 2016), while many of the insular woody species native to oceanic islands are often released from seasonal growth conditions (especially frost; Carlquist, 1974).

Potential candidates for the link between the flowering pathway and cell proliferation

The *soc1ful* genotype combines a late flowering phenotype with enhanced cell proliferation via the vascular cambium, leading to increased secondary growth (see also Sibout *et al.*, 2008). We found three potential candidates, emphasizing the link between the reproductive pathway and cell division activity. The first one is *XAL1*, which both activates the flowering time genes *SOC1*, *FT* and *LFY* and influences cell proliferation in RAM (Tapia-Lopez *et al.*, 2008). In addition to its function in RAM, *XAL1* is upregulated during cambium formation in the WT (Table 2). The second candidate is *AN3* (or *MOTHER OF FT*), which is upregulated during cambium formation in *soc1ful* (Table 2). *AN3* acts as a flowering time gene with *FT*-like properties (Yoo *et al.*, 2004) and it also coordinates cell proliferation in leaves (Kawade *et al.*, 2013). A third gene, *REM1*, is upregulated during wood formation and is normally expressed in SAM, in flower meristems and in lateral shoot meristems, while it also shows a co-expression with the flowering meristem identity gene LFY (Mantegazza *et al.*, 2014).

Arabidopsis thaliana soc1ful is a model for secondary growth

As mentioned before, previous transcriptome studies during stem formation in Arabidopsis inflorescences have identified multiple regulatory and structural genes involved in secondary cell wall formation and cambium formation (Oh *et al.*, 2003; Ko *et al.*, 2004; Ehlting *et al.*, 2005; Sehr *et al.*, 2010). However, the stems investigated did not show extensive wood development, and some of the studies used short-day conditions (Chaffey *et al.*, 2002; Tixier *et al.*, 2013), different types of stress agents (see 'Secondary growth in *A. thaliana* overlaps with stress-related genes') or hormonal treatments (Agusti *et al.*, 2011b) to stimulate cambium activity. We have shown that the combination of stem and leaf transcriptome remodeling using herbaceous WT and woody *soc1-ful* individuals in Arabidopsis represents an excellent approach to investigating the genetic aspects of secondary growth. Firstly, our network analyses show a high connectivity among the DE genes during cambium and wood formation, providing a biologically meaningful structure in both networks (Figure 3). Secondly, comparisons of transcript profiles indicate that metabolic processes in Arabidopsis *soc1ful* resemble those of the WT and non-related *Populus* and *Eucalyptus* trees (Figure S3). Finally, the wood anatomy of *soc1ful* and various related Brassicaceae shrubs and treelets growing in the wild is similar (Lens *et al.*, 2012). Altogether, this suggests that our transcriptome dataset is an excellent starting point for a comparative transcriptomics approach across woody species within the family.

Concluding thoughts

Our study shows that the molecular pathways involved in primary and secondary growth resemble each other closely and should be integrated with each other in plant developmental studies (Jouannet *et al.*, 2015). We also found expression of many mechanical stress-related genes during secondary growth, more particularly during wood formation, which can be related to increasing internal stem stress as a result of a wider wood cylinder pushing against the outer stem tissues. The complex and highly interconnected gene association networks observed during secondary growth may suggest that modification of any number of nodes within this network could potentially result in multiple habit shifts, although more experimental gene interactions and their effect on the global network must be further investigated (Taylor-Teeples *et al.*, 2015). To conclude, *A. thaliana* is a good model to investigate gene expression and interaction networks to better understand the genetic mechanism(s) behind derived woodiness, because: (i) network approaches in Arabidopsis have proven to be promising based on recent studies of the role of gene regulatory networks in secondary cell wall synthesis (Hussey *et al.*, 2013; Taylor-Teeples *et al.*, 2015), (ii) Arabidopsis is a member of the predominantly herbaceous family Brassicaceae, including several hundred species of woody (sub) shrubs, trees and lianas that have evolved multiple times (Al-Shehbaz, 1984), and (iii) the woody *soc1ful* mutant is an example of a secondarily woody shrub that can be compared with herbaceous WT plants in an isogenic background (Lens *et al.*, 2012). Finally, our transcriptome dataset provides baseline data for further comparative transcriptomics studies on secondarily woody Brassicaceae and beyond. Such studies may allow us to further assess the potential roles of genes and interactions among these genes in networks responsible for wood development, one of the most fundamental processes on Earth that has been reinvented multiple times within angiosperms.

EXPERIMENTAL PROCEDURES

Biological materials

We used *A. thaliana* (L.) Heynh. Columbia-0 (Col-0) ecotype (42-day-old individuals) and the woody mutant *soc1-6ful-7* (63-day-old individuals) for our transcriptome sequencing and analysis. For the double mutant, we crossed *soc1-6* (Salk_138131) and *ful-7* (Salk_033647) T-DNA insertion lines (Alonso *et al.*, 2003). We selected homozygous double mutant individuals based on their late flowering and short-silique phenotypes, and confirmed the presence of the T-DNA insertions by PCR. To genotype the plants, we used the primer sequences from the Signal database (http://signal.salk.edu): T-DNA Left Border LBb1.3 (5′-ATTTTGCCG ATTTCGGAAC-3′), soc1-6 forward (5′-AAAGGATGAGGTTTC AAGCG-3′), soc1-6 reverse (5′-ATGTGATTCCAVAAAAGGCC-3′), ful-7 forward (5′-GGAATTTTATGGGGGAAG-3′) and ful-7 reverse (5′-GCGAATTGTTGTGATCT-3′).

Before germination, we incubated the seeds of the double mutants and wild-type plants in Petri dishes with moistened filter paper for 7 days at 4°C. We subsequently transferred 20 seedlings per genotype in individual 1-L pots in a growth chamber at 22°C in long days (16 h:8 h, light:dark). To synchronize the time of flowering, we sowed the *soc1-6 ful7* individuals 21 days before Col-0, and randomly arranged the pots on the same growing table.

Wood developmental stages

To identify stem regions corresponding to developmental stages of interest, we harvested two plants per genotype 4 days before the final sampling. We sectioned main and side branches at different heights (about every 0.5 cm along the stem) using a sliding microtome with freezing unit (MICROM International HM 450, http://www.microm-online.de/). Based on the preliminary anatomical observations, we identified the position of initiation of the vascular cambium in both genotypes and initiation of the wood cylinder in the double mutant (Figure 1). We then harvested 10 comparable individuals per treatment for RNA sequencing (RNA-seq) and detailed anatomical observations. For the 10 double mutant individuals, three developmental stem stages (herbaceous, cambium and wood stage; Figure 1) and two stem stages of the WT (herbaceous and cambium stage; Figure 1) were harvested together with pooled basal rosette leaves. The samples for RNA-seq were immediately transferred to liquid nitrogen and stored at −80°C until RNA extraction. The stem parts just above and below the RNA-seq samples were stored in 70% ethanol for detailed anatomical analysis. We sampled the 20 individuals on three consecutive days from 10:00 h to 12:00 h to avoid diurnal variation, and finally selected three individuals per genotype based on the anatomical analysis of the stem parts.

Stem anatomy

We embedded the stem samples for detailed anatomical observation in epoxy resin (EMbed-812 kit, EMS, https://www.emsdiasum.com/) using the modified protocol from Luft (1961). In short, we dehydrated the tissue in a graded series of ethanol (80, 90, 96%), and twice in 100% ethanol at 4°C for 1 h under vacuum at each step then twice in 100% propylene oxide at room temperature (RT; about degrees) for 20 min. The samples were subsequently infiltrated for 2 h at RT with 1:1 propylene oxide:epoxy resin and overnight at RT with pure epoxy resin. We performed polymerization with fresh epoxy resin in flat embedding molds for 24 h at 60°C. Cross sections 5–7 μm thick were made with a Leica RM2165 rotary microtome (Leica Microsystems, http://www.leica-microsystems.com/); sections were stained in 1% toluidine blue O, mounted in epoxy resin, observed with a Leica DN2500 light microscope, and photographed with a DFC425c camera equipped with LAS software. After careful anatomical observation, three individuals per genotype were selected for RNA isolation. Special emphasis was placed on the initiation and the closure of the vascular cambium ring, which is crucial for selecting the cambium stage, and the initiation of the wood cylinder in the double mutant (Figures 1 and S2).

Total RNA isolation and sequencing

Total RNA was isolated from 21 stem and leaf samples using an EZNA Plant RNA Kit (Omega Bio-tek, http://www.omegabiotek.com/) including DNaseI treatment on the column (Qiagen, http://www.qiagen.com/) according to the manufacturer's instructions. Quality and quantity analysis of the RNA were performed using UV spectroscopy (NanoDrop 1000) and electropherogram (2100 Bioanalyzer, RNA 6000 Nano labChip, Agilent Technologies, http://www.agilent.com/). Poly A+ RNA purification and strand-specific cDNA library preparation for Illumina Hiseq 2000 were performed by the Leiden Genome Technology Center (Leiden, the Netherlands) (Parkhomchuk *et al.*, 2009). For each library, 100 bp were sequenced from both ends (PE100) yielding at least 20 million paired-end reads per sample. The raw Illumina sequences were deposited at the National Center for Biotechnology Information under accession numbers GSM1694242–GSM1694262.

RNA-seq data analyses

FastQ files were converted from Illumina 1.8 format to Sanger format using FASTQ GROOMER v.1.0.4 (Blankenberg *et al.*, 2010) and read quality and quantity was assessed using FASTQC v.0.5.1 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). Adapters and low-quality regions were trimmed using the pair-aware software TRIMMOMATIC v.0.0.3 (Bolger *et al.*, 2014): Illumina TruSeq adapters were filtered out, the first nine bases were removed, a sliding window over four bases with a quality threshold of Phred > 20 was applied corresponding to a base call accuracy of >99% and only reads in pairs and with a minimum length of 45 bp were retained. The high-quality reads were mapped to the *A. thaliana* genome (TAIR10, 06/11/2013) using TOPHAT v.2.0.8 (Kim *et al.*, 2013; using default parameters), the mapping statistics were assessed using FLAGSTAT v.1.0.0 (Blankenberg *et al.*, 2010) and insertion size metrics calculated with SAMTOOLS v.1.56.0 (Li *et al.*, 2009). Relative transcript abundance was calculated using FPKM for biological replicates, and fold changes were statistically evaluated against the null hypothesis using CUFFDIFF v.2.1.1 (using the following parameters: -p 20 -u -N -b). Exploration of the CUFFDIFF output and quality assessments of biological replicates were performed using CUMMERBUND running on R-3.1.0 (Goff *et al.*, 2012; R Development Core Team 2015).

Divergence between transcript abundance was performed using Jenson–Shannon calculations and illustrated in the FPKM dendrogram (Figure S1a). All stem samples from one *soc1ful* individual (m21 in Table S1) clustered together, as well as the cambium and herbaceous stem sample from the WT individual Col1 (Table S1, Figure S1b). We carried out Gene Ontology term enrichment analyses focusing on the genes involved in this aggregation and could not find any obvious reason for the clustering (e.g. plant stress, response to herbivores). Since no biological reason for this clustering was found, we therefore excluded these samples. Consequently, 16 out of 21 samples were included in downstream analyses (Figure S1c).

Gene filtering, annotation and categorization

We selected DE based on a two-fold difference in transcript abundance to reduce the number of false positives (Table S3). To account for multiple testing, changes were considered significant if their *q*-values (*P*-values corrected for multiple tests using the Benjamini–Hochberg false discovery rate) were below 0.05. Genes were annotated using TAIR10 annotations (http://www.arabidopsis.org/, 31/08/2013) and transcription factors identified against the Arabidopsis Plant TFDB v.3.0 (Jin *et al.*, 2014). We categorized genes involved in regulation and synthesis of secondary cell wall genes (lignin, cellulose/xylan) using gene lists (Hussey *et al.*, 2013; Myburg *et al.*, 2014) and genes involved in regulation and synthesis of hormones were matched with the Arabidopsis Hormone Database (Jiang *et al.*, 2011). We also investigate the overlap between this study and published microarray studies that applied stress by removing flower buds or bending the stem or the leaves (Oh *et al.*, 2003; Ko *et al.*, 2004; Lee *et al.*, 2005) (general expression table; Table S4). To improve the readability, standard TAIR gene symbols are used throughout the paper, with the full gene names and AGI codes being presented in Table S3.

Interaction network

To identify gene association networks during cambium or wood formation, the DE gene lists were mapped onto the STRING v.10 global Arabidopsis network (Szklarczyk *et al.*, 2015) in R (R Development Core Team 2015) using STRINGdb (with score_threshold = 400) and visualized with IGRAPH (Csárdi *et al.*, 2008). The STRING v.10 database is based on functional association, i.e. a specific and productive functional relationship between two proteins, probably contributing to a common biological purpose. The database integrates direct and indirect protein/gene interactions, *de novo* predicted interactions based on genomic information and co-expression, pathway knowledge and inferred interactions from other organisms based on pre-computed orthology relations. To evaluate both networks, we compared them with 40 randomized STRING networks using identical numbers of genes as input. The number of connected nodes and edges, average path length, connectivity, average degree (i.e. number of edges per node) and number of isolated networks was calculated for the observed and randomized networks, and statistically assessed using a one sample *t*-test or one-sample Wilcoxon signed rank test in R (Figure 3).

To assess structure within both networks, we performed community detection analyses using an information-based algorithm as suggested previously (Lancichinetti and Fortunato, 2009; Orman *et al.*, 2011) with the function infomap.community in IGRAPH. We visualized the various communities by applying force-directed graph drawing using the Fruchterman–Reingold algorithm (Fruchterman and Reingold, 1991). Gene annotation enrichment analysis for categorization of the largest gene communities within the network was performed using DAVID v.6.7 (Huang *et al.*, 2008).

ACKNOWLEDGEMENTS

We are grateful to Peter Klinkhamer, Karin van der Veen, Bertie Joan van Heuven, Gerda Lamers, Rob Langelaan and Martin van

Rongen for assistance and technical support. This study was financially supported by the Naturalis Biodiversity Center (FES 017/202) and the Alberta Mennega Stichting. Further financial support was provided by the Genome Canada Large-Scale Applied Research Program (POPCAN, project 168BIO), USDA National Institute of Food and Agriculture and AgBioResearch to PPE, a NSERC (Canada) Discovery Grant to CJD and NWO (Netherlands Science Foundation) VIDI and Ecogenomics grants to MES. The authors declare that there are no conflicts of interest. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Dendrograms presenting the relationship between replicates.

Figure S2. Cross sections of *soc1ful* and wild-type stem samples selected for RNA sequencing.

Figure S3. MapMan metabolism overview with a focus on cell wall and secondary metabolism as evidenced by xylem and leaf specific differentially expressed genes in wild-type and *soc1ful Arabidopsis thaliana*, *Populus trichocarpa* and *Eucalyptus grandis*.

Figure S4. Expression of candidate genes clustered by trend.

Figure S5. Expression of secondary cell wall-related genes clustered by trend.

Table S1. Summary of the number of raw reads sequenced and the percentage retained after quality filtering and mapping.

Table S2. Transcript abundance in genes related to meristem activity and xylem development, compiled from recent reviews.

Table S3. Transcript abundance in genes related to secondary cell wall and lignin synthesis.

Table S4. List of gene differentially expressed during cambium formation in the wild-type and *soc1ful* and wood formation.

REFERENCES

Abe, M., Takahashi, T. and Komeda, Y. (2001) Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. *Plant J.* **26**, 487–494.

Adio, A.M., Casteel, C.L., De Vos, M., Kim, J.H., Joshi, V., Li, B., Juery, C., Daron, J., Kliebenstein, D.J. and Jander, G. (2011) Biosynthesis and defensive function of N-acetylornithine, a jasmonate-induced arabidopsis metabolite. *Plant Cell*, **23**, 3303–3318.

Aguilar-Martinez, J.A., Poza-Carrion, C. and Cubas, P. (2007) Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. *Plant Cell*, **19**, 458–472.

Agusti, J., Herold, S., Schwarz, M. *et al.* (2011a) Strigolactone signaling is required for auxin-dependent stimulation of secondary growth in plants. *Proc. Natl Acad. Sci. USA*, **108**, 20242–20247.

Agusti, J., Lichtenberger, R., Schwarz, M., Nehlin, L. and Greb, T. (2011b) Characterization of transcriptome remodeling during cambium formation identifies MOL1 and RUL1 as opposing regulators of secondary growth. *PLoS Genet.* **7**, e1001312.

Aichinger, E., Kornet, N., Friedrich, T. and Laux, T. (2012) Plant stem cell niches. *Annu. Rev. Plant Biol.* **63**, 615–636.

Aloni, R. (1987) Differentiation of vascular tissues. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **38**, 179–204.

Alonso, J.M., Stepanova, A.N., Leisse, T.J. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.

Al-Shehbaz, I.A. (1984) The tribes of Cruciferae (Brassicaceae) in the southeastern United States. *J. Arnold. Arbor.* **65**, 343–373.

Baroja-Fernandez, E., Munoz, F.J., Li, J., Bahaji, A., Almagro, G., Montero, M., Etxeberria, E., Hidalgo, M., Sesma, M.T. and Pozueta-Romero, J. (2012) Sucrose synthase activity in the sus1/sus2/sus3/sus4 Arabidopsis mutant is sufficient to support normal cellulose and starch production. *Proc. Natl Acad. Sci. USA*, **109**, 321–326.

Baucher, M., El Jaziri, M. and Vandeputte, O. (2007) From primary to secondary growth: origin and development of the vascular system. *J. Exp. Bot.* **58**, 3485–3501.

Birchler, J.A. and Veitia, R.A. (2007) The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell*, **19**, 395–402.

Birchler, J.A. and Veitia, R.A. (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl Acad. Sci. USA*, **109**, 14746–14753.

Björklund, S., Antti, H., Uddestrand, I., Moritz, T. and Sundberg, B. (2007) Cross-talk between gibberellin and auxin in development of Populus wood: gibberellin stimulates polar auxin transport and has a common transcriptome with auxin. *Plant J.* **52**, 499–511.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A. and Galaxy, T. (2010) Manipulation of FASTQ data with galaxy. *Bioinformatics*, **26**, 1783–1785.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Carlquist, S.J. (1974) *Island Biology*. New York: Columbia University Press.

Chae, K., Gonong, B.J., Kim, S.C., Kieslich, C.A., Morikis, D., Balasubramanian, S. and Lord, E.M. (2010) A multifaceted study of stigma/style cysteine-rich adhesin (SCA)-like Arabidopsis lipid transfer proteins (LTPs) suggests diversified roles for these LTPs in plant growth and reproduction. *J. Exp. Bot.* **61**, 4277–4290.

Chaffey, N., Cholewa, E., Regan, S. and Sundberg, B. (2002) Secondary xylem development in Arabidopsis: a model for wood formation. *Physiol. Plant.* **114**, 594–600.

Chang, K.N., Zhong, S., Weirauch, M.T. *et al.* (2013) Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. *Elife*, **2**, e00675.

Csárdi, G., Strandburg, K., Zalányi, L., Tobochnik, J. and Érdi, P. (2008) *Estimating the Dynamics of Kernel-Based Evolving Networks*. Berlin: Springer Science + Business Media.

Du, J. and Groover, A. (2010) Transcriptional regulation of secondary growth and wood formation. *J. Integr. Plant Biol.* **52**, 17–27.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. and Lepiniec, L. (2010) MYB transcription factors in Arabidopsis. *Trends Plant Sci.* **15**, 573–581.

Ehlting, J., Mattheus, N., Aeschliman, D.S. *et al.* (2005) Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant J.* **42**, 618–640.

Esau, K. (1977) *Anatomy of Seed Plants*. New York: Wiley.

Eudes, A., Pollet, B., Sibout, R., Do, C.-T., Séguin, A., Lapierre, C. and Jouanin, L. (2006) Evidence for a role of AtCAD 1 in lignification of elongating stems of *Arabidopsis thaliana*. *Planta*, **225**, 23–39.

Feng, C., Andreasson, E., Maslak, A., Mock, H.P., Mattsson, O. and Mundy, J. (2004) Arabidopsis MYB68 in development and responses to environmental cues. *Plant Sci.* **167**, 1099–1107.

Franco-Zorrilla, J.M., Martin, A.C., Solano, R., Rubio, V., Leyva, A. and Paz-Ares, J. (2002) Mutations at CRE1 impair cytokinin-induced repression of phosphate starvation responses in Arabidopsis. *Plant J.* **32**, 353–360.

Fruchterman, T.M.J. and Reingold, E.M. (1991) Graph drawing by force-directed placement. *Software Pract. Exper.* **21**, 1129–1164.

Gao, Q.M., Venugopal, S., Navarre, D. and Kachroo, A. (2011) Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and WRKY51 proteins. *Plant Physiol.* **155**, 464–476.

Goff, L., Trapnell, C. and Kelley, D.R. (2012) CummeRbund: vizualisation and exploration of cufflinks high-throughput sequencing data. R package version 2.0.

Groover, A.T. (2005) What genes make a tree a tree? *Trends Plant Sci.* **10**, 210–214.

Groover, A.T., Mansfield, S.D., DiFazio, S.P., Dupper, G., Fontana, J.R., Millar, R. and Wang, Y. (2006) The populus homeobox gene ARBORKNOX1 reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Mol. Biol.* **61**, 917–932.

Hefer, C.A., Mizrachi, E., Myburg, A.A., Douglas, C.J. and Mansfield, S.D. (2015) Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: *Populus trichocarpa* and *Eucalyptus grandis*. *New Phytol.* **206**, 1391–1405.

Hejatko, J., Ryu, H., Kim, G.T. *et al.* (2009) The histidine kinases CYTOKININ-INDEPENDENT1 and ARABIDOPSIS HISTIDINE KINASE2 and 3 regulate vascular tissue development in Arabidopsis shoots. *Plant Cell*, **21**, 2008–2021.

Hertzberg, M., Aspeborg, H., Schrader, J. *et al.* (2001) A transcriptional roadmap to wood formation. *Proc. Natl Acad. Sci. USA*, **98**, 14732–14737.

Hu, Y., Poh, H.M. and Chua, N.-H. (2006) The Arabidopsis ARGOS-LIKE gene regulates cell expansion during organ growth. *Plant J.* **47**, 1–9.

Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

Hussey, S.G., Mizrachi, E., Creux, N.M. and Myburg, A.A. (2013) Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Front. Plant Sci.* **4**, 325.

Israelsson, M., Eriksson, M.E., Hertzberg, M., Aspeborg, H., Nilsson, P. and Moritz, T. (2003) Changes in gene expression in the wood-formaing tissue of transgenic hybrid aspen with increase secondary growth. *Plant Mol. Biol.* **52**, 893–903.

Israelsson, M., Sundberg, B. and Moritz, T. (2005) Tissue-specific localization of gibberellins and expression of gibberellin-biosynthetic and signaling genes in wood-forming tissues in aspen. *Plant J.* **44**, 494–504.

Jiang, Z., Liu, X., Peng, Z., Wan, Y., Ji, Y., He, W., Wan, W., Luo, J. and Guo, H. (2011) AHD2.0: an update version of Arabidopsis hormone database for plant systematic studies. *Nucleic Acids Res.* **39**, D1123–D1129.

Jin, J., Zhang, H., Kong, L., Gao, G. and Luo, J. (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* **42**, D1182–D1187.

Jouannet, V., Brackmann, K. and Greb, T. (2015) (Pro)cambium formation and proliferation: two sides of the same coin? *Curr. Opin. Plant Biol.* **23**, 54–60.

Kaufmann, K., Pajoro, A. and Angenent, G.C. (2010) Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat. Rev. Genet.* **11**, 830–842.

Kawade, K., Horiguchi, G., Usami, T., Hirai, M. Y. and Tsukaya, H. (2013) ANGUSTIFOLIA3 signaling coordinates proliferation between clonally distinct cells in leaves. *Curr. Biol.* **23**, 788–792.

Kazan, K. and Manners, J.M. (2012) JAZ repressors and the orchestration of phytohormone crosstalk. *Trends Plant Sci.* **17**, 22–31.

Kidner, C., Groover, A., Thomas, D.C., Emelianova, K., Soliz-Gamboa, C. and Lens, F. (2016) First steps in studying the origins of secondary woodiness in Begonia (Begoniaceae): combining anatomy, phylogenetics, and stem transcriptomics. *Biol. J. Linn. Soc.* **117**, 121–138.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36.

Ko, J.H., Han, K.H., Park, S. and Yang, J. (2004) Plant body weight-induced secondary growth in Arabidopsis and its transcription phenotype revealed by whole-transcriptome profiling. *Plant Physiol.* **135**, 1069–1083.

Kong, Y., Zhou, G., Yin, Y., Xu, Y., Pattathil, S. and Hahn, M.G. (2011) Molecular analysis of a family of Arabidopsis genes related to galacturonosyltransferases. *Plant Physiol.* **155**, 1791–1805.

Krichevsky, A., Zaltsman, A., Kozlovsky, S.V., Tian, G.-W. and Citovsky, V. (2009) Regulation of root elongation by histone acetylation in Arabidopsis. *J. Mol. Biol.* **385**, 45–50.

Lancichinetti, A. and Fortunato, S. (2009) Community detection algorithms: a comparative analysis. *Phys. Rev. E Stat. Nonlin. Soft Matter. Phys.* **80**, 056117.

Larson, P.R. (1994) *The Vascular Cambium: Development and Structure*. Berlin/Heidelberg: Springer.

Lee, D., Polisensky, D.H. and Braam, J. (2005) Genome-wide identification of touch- and darkness-regulated Arabidopsis genes: a focus on calmodulin-like and XTH genes. *New Phytol.* **165**, 429–444.

Lens, F., Smets, E. and Melzer, S. (2012) Stem anatomy supports *Arabidopsis thaliana* as a model for insular woodiness. *New Phytol.* **193**, 12–17.

Lens, F., Davin, N., Smets, E. and del Arco, M. (2013a) Insular woodiness on the Canary Islands: a remarkable case of convergent evolution. *Int. J. Plant Sci.* **174**, 992–1013.

Lens, F., Tixier, A., Cochard, H., Sperry, J.S., Jansen, S. and Herbette, S. (2013b) Embolism resistance as a key mechanism to understand adaptive plant strategies. *Curr. Opin. Plant Biol.* **16**, 287–292.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and **1000 Genome Project Data Processing Subgroup** (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Little, C.H.A., MacDonald, J.E. and Olsson, O. (2002) Involvement of indole-3-acetic acid in fascicular and interfascicular cambial growth and interfascicular extraxylary fiber differentiation in *Arabidopsis thaliana* inflorescence stems. *Int. J. Plant Sci.* **163**, 519–529.

Liu, Q. and Wen, C.K. (2012) Arabidopsis ETR1 and ERS1 differentially repress the ethylene response in combination with other ethylene receptor genes. *Plant Physiol.* **158**, 1193–1207.

Love, J., Bjorklund, S., Vahala, J., Hertzberg, M., Kangasjarvi, J. and Sundberg, B. (2009) Ethylene is an endogenous stimulator of cell division in the cambial meristem of Populus. *Proc. Natl Acad. Sci. USA*, **106**, 5984–5989.

Luft, J.H. (1961) Improvement in epoxy resin embedding methods. *J. Cell Biol.* **9**, 409–414.

Mantegazza, O., Gregis, V., Mendes, M.A., Morandini, P., Alves-Ferreira, M., Patreze, C.M., Nardeli, S.M., Kater, M.M. and Colombo, L. (2014) Analysis of the arabidopsis REM gene family predicts functions during flower development. *Ann. Bot.* **114**, 1507–1515.

Matsumoto-Kitano, M., Kusumoto, T., Tarkowski, P., Kinoshita-Tsujimura, K., Vaclavikova, K., Miyawaki, K. and Kakimoto, T. (2008) Cytokinins are central regulators of cambial activity. *Proc. Natl Acad. Sci. USA*, **105**, 20027–20031.

Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A. and Beeckman, T. (2008) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nat. Genet.* **40**, 1489–1492.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A. *et al.* (2014) The genome of Eucalyptus grandis. *Nature*, **510**, 356–362.

Nilsson, J., Karlberg, A., Antti, H., Lopez-Vernaza, M., Mellerowicz, E., Perrot-Rechenmann, C., Sandberg, G. and Bhalerao, R.P. (2008) Dissecting the molecular basis of the regulation of wood formation by auxin in hybrid aspen. *Plant Cell*, **20**, 843–855.

Oh, S., Park, S. and Kan, K. (2003) Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. *J. Exp. Bot.* **54**, 2709–2722.

Orman, G.K., Labatut, V. and Cherifi, H. (2011) On accuracy of community structure discovery algorithms. *J. Convergence Inf. Technol.* **6**, 283–292.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H. and Soldatov, A. (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37**, e123–e123.

Pre, M., Atallah, M., Champion, A., De Vos, M., Pieterse, C.M.J. and Memelink, J. (2008) The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiol.* **147**, 1347–1357.

R Development Core Team (2015) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ragni, L., Nieminen, K., Pacheco-Villalobos, D., Sibout, R., Schwechheimer, C. and Hardtke, C.S. (2011) Mobile gibberellin directly stimulates arabidopsis hypocotyl xylem expansion. *Plant Cell*, **23**, 1322–1336.

Rashid, A. and Deyholos, M.K. (2011) PELPK1 (At5g09530) contains a unique pentapeptide repeat and is a positive regulator of germination in *Arabidopsis thaliana*. *Plant Cell Rep.* **30**, 1735–1745.

Rowe, N. and Paul-Victor, C. (2012) Herbs and secondary woodiness – keeping up the cambial habit. *New Phytol.* **193**, 3–5.

Sanchez, P., Nehlin, L. and Greb, T. (2012) From thin to thick: major transitions during stem development. *Trends Plant Sci.* **17**, 113–121.

Sankar, M., Nieminen, K., Ragni, L., Xenarios, I. and Hardtke, C.S. (2014) Automated quantitative histology reveals vascular morphodynamics during Arabidopsis hypocotyl secondary growth. *Elife*, **3**, e01567.

Schrader, J., Nilsson, J., Mellerowicz, E., Berglund, A., Nilsson, P., Hertzberg, M. and Sandberg, G. (2004) 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.

Schröder, F., Lisso, J., Lange, P. and Müssig, C. (2009) The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. *BMC Plant Biol.* **9**, 20.

Schuetz, M., Smith, R. and Ellis, B. (2013) Xylem tissue specification, patterning, and differentiation mechanisms. *J. Exp. Bot.* **64**, 11–31.

Sehr, E.M., Agusti, J., Lehner, R., Farmer, E.E., Schwarz, M. and Greb, T. (2010) Analysis of secondary growth in the Arabidopsis shoot reveals a positive role of jasmonate signalling in cambium formation. *Plant J.* **63**, 811–822.

Servet, C., Conde e Silva, N. and Zhou, D.-X. (2010) Histone acetyltransferase AtGCN5/HAG1 is a versatile regulator of developmental and inducible gene expression in arabidopsis. *Mol. Plant*, **3**, 670–677.

Sibout, R., Plantegenet, S. and Hardtke, C.S. (2008) Flowering as a condition for xylem expansion in Arabidopsis hypocotyl and root. *Curr. Biol.* **18**, 458–463.

Spicer, R. and Groover, A. (2010) Evolution of development of vascular cambia and secondary growth. *New Phytol.* **186**, 577–592.

Strabala, T.J., O'Donnell P, J., Smit, A.M., Ampomah-Dwamena, C., Martin, E.J., Netzler, N., Nieuwenhuizen, N.J., Quinn, B.D., Foote, H.C. and Hudson, K.R. (2006) Gain-of-function phenotypes of many CLAVATA3/ESR genes, including four new family members, correlate with tandem variations in the conserved CLAVATA3/ESR domain. *Plant Physiol.* **140**, 1331–1344.

Szklarczyk, D., Franceschini, A., Wyder, S. *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452.

Tapia-Lopez, R., Garcia-Ponce, B., Dubrovsky, J.G., Garay-Arroyo, A., Perez-Ruiz, R.V., Kim, S.H., Acevedo, F., Pelaz, S. and Alvarez-Buylla, E.R. (2008) An AGAMOUS-related MADS-box gene, XAL1 (AGL12), regulates root meristem cell proliferation and flowering transition in Arabidopsis. *Plant Physiol.* **146**, 1182–1192.

Taylor-Teeples, M., Lin, L., de Lucas, M. *et al.* (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature*, **517**, 571–575.

Tixier, A., Cochard, H., Badel, E., Dusotoit-Coucaud, A., Jansen, S. and Herbette, S. (2013) *Arabidopsis thaliana* as a model species for xylem hydraulics: does size matter? *J. Exp. Bot.* **64**, 2295–2305.

Turchi, L., Carabelli, M., Ruzza, V. *et al.* (2013) Arabidopsis HD-Zip II transcription factors control apical embryo development and meristem function. *Development*, **140**, 2118–2129.

Uggla, C., Moritz, T., Sandberg, G. and Sundberg, B. (1996) Auxin as a positional signal in pattern formation in plants. *Proc. Natl Acad. Sci. USA*, **93**, 9282–9286.

Uggla, C., Mellerowicz, E.J. and Sundberg, B. (1998) Indole-3-acetic acid controls cambial growth in scots pine by positional signaling. *Plant Physiol.* **117**, 113–121.

Wasternack, C. (2007) Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann. Bot.* **100**, 681–697.

Wasternack, C. and Hause, B. (2013) Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Ann. Bot.* **111**, 1021–1058.

Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H. and Schmulling, T. (2003) Cytokinin-deficient transgenic arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell*, **15**, 2532–2550.

Yaginuma, H., Hirakawa, Y., Kondo, Y., Ohashi-Ito, K. and Fukuda, H. (2011) A novel function of TDIF-related peptides: promotion of axillary bud formation. *Plant Cell Physiol.* **52**, 1354–1364.

Yoo, S.Y., Kardailsky, I., Lee, J.S., Weigel, D. and Ahn, J.H. (2004) Acceleration of flowering by overexpression of MFT (MOTHER OF FT AND TFL1). *Mol. Cells*, **17**, 95–101.

Yoshida, H. and Ueguchi-Tanaka, M. (2014) DELLA and SCL3 balance gibberellin feedback regulation by utilizing INDETERMINATE DOMAIN proteins as transcriptional scaffolds. *Plant Signal. Behav.* **9**, e29726.

Zander, M., La Camera, S., Lamotte, O., Métraux, J.-P. and Gatz, C. (2010) *Arabidopsis thaliana* class-II TGA transcription factors are essential activators of jasmonic acid/ethylene-induced defense responses. *Plant J.* **61**, 200–210.

Zhang, J., Nieminen, K., Serra, J.A.A. and Helariutta, Y. (2014) The formation of wood and its control. *Curr. Opin. Plant Biol.* **17**, 56–63.

Zhou, J., Lee, C., Zhong, R. and Ye, Z.H. (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. *Plant Cell*, **21**, 248–266.