



Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in *Populus*

Hui Li^{1†} , Xinren Dai^{1†} , Xiong Huang^{1,2} , Mengxuan Xu¹ , Qiao Wang^{1,2} , Xiaojing Yan¹ , Ronald R. Sederoff³  and Quanzi Li^{1*} 

1. State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China

2. Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China

3. Forest Biotechnology Group, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA

†These authors contributed equally to this work.

*Correspondence: Quanzi Li (liqz@caf.ac.cn)



Hui Li



Quanzi Li

ABSTRACT

High-throughput single-cell RNA sequencing (scRNA-seq) has advantages over traditional RNA-seq to explore spatiotemporal information on gene dynamic expressions in heterogenous tissues. We performed Drop-seq, a method for the dropwise sequestration of single cells for sequencing, on protoplasts from the differentiating xylem of *Populus alba* × *Populus glandulosa*. The scRNA-seq profiled

9,798 cells, which were grouped into 12 clusters. Through characterization of differentially expressed genes in each cluster and RNA *in situ* hybridizations, we identified vessel cells, fiber cells, ray parenchyma cells and xylem precursor cells. Diffusion pseudo-time analyses revealed the differentiating trajectory of vessels, fiber cells and ray parenchyma cells and indicated a different differentiation process between vessels and fiber cells, and a similar differentiation process between fiber cells and ray parenchyma cells. We identified marker genes for each cell type (cluster) and key candidate regulators during developmental stages of xylem cell differentiation. Our study generates a high-resolution expression atlas of wood formation at the single cell level and provides valuable information on wood formation.

Keywords: differentiating trajectory, differentiating xylem, marker genes, *Populus alba* × *Populus glandulosa*, single-cell RNA-seq, wood formation

Li, H., Dai, X., Huang, X., Xu, M., Wang, Q., Yan, X., Sederoff, R.R., and Li, Q. (2021). Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in *Populus*. *J. Integr. Plant Biol.* **63**: 1906–1921.

INTRODUCTION

Wood is the major renewable resource for timber, pulp, and many other industrial wood and paper products (Plomion et al., 2001). Many bioenergy traits, such as biomass yield, physical and chemical composition of wood, etc. affect wood utilization efficiency. With the increased conservation of natural forests, planted forests have attracted more attentions and investments. Genetic tree improvement, traditionally focused on growth and yield, has expanded to include specialized traits for specific uses, through clonal forestry or future genetic engineering. Trees with specific ideal wood properties are an urgent need (Guerriero et al., 2014; Li et al., 2014).

Wood is formed by the perennial activity of the vascular cambium, a secondary meristem, which generates xylem and phloem mother cells at inward and outward sides, respectively (Evert, 2006). The cells in the cambium undergo longitudinal division to create more fusiform initials, and also highly asymmetric periclinal cell division to produce the nearly isodiametric ray initials and long fibrous fusiform initials. Fusiform initials are spindle shaped, many times longer than their width, and give rise to all of the xylem cells that are longitudinal and parallel to the long axis of the tree, which in angiosperm includes tracheary vessel elements, fibers and axial parenchyma. Ray initials give rise to the radial system of ray cells, which is quite elaborate in some species. A cambial initial divides to produce another initial and a mother cell,

which expands and undergoes a second round of division and expansion to create a canonical block of four cells, known as Sanio's four. The four cells include a continuing initial, a xylem mother cell, and two daughter cells that divide one more before expanding and differentiating as xylem cells. Sanio's four can be identified in the orderly differentiating xylem of conifers, but it is more difficult to identify the groups of four in angiosperms where xylogenesis is less orderly. Xylem mother cells undergo cell expansion, secondary cell wall (SCW) thickening and programmed cell death (PCD), as the final steps of differentiation into secondary xylem (wood). Evert (2006) argues that identification of early cell types is made difficult by less orderly division and “the nearly total ignorance of molecular events”. Single cell sequencing provides new technology for such investigations.

The xylem in angiosperms consists of vessel elements for vertical water transport, fiber cells for mechanical support and parenchyma cells for radial transport and storage of nutrients (ray cells) and also through longitudinal parenchyma. Vessel elements fuse to create long vertical channels. Vessels and fiber cell have thick lignified SCWs. The major components of the SCWs are cellulose, lignin and hemicelluloses, which are chemically bonded to form a lignin-carbohydrate complex, structurally analogous to a reinforced multilayered composite. The deposition of cellulose, hemicelluloses and lignin confers mechanical strength to fiber cells for normal vertical growth, and the lignin provides a hydrophobic surface of the SCW, facilitating water transport in vessels.

The deposition of cellulose, hemicelluloses and lignin is highly ordered, which needs the precise transcriptional control of three independent biosynthetic pathway genes (Mellerowicz and Sundberg, 2008). Hundreds of transcription factors (TFs) and microRNAs controlling xylogenesis have been identified in the model plants (Zhong and Ye, 2007). A comprehensive building a hierarchical genetic regulatory network (hGRN) is needed to understand how the cambial meristem responds to endogenous developmental signals and to exogenous environmental stimuli to initiate xylem differentiation and SCW formation. The number and nature of regulatory elements and their interactions within the xylo-genic hGRN remain unknown.

Currently, the methods used for the deciphering the xylo-genic hGRNs include transcriptome sequencing, Graphical Gaussian Modeling (GGM)-based algorithms, yeast one hybridization (Y1H), and chromatin immunoprecipitation (ChIP-seq). High throughput Y1H assays were used to identify protein–DNA interactions in both *Arabidopsis* and poplar (Taylor-Teeple et al., 2015; Petzold et al., 2018; Yeh et al., 2019). In poplar, the hGRN directed by a SCW-associated NAC-domain protein PtrSND1-B1 was built by integrating RNA-seq, top-down GGM-based algorithms and ChIP-seq (Lin et al., 2013; Chen et al., 2019). Many protein–protein interactions among xylem-active TFs during wood formation were identified by yeast two hybridization (Y2H) (Petzold et al., 2018).

However, the large scale of yeast assays on TFs, which are derived from multiple xylem cells, may generate some

false positives. ChIP-seq using xylem protoplasts, also generates some false positives due to the 35S promoter-driven gene overexpression. Information is limited regarding expression of candidate genes in the same cell because a protein–protein interaction identified by Y2H may involve candidates expressed in different cells, giving rise to a false Y2H positive. RNA *in situ* hybridization (RISH) for gene expression analysis is tedious and depends on the probe specificity. Laser capture dissection to collect xylem cells for RNA sequencing is also tedious, and the sequencing is not always from a single cell (Lin et al., 2017; Shi et al., 2017). Although a series of cryosections were isolated from stems and their transcriptome provides potential gene co-expression information (Jokipii-Lukkari et al., 2017; Sundell et al., 2017), whether the co-expression is actually within a single cell is not yet known.

Recent advances in single-cell RNA sequencing (scRNA-seq), including robotics, microfluidics and hydrogel droplets (Zhang et al., 2019), resolve the cell specificity problem. The “drop-seq” method can analyze gene expression in thousands of individual cells from a heterogeneous tissue (Zheng et al., 2017). In this method, each cell is encapsulated in one oil droplet, which is barcoded to distinguish transcripts from different cells. Each gene is provided with a unique molecular identifier (UMI), allowing the counting of individual transcripts on a single cell basis (Kivioja et al., 2011). This technique has been widely used in human cancer cells, nematode worms, and viruses and has yielded complete cell type-specific information (Jean-Baptiste et al., 2019). In plants a number of scRNA sequencing efforts have been conducted in different tissues of model plants, including roots, lateral roots, female gametophytes, shoot apex, leaf phloem and stomas of *Arabidopsis* (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Turco et al., 2019; Zhang et al., 2019; Liu et al., 2020; Song et al., 2020; Kim et al., 2021; Serrano-Ron et al., 2021; Zhang et al., 2021b), rice roots (Liu et al., 2021; Zhang et al., 2021a), the ears, shoot apex and early meiotic cells of maize (Nelms and Walbot, 2019; Satterlee et al., 2020; Xu et al., 2021). A study of scRNA-seq for wood formation is lacking.

Here, we report the gene expression profiling of 9,798 wood-forming cells at a single-cell level in a hybrid poplar. This compendium includes some common cell types reported in previous studies and identifies highly specific marker genes for each population of profiled cells. We used pseudotime analysis to characterize gene expression changes during xylogenesis, showing cell specific differentiation at high resolution.

RESULTS

scRNA-seq of differentiating xylem in *P. alba* × *P. glandulosa*

We performed scRNA-seq using the 10x Genomics Chromium platform (Zheng et al., 2017) to study wood formation in poplar. Protoplasts were prepared for the differentiating

xylem from a stem of *P. alba* × *P. glandulosa*. Stem sectioning showed that about 10 layers of differentiating xylem cells were released by enzyme digestion (Figure 1). In total, 10,646 cells were captured for library construction and paired-end sequencing. After filtration removal of the cells with gene numbers less than 500 and over 10,000 (Figure S1A) and the cells with more than 30,000 UMIs (Figure S1B), a total of 9,798 high quality cells were retained for further analysis. In total, 400,645,657 sequencing reads were obtained, with 97.0% valid barcodes. 83.3% of the sequencing reads could be mapped to the *P. alba* × *P. glandulosa* genome (Qiu et al., 2019; Table S1). We identified 50,762 genes in these 9,798 cells, with 2,190 median genes per cell, and the UNI number per cell was 3,673 (Figure S2; Table S1). These values were well within the range reported in previous droplet-based scRNA-seq. To further evaluate to what extent the gene expression profile in scRNA-seq represents the bulk differentiating xylem tissue, we compared the expressed genes between scRNA-seq and bulk differentiating xylem RNA-seq. 85.84% of the differentiating xylem-expressed genes were detected in scRNA-seq (Figure S1C). Pearson correlation analysis was conducted to compare the gene expression between scRNA-seq and three biological replicates of bulk differentiating xylem RNA-seq, and a high correlation coefficient of over 0.8 and a high goodness of fit (R^2 value = 0.82, P value = 2.2e-16) were obtained in the three comparisons (Figure S1D). Therefore the scRNA-seq data appear valid and reflect the transcriptome of differentiating xylem of *P. alba* × *P. glandulosa*.

Unsupervised analysis identified 12 clusters with high cell heterogeneity

We conducted unsupervised clustering analysis of the 9,798 cells with the canonical correlation analysis function of Seurat (Satija et al., 2015), which yielded 12 clusters (Figure 2A, B). The cell numbers distributed in each cluster ranged from 189 to 2,023 (Figure S2A). To investigate whether the cells within each cluster have high homogeneity, we conducted weighted gene co-expression network analysis (WGCNA). Through 15 iterative computations, all genes in different clusters were divided into 82 modules (Figures S3A, B). After merging, 11 modules were obtained (Figure S3B–D). Except for DarkOliveGreen module that had higher correlations with two clusters (2 and 6), the other 10 modules were highly correlated with one cluster each among 10 clusters (Figure S3C, D). The genes in each module had higher expression levels in the corresponding cluster than in other clusters (Figure S4). WGCNA showed that the genes were co-expressed within each cluster (Figure S5). The cells in each cluster have a high homogeneity, and each cluster has its own characteristic cell type, except that Clusters 2 and 6 share common characters with similar gene expression patterns.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses of 12 clusters

To gain basic information of cell types in the 12 clusters, we identified upregulated differentially expressed genes (DEGs)

in each cluster (using a fold change FC > 2 and a P value < 0.05, compared with the other 11 clusters) and examined their potential functions and pathways according to Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses. In GO analysis, we observed that four clusters (2, 6, 7, and 10) were significantly enriched in xylan biosynthesis (GO: 0045491), hemicellulose biosynthesis (GO: 0010410), cell wall polysaccharide biosynthesis (GO: 0010383), and cell wall macromolecule biosynthesis (GO: 0044036) (Figure 2C; Dataset S1), all of which are involved in cell wall formation, which is essentially a wood formation process. In KEGG, three clusters (2, 6, and 7) were significantly enriched in phenylalanine metabolism (ko00360) and pentose phosphate pathway (ko00030) (Figure 2D). Clusters 2 and 7 were significantly enriched in phenylpropanoid biosynthesis (ko00940) (Dataset S2), and these enriched genes encode all necessary enzymes for lignin biosynthesis (Figure S6), indicating lignin biosynthesis in Clusters 2 and 7. Clusters 2 and 7 cells are in the stage of SCW thickening of either vessels or fiber cells.

Four clusters (0, 4, 9, and 10) were significantly enriched in the formation of ribonucleoprotein complexes (GO: 1990904) in GO analysis (Figure 2C) and ribosome biogenesis (ko03008) in KEGG analysis (Figure 2D). Clusters 3, 5, 8, and 11 were enriched in protein processing in the endoplasmic reticulum (ko04141), and in the MAPK signaling pathway

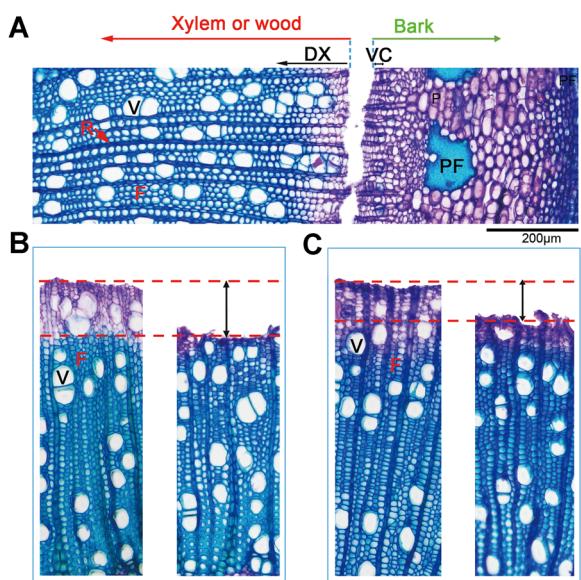


Figure 1. Differentiating xylem (DX) protoplast preparation by the stem dipping approach

(A) A cross-section of the stem with the bark is peeled, showing that cambium is attached to the bark side. (B) Cross-sections of stems showing that about 15 layers of DX cells were scraped out for the RNA-seq. Left: before scraping; Right: after scraping. (C) Cross sections of stems showing that about 10 layers of DX cells were released for the single-cell RNA sequencing (scRNA-seq). Left: before digestion; right: after digestion. F: fiber; PF: phloem fiber; R: ray cell; V: vessel; VC: vascular cambium.

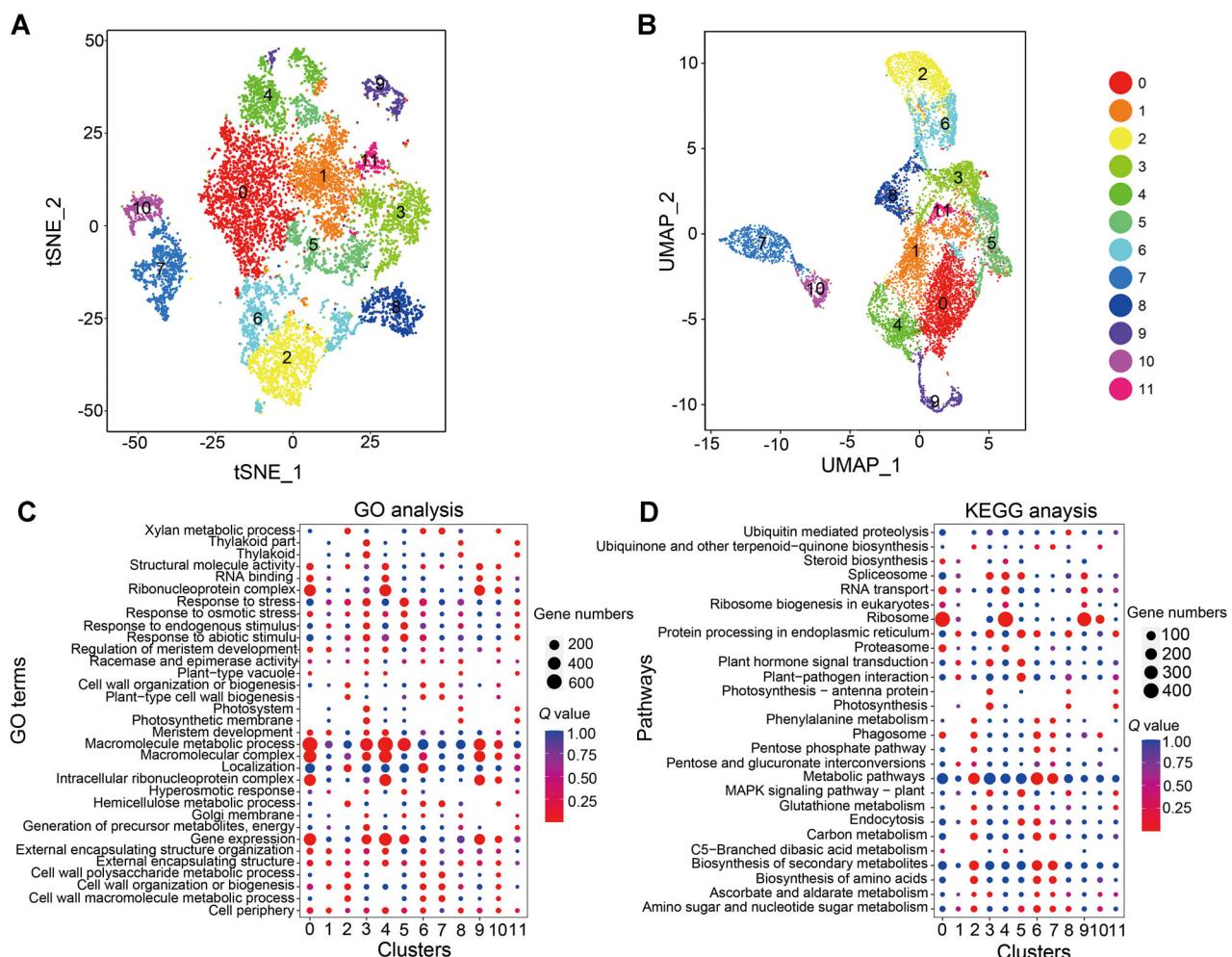


Figure 2. Cluster analysis of single-cell transcriptome from the differentiating xylem in *Populus alba* × *Populus glandulosa*

12 clusters were obtained by the unsupervised clustering using Seurat and visualized using t-SNE (A) and UMAP (B) algorithms. Clusters are shown in different colors, and each dot represents a single cell. 12 clusters were analyzed by Gene Ontology (GO) (C) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (D). Top 5 GO terms and pathways are shown in each cluster. The circle size represents enriched gene numbers. Red color and blue color represent significant and insignificant enrichments, respectively, and the degree of significant enrichments on GO terms and KEGG pathways is shown by Q value.

(ko04016) in KEGG analysis (Figure 2D). Clusters 3 and 5 were mainly enriched in plant hormone signal transduction (ko04075) (Figure 2D). Based on the KEGG annotation, these 12 clusters belong to three major groups. Group A (Clusters 2, 6, and 7) contains the cells with SCW thickening. Group B (Clusters 0, 3, 4, 5, 8, 9, 10, and 11) contains the cells without cell wall thickening that should be either early differentiating cells before SCW thickening or parenchyma cells in the xylem. Group C (Cluster 1) has few enrichments on both GO terms and KEGG pathway compared to other clusters (Datasets S1, S2), suggesting that it belongs to the earliest cell stage among the 12 clusters.

Cell wall metabolic genes in Clusters 2, 6, and 7

To gain further understanding of the SCW formation in Clusters 2, 6, and 7, we analyzed the DEGs in these three clusters and focused on pathway genes in the biosynthesis of cellulose, lignin, and hemicelluloses. DEGs were obtained by

comparing the gene expression between Clusters 2, 6, 7, and the other nine clusters (as a group), with FC > 2 and P value < 0.05. The DEG (upregulated) numbers in Clusters 2, 6, 7 were 667, 961, and 536, respectively. 301 DEGs were upregulated in all three clusters (Dataset S3).

In *P. alba* × *P. glandulosa*, *PagCesA4*, 7-A/B, and 8-A/B, which encode cellulose synthases for SCWs (Joshi et al., 2011; Xi et al., 2017; Abbas et al., 2020), were abundantly expressed in differentiating xylem compared to phloem and leaves (Dataset S4; Figure S7). Except *PagCesA7-B*, the other four *CesAs*, *PagCesA4*, *PagCesA7-A*, *PagCesA8-A*, and *PagCesA8-B* were up-regulated significantly in Clusters 2, 6, and 7 (Dataset S3), showing SCW cellulose synthesis in these three clusters.

Monolignol biosynthesis involves 11 enzyme families (Vanholme et al., 2019). Using the RNA-seq data, we identified the specific monolignol biosynthetic genes (Dataset S4; Figure S7) and examined their expression in the 12

clusters. DEGs analysis showed that 13 genes (*PagC3H3*, *PagCAD1*, *PagC4H1*, *PagC4H2*, *PagCCoAOMT1*, *PagCCoAOMT2*, *Pag4CL3*, *PagCOMT2*, *PagHCT1*, *PagCSE1*, *PagCSE2*, *PagPAL2*, and *PagPAL4/5*) belonging to nine families were all upregulated significantly in Clusters 2, 6, and 7 (Dataset S3). Two CAld5H family members, *PagCAld5H1* and *PagCAld5H2* had increased transcripts in Clusters 2, 6, and 7 compared to other clusters, and the increase was significant for *PagCAld5H1* in Clusters 7 (2^6 fold) and for *PagCAld5H2* in Cluster 2 (2^7 fold). The DEG analysis showed the enrichment of genes for monolignol biosynthesis, with enrichment of phenylpropanoid biosynthesis in the KEGG analysis.

We looked for the expression of glycosyltransferase genes involved in xylan biosynthesis (Rennie and Scheller, 2014) and observed that 14 such genes were significantly upregulated in Clusters 2, 6, and 7 (Dataset S3), including *PagGT43A*, *PagGT43B*, *PagGT43C*, *PagiRX10-1*, *PagiRX10-2*, *PagiRX10-L-A1* for xylan backbone biosynthesis, *PagPARVUS-2*, *PagGT8D2*, and *PagFRA8* for reducing end biosynthesis, *PagGUX1a* and *PagGUX2* for adding GlcA, *PagGXM1*, *PagGXM2*, and *PagiRX15-L* for adding 4-O-methyl groups to GlcA residues of xylan.

PagCsIA1, encoding glucomannan synthase for the synthesis of (1,4) β -D-glucosidase (Suzuki et al., 2006), was upregulated in Clusters 2, 6, and 7 (Dataset S3), indicating the deposition of mannan in the cell walls of these three cluster cells. All these DEGs analyses showed that essential enzyme-encoding genes for cellulose, hemicellulose and lignin biosynthesis were upregulated in Clusters 2, 6, and 7, not in other nine clusters, indicating that the cells in these three clusters are undergoing SCW thickening.

Cell types of 12 clusters

To distinguish vessel and fiber cell types in Clusters 2, 6, and 7, we checked the vessel marker genes among the DEGs that were significantly and specifically upregulated in Clusters 2, 6, and 7 (Dataset S3). Different from fiber cells, vessel elements are a special type of tracheary elements (TEs) that die after SCW thickening. TEs undergo PCD with a series of proteolytic and hydrolytic processes. The tonoplast first undergoes a change in transport capability, and then its implosion initiates final TE clearing or mega-autolysis (Kuriyama, 1999). Cysteine proteases and other hydrolases are loaded into the central vacuole and then released by its implosion to carry out TE mega-autolysis (Kuriyama and Fukuda, 2002). Several proteolytic enzymes are involved in this process (Beers et al., 2000; Kuriyama and Fukuda, 2002; Avci et al., 2008; Bolhoner et al., 2013; Rennie and Scheller, 2014). In *Arabidopsis*, genes encoding XYLEM CYSTEINE PROTEASEs (XCP1 and XCP2) that are involved in vacuole autolysis and later tonoplast implosion, are specifically expressed in vessel elements (Funk et al., 2002; Avci et al., 2008; Yamaguchi et al., 2011). Our results showed that *PagXCP1* and *PagXCP2* were significantly and specifically upregulated in Cluster 7 (Dataset S3). Type II metacaspase

degrades vessel cell content (Ito and Fukuda, 2002; Bolhoner et al., 2013) and endonuclease ZEN1 in zinnia mediates DNA degradation after vacuole rupture during PCD (Ito and Fukuda, 2002). Significant increases of gene expression in Cluster 7 were observed for a metacaspase gene *PagMC9* (FC = $>2^7.86$), a S-like nuclease family protein gene *PagEN1* (FC > $2^6.06$), and the cysteine proteinase gene *PagCP* (FC > $2^3.81$) (Dataset S3). Other genes, encoding serine carboxypeptidase *PagSCP*, aspartyl protease *PagASP1* and the vascular processing enzyme *PagAEP3*, were also upregulated in Cluster 7 (Dataset S3). The DEG analysis showed that PCD-related proteolytic enzyme encoding genes were specifically and significantly expressed in Cluster 7, indicating that cells in Cluster 7 undergo PCD and belong to vessel elements.

Among the three clusters (2, 6, and 7), Clusters 2 and 6 shared more common DEGs (Dataset S3). Comparison of the DEGs between Clusters 2 and 6 showed that most of the DEGs in one cluster were also upregulated in the other cluster compared to other nine clusters, although the upregulation was not significant, indicating these two clusters may belong to the same cell type, just at different developmental stages. *CAld5H2* was identified as fiber-specific in *P. trichocarpa* by laser microdissection (LMD) (Shi et al., 2017). This gene *PagCAld5H2* was significantly upregulated in Cluster 2 (Datasets S3, S5), showing its involvement in the S monolignol biosynthesis in fiber cells. Two fiber-specific genes *Potri.001G099800* (*MYB010*) and *Potri.008G136600* (*CCoAOMT3*) (Shi et al., 2017), were significantly upregulated in Clusters 2 and 6. The fiber marker *MYB148* gene (*Potri.012G084100*) (Shi et al., 2017) was significantly upregulated in Cluster 6 (Datasets S3, S5). Although *PagCsIA1* was upregulated in Clusters 2, 6, and 7, it had the highest expression level in Cluster 2 (twice as in Cluster 7) (Dataset S6), consistent with the result that mannan is richer in wood fibers than in vessels (Kim and Daniel, 2012). These results suggest that Clusters 2 and 6 belong to fiber cells. The expression of β -glucuronidase *GUS* gene driven by a *Eucalyptus CAD* gene is detected in the xylem rays, suggesting that parenchyma cells expressing *CAD* may provide lignin precursor to the adjacent lignified elements (Feuillet et al., 1995). In our study, *PagCAD7* (*Pop_G09G031724*) was significantly upregulated in Cluster 5 (Datasets S3, S6).

There are limited numbers of markers available to characterize the remaining nine clusters which do not have SCW. To aid cell type identification, we selected one gene from each cluster that is specifically upregulated in its cluster for RISH. Consistent with the above characterizations of Clusters 2, 6, and 7, *Pop_A07G022645* (Cluster 2) and *Pop_A13G054260* (Cluster 6) were expressed in fiber cells by RISH (Figure 3B), supporting the view that these two clusters comprise fiber cells. *Pop_A09G026733* (*PagLAC22*) is the ortholog of *Arabidopsis AtLAC11*, which is co-expressed with *AtXCP1* (direct target of vessel-specific master regulator *AtVND7*) (Yamaguchi et al., 2011). *Pop_A09G026733* (Cluster

7) was expressed in vessels by RISH (Figure 3B), supporting the vessel identity of Cluster 7.

Based on the scRNA-seq data, Pop_A10G001185 (for Cluster 1) showed nearly even expression among all 12 selected genes (Figure 3A; Dataset S6). Consistently RISH detected its expression level in all cell types, including vessels, ray parenchyma cells and fiber cells in both xylem and phloem (Figure 3B), indicating that Cluster 1 has a low heterogeneity. Pop_A01G004551 (for Cluster 8) and

Pop_A06G063029 (for Cluster 11) were expressed in both fiber cells and ray parenchyma cells. Pop_A19G053004 (for Cluster 3) and Pop_G01G010746 (for Cluster 5) displayed a high RNA abundance in ray parenchyma cells and no signals in phloem fibers, indicating that Clusters 3 and 5 belong to ray parenchyma cells at different developmental stages. Obvious RNA signals were detected for Pop_A14G000834 (for Cluster 10) in vessels, suggesting Cluster 10 was in the process of differentiating into

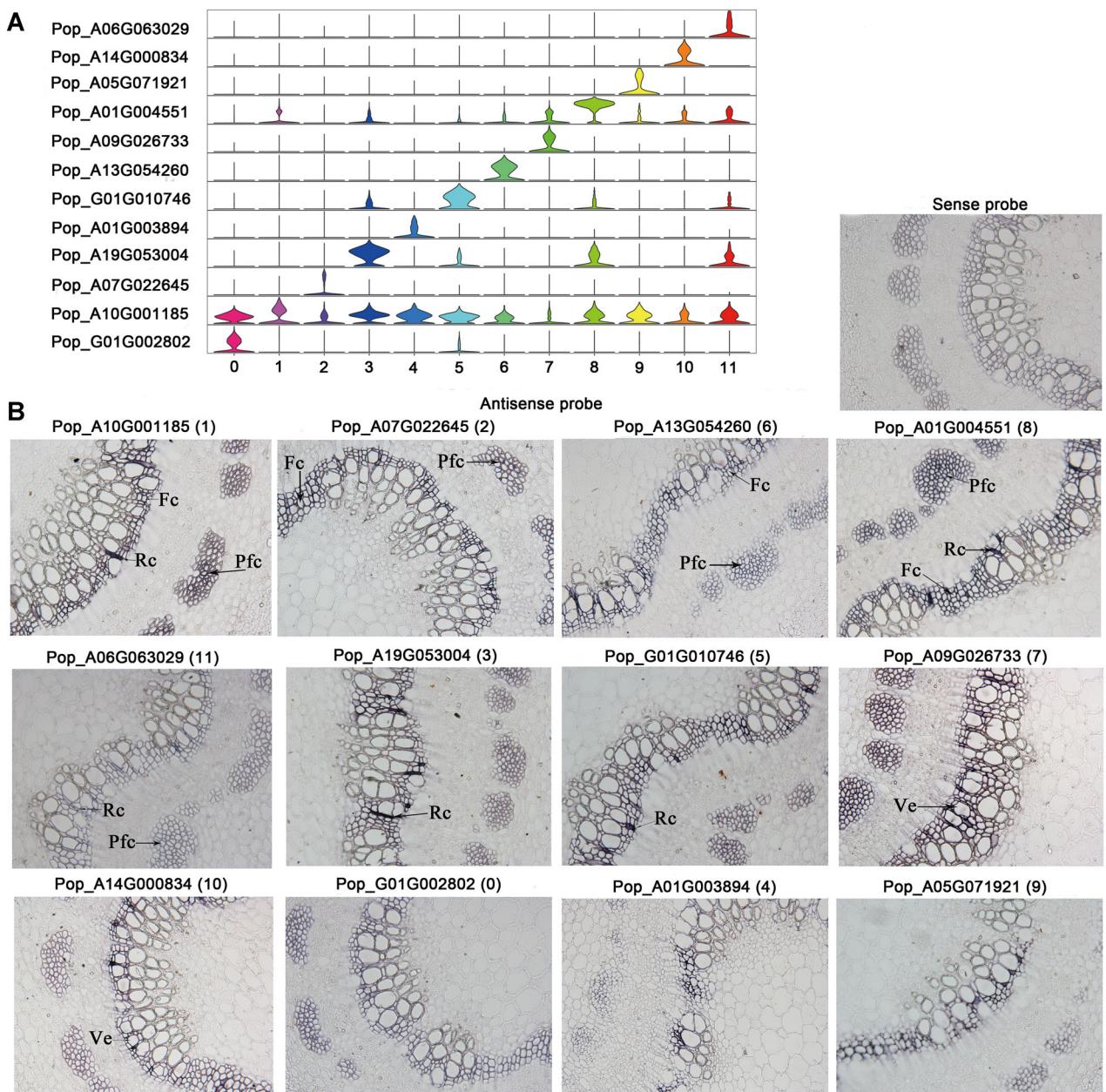


Figure 3. Expression patterns of selected genes in each cluster

(A) Violin graph showing expression levels for each selected gene. The height of the violin represents the gene expression level, and the width of the violin represents the proportion of cells expressing in the cluster. (B) RNA *in situ* hybridization of 12 selected genes, including Pop_G01G002802 in Cluster 0, Pop_A10G001185 in Cluster 1, Pop_A07G022645 in Cluster 2, Pop_A19G053004 in Cluster 3, Pop_A01G003894 in Cluster 4, Pop_G01G010746 in Cluster 5, Pop_A13G054260 in Cluster 6, Pop_A09G026733 in Cluster 7, Pop_A01G004551 in Cluster 8, Pop_A05G071921 in Cluster 9, Pop_A14G000834 in Cluster 10, and Pop_A06G063029 in Cluster 11. Fc: fiber cell; Rc: ray cell; Pfc: phloem fiber cell; Ve: vessel.

vessels. Pop_G01G002802 (for Cluster 0) was expressed in the cells with thin walls in both xylem and phloem. RNA signals for Pop_A01G003894 (for Cluster 4) and Pop_A05G071921 (for Cluster 9) were mainly observed in the cells that may belong to parenchyma cells. We were unable to define the identities of Clusters 0, 4, and 9. Based on the available marker genes and their expression, we attributed these 12 clusters to four main cell types, including xylem precursor cells (Xpc; Cluster 1) that contains fusiform initials and ray initials, fiber cells (Fc; Clusters 2 and 6), vessels (Ve; Clusters 7 and 10), ray parenchyma cells (Rc; Clusters 3 and 5), at different development stages. Clusters 8 and 11 comprise either fiber cells or ray parenchyma cells.

Cell differentiation classified by pseudotime trajectory analysis

To study cell differentiation process during xylem formation, we performed pseudotime trajectory analysis using all 12 clusters (Figure S8A). The developmental process was divided into four states (Figure S8B). Clusters 2, 6, 7, and 10 were mainly distributed in State 1. Clusters 2, 6, and 7 are in the process of SCW thickening, showing that State 1 represents a late stage of vessel and fiber cell SCW formation. Clusters 3 and 5 were mainly distributed in State 3 which was

in the process of ray parenchyma cell differentiation. Clusters 1 and 0 were distributed in States 4 and 5, which were in an early stage of xylogenesis.

Based on the observed gene expression of Clusters 7 and 10, we could conclude that cells in Cluster 10 were at an early stage of vessel differentiation. Consistently, pseudotime analysis using Clusters 1, 7, and 10 showed that the differentiation trajectory is from 1 to 10 to 7 (Figure 4A). We defined Cluster 10 as Vessel Stage 1 (Vel) and Cluster 7 as VeII (Figure 4C).

We used the cells of seven clusters (not including vessel-specific Clusters 7 and 10, and three unknown clusters 0, 4, 9) to study the differentiation trajectories of fiber cells and ray parenchyma cells. In the pseudotime analysis (Figure 4B), Clusters 3 and 5 were concentrated in State 2, and Clusters 2 and 6 were concentrated in States 3 and 4, indicating that States 2 and 4 are the differentiation endpoints for ray parenchyma cells and fiber cells, respectively. Cells in Cluster 11 were mainly distributed in State 2, indicating that it is inclined to ray parenchyma cell differentiation. Cells in Cluster 8 were mainly distributed in both States 2 and 3 and fewer cells were distributed in State 4, suggesting that the cells in Cluster 8 may differentiate into either ray parenchyma cells or fiber cells. By integration of the gene expression results of RISH, the position of clusters in UMAP, and the cell

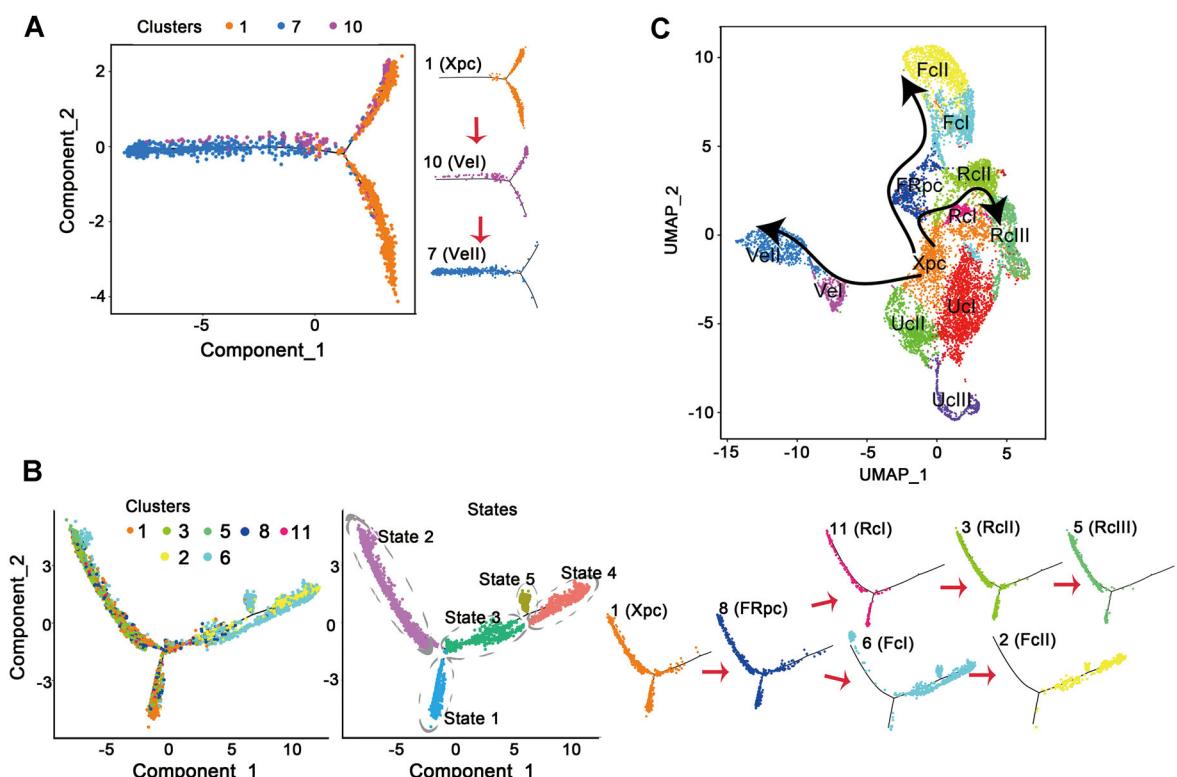


Figure 4. Pseudotime trajectory of vessels, fiber cells and ray parenchyma cells

(A) Pseudotime analysis of Clusters 1, 7, and 10. (B) Pseudotime analysis of Clusters 1, 2, 3, 5, 6, 8, and 11. (C) UMAP graph showing the cell differentiation directions (arrows). Fc: fiber cell; FRpc: fiber and ray parenchyma cell precursor cell; Rc: ray parenchyma cell; Uc: unidentified cell type; Ve: vessel; Xpc: xylem precursor cell. I, II, and III represent different developmental stages. Each dot represents a single cell, and different colors mean different cell types (clusters) or states.

distributions by pseudotime analysis, we could infer the differentiation trajectory for ray parenchyma cell, fiber cell and vessel formation (Figure 4C) and define the cell types of these six clusters, including Cluster 6 (Fiber stage I, Fcl), Cluster 2 (FcII), Cluster 8 (Fiber and ray parenchyma cell precursor cell, FRpc), Cluster 11 (Ray parenchyma cell Stage I, Rcl), Cluster 3 (RcII), and Cluster 5 (RcIII) (Figure 4C).

Marker genes in different cell types

Few marker genes for wood formation are available. To identify more marker genes for xylem formation, we examined the significantly upregulated genes ($\log_2(\text{FC}) \geq 1$, $P \leq 0.05$) in each cluster. We selected 65 markers genes based on their high expression levels in a specific cell type and high fold change comparing to other clusters (Figure 5A; Dataset S7). The markers for Xpc (Cluster 1) and Cluster 0 were not ideal, because they also displayed relatively high expression levels in other clusters. Clusters 6 (Fcl) and 2 (FcII) represent early and late stages in the fiber cell SCW thickening process, respectively. Several genes were selected as markers for different developmental stages during fiber cell SCW thickening (Dataset S7). *PagTUB8* was upregulated in both Fcl

and FcII (Figure 5B), and it can be a marker for the whole SCW thickening process of fiber cells. Several genes, including *PagXCP1* (Figure 5B), *PagXCP2*, *PagMC9* mentioned above (Dataset S3), were specific for SCW thickening in vessels. While *PagLAC12* was upregulated in Fcl, FcII and VelI (Datasets S6, S7), and it could be a marker for SCW thickening, regardless of cell type of vessels or fiber cells. *PagEXP6* was upregulated in Vel, a stage before SCW thickening, showing the cells are differentiating before SCW thickening for vessel formation.

Key regulators of xylogenesis along a pseudotime trajectory and at a branch point

To identify genes controlling cell differentiation, we examined the gene expression along the differentiation trajectory of vessels, fiber cells and ray parenchyma cells. Expression heatmaps showed that vessels, fiber cells and ray parenchyma cells were divided into three, five, and seven gene clusters, respectively (Figure 6A–C). Genes were upregulated in different gene clusters, indicating their involvement at different developmental stages, such as expansin encoding genes (*PagEXP4*, 6, 9) expressed at the early stage for cell

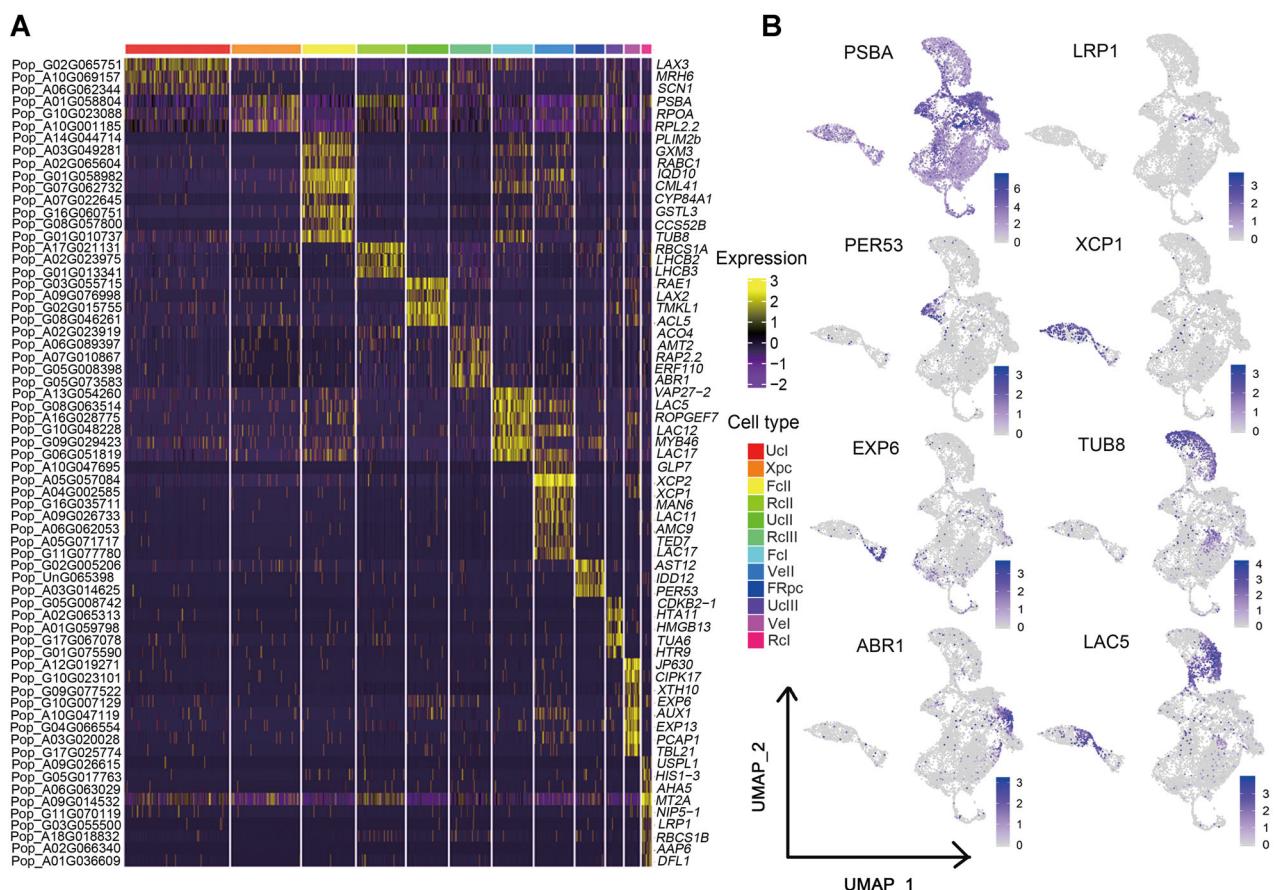


Figure 5. Expression marker genes in 12 clusters

(A) Heatmap showing the expression of marker genes in 12 clusters. High and low expression levels are shown by yellow and purple respectively. **(B)** UMAP graph showing the distribution and expression level of some marker genes. Grey represents low expression levels, and dark blue represents high expression levels.

elongation. *PagXCP1* and *PagLAC17* were expressed at a later stage (PCD) of vessel formation, and cell wall biosynthetic genes were expressed at the stage of SCW thickening. *PagHB15*, encoding a class III homeo-

domain leucine-zipper protein, was upregulated in Cluster 1 (Dataset S8). HD-Zip III proteins are key regulators in procambium and xylem cell differentiation in *Arabidopsis* (Ohashi-Ito and Fukuda, 2003; Ohashi-Ito et al., 2005), and

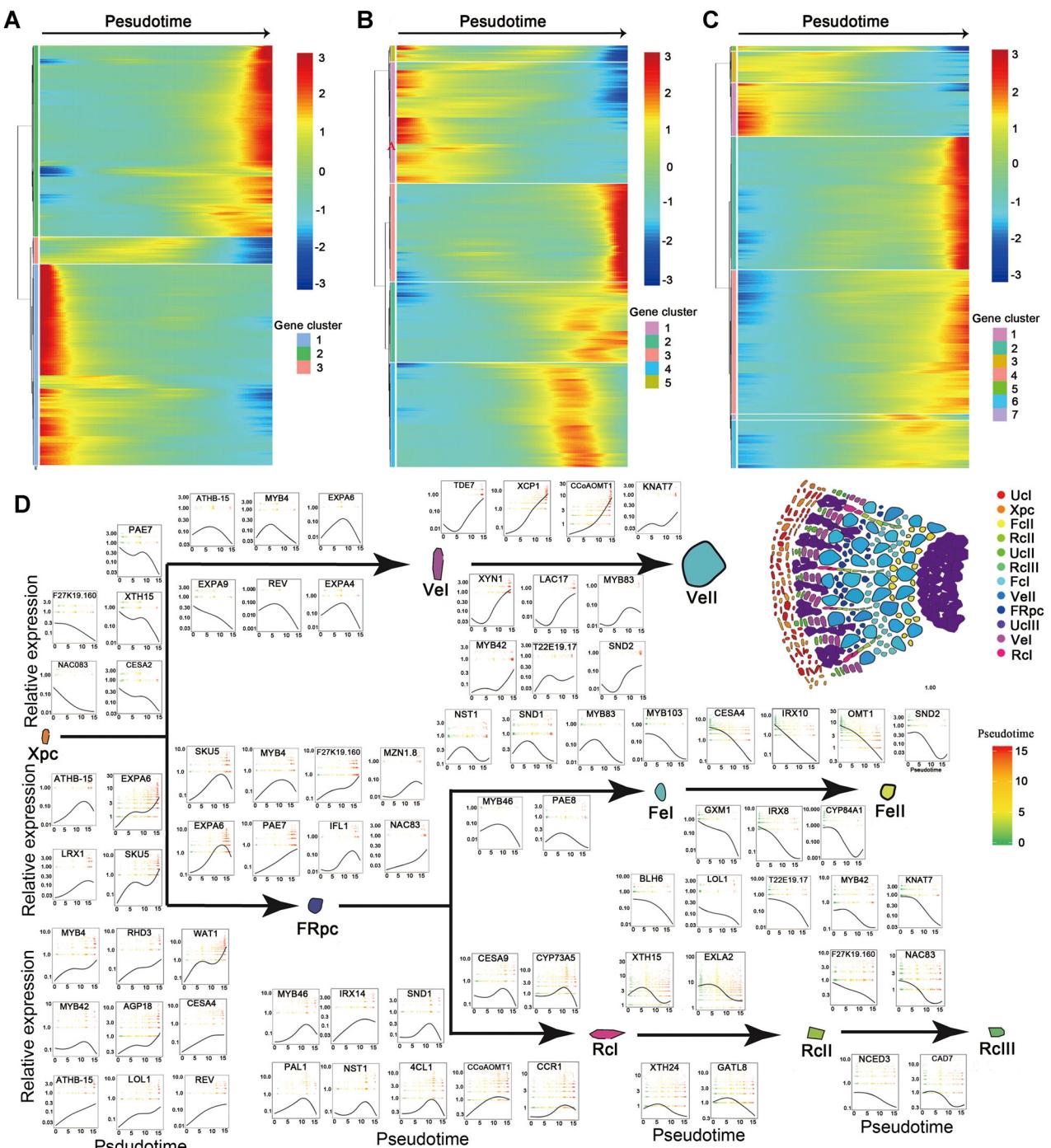


Figure 6. Expression heatmap and key regulators along trajectories during cell differentiation

(A) Gene expression heatmap for vessel-related clusters Xpc (1), Vel (10), and Vell (7). (B) Gene expression heatmap for fiber cell-related clusters Xpc (1), FRpc (8), Fcl (6), and FcII (2). (C) Gene expression heatmap for ray parenchyma cell-related cluster Xpc (1), FRpc (8), Rcl (11), Rcll (3), and Rclll (5). In the heatmaps, blue and red colors represent low and high gene expression levels, respectively. Different gene clusters represent different differentiation states along a pseudotime trajectory. (D) Expression trend graph of selected differential expressed genes along pseudotime trajectory during differentiation for each cell type. Each point in the expression trajectory represents a single cell. Green and red colors represent short and long pseudotime, respectively. All Y-axis labels in (D) are “Relative expression,” and all X-axis labels in (D) are “Pseudotime.”

member ATHB-15 is a pivotal transcriptional regulator for early vascular development, supporting that Cluster 1 consists of xylem precursor cells (Xpc). SND1 is a master regulator activating fiber cell SCW formation (Zhong et al., 2006). As expected, *PagSND1* was upregulated in Fcl, the early stage of fiber SCW biosynthesis. Upregulation of *PagSND1* first appeared in FRpc (Dataset S8), indicating that *PagSND1* started its activation on SCW thickening in the cells of FRpc. Consistently, the SND1 targets *PagMYB46/83* were upregulated following SND1 in Fcl (Figure 6D; Dataset S8).

To identify the key genes controlling the transition to differentiated states for vessels, fiber cells and ray parenchyma cells, we performed Branching Expression Analysis Modeling (BEAM), which uses penalized splines to infer the individual gene branching time (Qiu et al., 2017a). Vessel, fiber cell and ray parenchyma cell differentiation trajectories had one, two, three branches, respectively (Figures S9–11). We examined the TFs that were differentially expressed at the branch points and identified some key regulator candidates (Dataset S9). *PagSND2* and *PagMYB42* may be important regulators in vessel differentiation. During fiber cell differentiation, *PagSND1* may initiate fiber cell differentiation and SCW thickening at two branch points during fiber cell differentiation. Further studies are needed to verify the functions of the TFs identified in both differentiation trajectories and branch points in controlling xylem cell fate.

DISCUSSION

In this study, scRNA-seq was carried out for cells in differentiating xylem of *P. alba* × *P. glandulosa* using xylem protoplasts for single cell isolation. The protoplast preparation time was reduced to maintain cell original differentiated states. Cell wall digestion to release protoplast is the major process and its time length determines whole process time for protoplast preparation. A simple xylem protoplast system in *P. trichocarpa* has been generated, with a short cell wall digestion of 20 min (Lin et al., 2014). Our preliminary experiments on protoplast isolation in *P. alba* × *P. glandulosa* showed that longer cell wall digestion times would give more impurities and shorter times would produce fewer protoplasts. We designated the cell wall digestion time to be 1 h, which generated a high yield of protoplasts with fewer impurities. Comparison between scRNA-seq and bulk xylem RNA-seq showed that xylem protoplasts share a high gene identity (85.84%) of their transcriptomes and a high correlation ($K = 0.838$, $R^2 = 0.82$) of gene expression with bulk xylem, indicating that scRNA-seq can reasonably represent the transcriptional state of native xylem in *P. alba* × *P. glandulosa*.

Through examining the upregulated genes and available markers in the 12 clusters, we ascertained that two clusters are composed of cells undergoing SCW thickening of fiber cells, one cluster is composed of the cells undergoing SCW thickening of vessels, and these three clusters are at the later stages of fiber cell and vessel differentiation, respectively.

Further RISH and pseudotime analysis helped us to identify the cell types for 12 clusters, except for Clusters 0, 4, and 9. Whether these Clusters 0, 4, and 9 belong to parenchyma cells needs further investigation. We compared the cell type-specific genes between the Drop-seq method and the LMD method (Shi et al., 2017) and found some fiber- or vessel-specific genes identified by LMD were upregulated in both vessels and fiber cells in the scRNA-seq analysis. The LMD method (Shi et al., 2017) may have contaminations with other types of cells collected, and RNA is needed to be amplified before library construction that may have biased amplifications. scRNA-seq could obtain a true single cell level sequencing compared to LMD-based sequencing. scRNA-seq data provides the information of upregulated genes in each cluster, allowing identification of marker genes for different developmental stages for each cell type. Based on the gene expression in different clusters, it is now possible to investigate the role of specific genes during different developmental stages. For example, *CAD2* (also termed *SAD*) has been proposed to have a role in syringyl monolignol biosynthesis (Li et al., 2001), but its expression level is very low in *P. trichocarpa* xylem (Shi et al., 2010). Although *PagCAD2* was expressed in xylem of *P. alba* × *P. glandulosa*, at a level of 26.82% of *PagCAD1* (Figure S7), it was not upregulated in Clusters 2, 6, and 7 compared to other clusters (Dataset S6). This result indicates that *PagCAD2* is not functioning during the process of SCW thickening in *P. alba* × *P. glandulosa*, in accordance with the conclusion in tobacco (Barakate et al., 2011).

scRNA-seq analysis also provides information on the functional divergence of different family members during wood formation. We observed that specific members in a gene family were always upregulated in specific clusters, showing their involvement in specific developmental stages. For example, *P. alba* × *P. glandulosa* has 32 LAC members in the laccase gene family (Dataset S10), among which 16 were upregulated in the 12 clusters (Dataset S6). Most laccase genes were upregulated in Clusters 2, 6, and 7, supporting their involvement in the deposition of lignin during SCW formation. These laccase genes include *PagLAC7, 11, and 12* (orthologs of *AtLAC17*), *PagLAC14, 21, 40, 4* (orthologs of *AtLAC4*), *PagLAC18, 26* (orthologs of *AtLAC5*), *PagLAC20* (ortholog of *AtLAC2*). *PagLAC30* (ortholog of *AtLAC17*) and *PagLAC22* (ortholog of *AtLAC11*) were upregulated only in Cluster 7 (Vell), indicating these two laccases function in vessel SCW lignin biosynthesis. *PagLAC25* was upregulated in Clusters 2 (Fcll) and 6 (Fcl), and *PagLAC19* was upregulated in Fcl, showing that these two laccases are fiber-specific for monolignol polymerization. The results not only support the redundancy of laccases in lignin biosynthesis in vessels and fiber cells (Berthet et al., 2011; Zhao et al., 2013; Schuetz et al., 2014), but also show vessel- and fiber-specific laccases in monolignol polymerization and that these cell-type specific laccases may polymerize specific lignin monomers.

A major advantage of scRNA-seq is to study the differentiation trajectory. In this study, the differentiation trajectory for vessel elements, fiber cells and ray parenchyma cells are examined. Re-clustering analysis using the nine clusters (excluding Clusters 2, 6, and 7) that do not have SCW thickening grouped Clusters 1, 3, 5, 8, and 11 into Cluster 0' (**Dataset S11**; **Figure S12**), indicating that these five clusters have close relationships. Pseudotime analysis showed that cells in Cluster 8 could differentiate to both fiber cells and ray parenchyma cells, indicating that it has precursor cells for both fibers and rays (**Figure 4**). We observed that *PagSND1* was significantly upregulated in Clusters 8 and 6 (**Datasets S3; S8**), which are destined for fiber cell differentiation. In *Arabidopsis* SND1 is a specific regulator for fiber cell formation. Although there is cross talk between SND and VND in poplar ([Lin et al., 2017](#)), over-expression of a *PtrSND1-A2* splice variant under the *PtrSND1-A2* promoter specifically repressed fiber cell wall thickening ([Li et al., 2012](#); [Zhao et al., 2014](#)), showing that SDN1 function in poplar is specific for fiber cells but not for vessels. Knockout of SND genes in poplar specifically affects SCW of both fiber cells and ray parenchyma cells ([Takata et al., 2019](#)). Consistently, pseudotime analyses suggest that fiber cells and ray parenchyma cells undergo a similar differentiation process even if they are divided from different initial cells, and vessels and fiber cells undergo different differentiation processes from the early stage of their development.

Preparation of protoplasts from phloem was attempted in study, but yield was low and the impurities were high, making it unsuitable for scRNA-seq. Because cambium is attached to the phloem side after bark was peeled, our scRNA-seq is focused on differentiating xylem, not including cambium and phloem. A new protocol for phloem protoplast preparation is needed. We observed that about 10 layer cells were released by 1 h enzyme digestion, which is less than the 15 layer cells that were scraped from the stem for RNA-seq, showing that there are some leftovers for the 1 h enzyme digestion. It is necessary to establish the method of single-nucleus RNA sequencing (sNucRNA-seq) to avoid the problem of leftovers. The more extensive scRNA-seq on cambium, phloem and xylem will provide valuable information on the maintenance of cambium activity and the cell differentiation from cambium to xylem and phloem. The combination of scRNA-seq and spatial transcriptome profiling will aid in the cell differentiation trajectory analysis. The application of this technique may also be expected to provide new insights into cell autonomy and cell interactions during differentiation and xylogenesis *in vivo* and *in vitro*.

MATERIAL AND METHODS

Plant materials

Populus alba × *P. glandulosa* hybrid plants were used in this study. They were grown in a room with long-day conditions (16 h light/8 h dark) at 25 ± 1 °C, with a light intensity of 80 mmol m⁻² s⁻¹.

RNA-seq library construction and sequencing

Samples of leaves, xylem and phloem were collected from three 5-month-old trees. The stem of 80 cm long beneath the 15th node was peeled, and xylem and phloem were scraped with single-edge razors from the xylem and bark sides, respectively. Total RNAs from leaves, phloem and xylem were extracted from three individual trees using the CTAB method ([Chang et al., 1993](#)) for RNA-seq library construction following the protocol of Next Ultra Directional RNA Library Prep Kit (NEB, USA). Nine libraries were sequenced using the Illumina HiSeq 2500 platform with a read length of 150 bp at both sides. The normalized gene transcript abundances in each RNA-seq library were calculated using the TMM (trimmed mean of *M*-values) method via RSEM pipeline ([Li and Dewey, 2011](#)).

Protoplast preparation for differentiating xylem

Protoplasts were prepared from differentiating xylem of a 5-month-old tree as described previously ([Lin et al., 2014](#)). Briefly, after the bark of a stem was peeled off, the stem of 80 cm long below the 15th node was divided into eight fragments and immersed into enzyme solution (20 mM MES, pH 5.7, 0.5 M Mannitol, 20 mM KCl, 1.5% (w/v) Cellulase R10 (Yakult Pharmaceutical Industry, Japan), 0.4% (w/v) Macerozyme R10 (Yakult), 10 mM CaCl₂, 0.1% BSA (w/v)). After digestion at room temperature for 1 h, the stems were transferred into MMG solution (4 mM MES, pH 5.7, 0.5 M Mannitol, 15 mM MgCl₂) to release protoplasts by gently shaking. The protoplast solution was filtered through a 40 µm nylon mesh, centrifuged at 300 g at 4 °C for 5 min, and the pellet was resuspended in 20 mL pre-cooled W1 solution (4 mM MES, pH 5.7, 0.5 M Mannitol, 20 mM KCl) with shaking. After re-centrifugation twice, the protoplasts were resuspended in 500 µL W1 solution.

scRNA-seq library construction and sequencing

The library was constructed using the Chromium Controller and Chromium Single Cell 30 Reagent Kit V3 (10X Genomics, USA). Protoplasts were loaded on a Chromium Controller to generate single-cell gel beads in emulsion (GEMs). scRNA-seq library was prepared using the Chromium single cell 30 gel bead and library kit (P/N #120236, 120237, 120262). The library was paired-end sequenced on an Illumina Hiseq 4000 sequencer. Raw scRNA-seq dataset comprises Read1, Read2, and i7 index read. The 26 bp read length of Read1 contains the sequence of 16 bp 10x Barcode and 10 bp UMI which locate at the 3'UTR region for gene identification. The 98 bp read length of Read2 is cDNA sequences.

Optimization of gene annotation in *P. alba* × *P. glandulosa* genome

P. alba × *P. glandulosa* genome annotation lacks 5' and 3' UTR sequences ([Qiu et al., 2019](#)), limiting gene identification in our scRNA-seq analysis. We augmented the UTR sequences to improve transcript assembly. HISAT2 ([Kim et al., 2015](#)) was used to align the bulk RNA data to the original

genome file, and Stringtie (V2.0) (Pertea et al., 2015) was used to reconstruct the transcripts to produce a gtf file. Cuffcompare (V2.0.2) (Roberts et al., 2011) was used to compare the original gtf file with the new one, generating a new gtf file.

Cell clustering and DEGs analysis by Seurat

We employed the Seurat package for cell clustering analysis. Cells with the QC metrics of gene counts between 500 and 10,000 per cell and UMI counts less than 30,000 per cell were retained. Gene expression was normalized using the global-scaling normalization method LogNormalize according to the following formula:

$$\text{GeneA}_{\text{Expression}} = \log \left(1 + \frac{\text{UMI}_A}{\text{UMI}_{\text{Total}}} \times 10,000 \right)$$

In this formula, UMI represents Unique Molecular Identifier, UMI_A is Unique Molecular Identifier of geneA, $\text{UMI}_{\text{Total}}$ is Unique Molecular Identifier of total genes.

The principal component analysis (PCA) was used to overcome technical noise and reduce dimensionality. The JackStraw procedure was used to construct the null distribution of gene scores and to identify significant principal components (PCs) (Saadatpour et al., 2015). For PC analysis, the scaled data was reduced to 51 approximate PCs depending on the 25,011 highly variable genes (set n_{pc}s = 51). Cell distances were calculated based on PC results.

We used "FindVariableFeatures" with "mean.cutoff (0.01–8) and dispersion.cutoff (0.5, Inf)" to calculate highly variable genes. A total of 25,011 genes were used for clustering analysis. Clusters were identified using the Seurat function "FindClusters" with "resolution = 0.6". The data structures and cell trajectories were separately visualized and explored by t-SNE (run the "RunTSNE" function with "dims = 44") and UMAP (run the "RunUMAP" function with "n.neighbors = 30 L", "metric = cosine", and "min.dist = 0.3"). Cluster marker genes (cluster-enriched genes) were identified using Seurat function "FindAllMarkers" and tested by "bimod". The cluster-enriched genes were detected by parameters of "min.pct = 0.25" and "logfc.threshold = 1", which means that the minimum cell percentage for marker genes is more than 0.25 and the log₂ fold change of average expression is more than 1.

GO and KEGG analyses of the 12 clusters

Gene Ontology (GO) enrichment analysis of DEGs of the 12 clusters was conducted by Goseq and topGO (Young et al., 2010). We mapped DEGs to GO terms in the database (<http://geneontology.org/>), calculated the gene numbers for each term, and found significantly enriched GO terms. Candidate target genes were compared to the reference gene background using a hypergeometric test (sampling without replacement). According to GO annotation, DEGs were classified. The Phyper function in the R software package was used for function enriched analysis.

Xylem cell differentiation by scRNA-seq

DEGs were analyzed in the KEGG database (<https://www.kegg.jp/>) to identify enriched metabolic or signal transduction pathways. Target gene candidates compared to the whole reference gene background using hypergeometric test. *p* value in GO and KEGG analysis was calculated by the following formula:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

In the formula above, N is the number of all genes with GO and KEGG annotation, n is the number of target gene candidates in N , M is the number of all genes that are annotated to a certain GO term and KEGG pathway, and m is the number of target gene candidates in M . We corrected *p* values using the Bonferroni correction as a Q value.

Pseudo-time trajectories construction, differential expression and branching expression analyses for putative cell types

The Monocle (Version 2.10.1) (Trapnell et al., 2014) package was used to construct cell differentiation trajectories for selected clusters. The raw count in the Seurat object was first converted into the CellDataSet with the importCDS (object, import_all = F). We then used the estimateSizeFactors () function to normalize the differences in mRNA recovered across cells and the estimateDispersions() function to perform the differential expression analysis re-normalize the data. High variant genes are used for identifying cell subpopulations or ordering cells along a trajectory. We selected these genes via the disp_table() and ordering_genes() with standard: mean_expression ≥ 0.5 and dispersion_empirical ≥ dispersion values.

The ordered genes were marked with the setOrderingFilter () function. The dimensional reduction clustering analysis was performed with the reduce Dimension() function, max_components = 2, reduction_method = "DDRTree," then with the trajectory inference ('orderCells' function) with default parameters. Gene expression was then plotted as a function of pseudo-time in Monocle to track changes across pseudo-time. Gene expression tracks were visualized by plot_genes_in_pseudotime(), gene expression heatmaps were visualized by plot_pseudotime_heatmap(). To illustrate cell differentiation process and differentiation of cells into different branches, Branching Expression Analysis Modeling (BEAM) algorithm (Qiu et al., 2017b) in Monocle software was used to test the branch-dependent gene expression between two negative binomial generalized linear models (GLMs). Branch-dependent gene expression heatmaps were visualized by plot_genes_branches_heatmap().

RISH

The 8th internode stems were used for RISH. PCR fragments about 250 bp were amplified using gene-specific primers

(Dataset S12) to synthesize antisense and sense probes. Hybridization and immunological detection were conducted as previously described (Liu et al., 2018).

Weighted gene co-expression network analysis (WGCNA) within clusters

The WGCNA package in R software was used in co-expression analysis to detect relative relationships among genes. A weighted adjacency matrix was created following unsupervised hierarchical clustering analysis of genes, and the soft threshold power (β) was set at 15 to analyze scale-free topology. Modules were identified by the following parameters: "power = 15 minModuleSize = 30, mergeCutHeight = 0.25." The co-expression network was visualized using Cytoscape3.7.1.

Data availability statement

RNA-seq and scRNA-seq data have been deposited in GenBank (BioProject ID: PRJNA703312).

ACKNOWLEDGEMENTS

We thank Guangzhou Genedenovo Biotechnology for assisting in sequencing and bioinformatics analysis. This work was supported by grants from Fundamental Research Funds of Chinese Academy of Forestry (CAFYBB2018ZY001-5 and CAFYBB2017ZY001) and the National Natural Science Foundation of China (31670667).

AUTHOR CONTRIBUTIONS

Q.L. conceived the project; H.L., M.X., and Q.W. performed protoplast isolation; H.L., X.H., X.Y., R.R.S., and Q.L. conducted data analyses; X.D. and H.L. performed RISH experiment; H.L., R.R.S., and Q.L. wrote the manuscript. All authors read and approved of this manuscript.

Edited by: Jianbing Yan, Huazhong Agricultural University, China

Received May 18, 2021; **Accepted** Aug. 2, 2021; **Published** Aug. 4, 2021

REFERENCES

- Abbas, M., Peszlen, I., Shi, R., Kim, H., Katahira, R., Kafle, K., Xiang, Z., Huang, X., Min, D., Mohamadamin, M., Yang, C., Dai, X., Yan, X., Park, S., Li, Y., Kim, S.H., Davis, M., Ralph, J., Sederoff, R.R., Chiang, V.L., and Li, Q. (2020). Involvement of CesA4, CesA7-A/B and CesA8-A/B in secondary wall formation in *Populus trichocarpa* wood. *Tree Physiol.* **40**: 73–89.
- Avci, U., Petzold, H.E., Ismail, I.O., Beers, E.P., and Haigler, C.H. (2008). Cysteine proteases XCP1 and XCP2 aid micro-autolysis within the intact central vacuole during xylogenesis in *Arabidopsis* roots. *Plant J.* **56**: 303–315.
- Barakate, A., Stephens, J., Goldie, A., Hunter, W.N., Marshall, D., Hancock, R.D., Lapierre, C., Morreel, K., Boerjan, W., and Halpin, C. (2011). Syringyl lignin is unaltered by severe sinapyl alcohol dehydrogenase suppression in tobacco. *Plant Cell* **23**: 4492–4506.
- Beers, E.P., Woffenden, B.J., and Zhao, C.S. (2000). Plant proteolytic enzymes: Possible roles during programmed cell death. *Plant Mol. Biol.* **44**: 399–415.
- Berthet, S., Demont-Caulet, N., Pollet, B., Bidzinski, P., Cézard, L., Le Bris, P., Borrega, N., Hervé, J., Blondet, E., Balzergue, S., Lapierre, C., and Jouanin, L. (2011). Disruption of *LACCASE4* and 17 results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* **23**: 1124–1137.
- Bollhoner, B., Zhang, B., Stael, S., Denancé, N., Overmyer, K., Goffner, D., Breusegem, F.V., and Tuominen, H. (2013). Post mortem function of AtMC9 in xylem vessel elements. *New Phytol.* **200**: 498–510.
- Chang S., Puryear J., and Cairney J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11**: 113–116.
- Chen, H., Wang, J.P., Liu, H., Li, H., Lin, Y.-C.J., Shi, R., Yang, C., Gao, J., Zhou, C., Li, Q., Sederoff, R.R., Li, W., and Chiang, V.L. (2019). Hierarchical transcription factor and chromatin binding network for wood formation in black cottonwood (*Populus trichocarpa*). *Plant Cell* **31**: 602–626.
- Denyer, T., Ma, X., K Lesen, S., Scacchi, E., and Timmermans, M. (2019). Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single cell RNA sequencing. *Dev. Cell* **48**: 840–852.
- Evert, R.F. (2006). *Esau's Plant Anatomy: Meristem, Cells, and Tissue of the Plant Body: Their Structure, Function, and Development*. 3rd edition. Hoboken, NJ: Wiley & Sons.
- Feuillet, C., Lauvergeat, V., Desearche, C., Pilate, G., Boudet, A., and Grima-Pettenati, J. (1995). Tissue- and cell-specific expression of a cinnamyl alcohol dehydrogenase promoter in transgenic poplar plants. *Plant Mol. Biol.* **27**: 651–667.
- Funk, V., Kositsup, B., Zhao, C., and Beers, E.P. (2002). The *Arabidopsis* xylem peptidase XCP1 is a tracheary element vacuolar protein that may be a papain ortholog. *Plant Physiol.* **128**: 84–94.
- Guerriero, G., Sergeant, K., and Hausman, J.F. (2014). Wood biosynthesis and typologies: A molecular rhapsody. *Tree Physiol.* **34**: 839–855.
- Ito, J., and Fukuda, H. (2002). ZEN1 is a key enzyme in the degradation of nuclear DNA during programmed cell death of tracheary elements. *Plant Cell* **14**: 3201–3211.
- Jean-Baptiste, K., McFainline-Figueroa, J.L., Alexandre, C.M., Dorrrity, M.W., and Cuperus, J.T. (2019). Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* **31**: 993–1011.
- Jokipii-Lukkari, S., Sundell, D., Nilsson, O., Hvidsten, T.R., Street, N.R., and Tuominen, H. (2017). NorWood: A gene expression resource for evo-devo studies of conifer wood development. *New Phytol.* **216**: 482–494.
- Joshi, C.P., Thammannagowda, S., Fujino, T., Gou, J.Q., Avci, U., Haigler, C.H., McDonnell, L.M., Mansfield, S.D., Mengesha, B., Carpita, N.C., Harris, D., Debolt, S., and Peter, G.F. (2011). Perturbation of wood cellulose synthesis causes pleiotropic effects in transgenic aspen. *Mol. Plant* **4**: 331–345.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**: 357–360.
- Kim J.S., and Daniel G. (2012). Immunolocalization of hemicelluloses in *Arabidopsis thaliana* stem. Part II: Mannan deposition is regulated by phase of development and its patterns of temporal and spatial distribution differ between cell types. *Planta* **236**: 1367–1379.
- Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zollner, N., Hartwig, T., Wudick, M.M., Lercher, M., Chen, L.Q., Timmermans, M.C.P., and Frommer, W.B. (2021). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* **33**: 511–530.

- Kivioja, T., VaHaraatio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J.** (2011). Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**: 72–74.
- Kuriyama, H.** (1999). Loss of tonoplast integrity programmed in tracheary element differentiation. *Plant Physiol.* **121**: 763–774.
- Kuriyama, H., and Fukuda, H.** (2002). Developmental programmed cell death in plants. *Curr. Opin. Plant Biol.* **5**: 568–573.
- Li, B., and Dewey, C.N.** (2011). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.
- Li, L., Cheng, X., Leshkevich, J., Umezawa, T., Harding, S.A., and Chiang, V.** (2001). The last step of syringyl monolignol biosynthesis in angiosperm is regulated by a novel gene encoding synapyl alcohol dehydrogenase. *Plant Cell* **13**: 1567–1586.
- Li, Q., Lin, Y.-C., Sun, Y.H., Song, J., Chen, H., Zhang, X.H., Sederoff, R.R., and Chiang, V.L.** (2012). Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. USA* **109**: 14699–14704.
- Li, Q., Song, J., Peng, S., Wang, J.P., Qu, G.Z., Sederoff, R.R., and Chiang, V.L.** (2014). Plant biotechnology for lignocellulosic biofuel production. *Plant Biotechnol. J.* **12**: 1174–1192.
- Lin, Y.C., Li, W., Chen, H., Li, Q., Sun, Y.H., Shi, R., Lin, C.Y., Wang, J.P., Chen, H.C., Chuang, L., Qu, G.-Z., Sederoff, R.R., and Chiang, V.L.** (2014). A simple improved-throughput xylem protoplast system for studying wood formation. *Nat. Protoc.* **9**: 2194–2205.
- Lin, Y.C., Li, W., Sun, Y.H., Kumari, S., Wei, H., Li, Q., Tunlaya-Anukit, S., Sederoff, R.R., and Chiang, V.L.** (2013). SND1 transcription factor-directed quantitative functional hierarchical genetic regulatory network in wood formation in *Populus trichocarpa*. *Plant Cell* **25**: 4324–4341.
- Lin, Y.-C.J., Chen, H., Li, Q., Li, W., Wang, J.P., Shi, R., Tunlaya-Anukit, S., Shuai, P., Wang, Z., Ma, H., Li, H., Sun, Y.H., Sederoff, R., and Chiang, V.L.** (2017). Reciprocal cross-regulation of VND and SND multigene TF families for wood formation in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. USA* **114**: E9722–E9729.
- Liu, C., Xue, Z., Tang, D., Shen, Y., Shi, W., Shi, W., Ren, L., Du, G., Li, Y., and Cheng, Z.** (2018). Ornithine δ-aminotransferase is critical for floret development and seed setting through mediating nitrogen re-utilization in rice. *Plant J.* **96**: 842–854.
- Liu, Q., Liang, Z., Feng, D., Jiang, S., Wang, Y., Du, Z., Li, R., Hu, G., Zhang, P., Ma, Y., Lohmann, J.U., and Gu, X.** (2021). Transcriptional landscape of rice roots at the single cell resolution. *Mol. Plant* **14**: 384–394.
- Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J., Wu, R., Zhang, B., Hu, Y., Sun, Y., Shangguan, Y., Li, W., Li, T., Hu, Y., Guo, C., Rochaix, J.-D., Miao, Y., and Sun, X.** (2020). Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol. Plant* **13**: 1178–1193.
- Mellerowicz, E.J., and Sundberg, B.** (2008). Wood cell walls: Biosynthesis, developmental dynamics and their implications for wood properties. *Curr. Opin. Plant Biol.* **11**: 293–300.
- Nelms, B., and Walbot, V.** (2019). Defining the developmental program leading to meiosis in maize. *Science* **364**: 52–56.
- Ohashi-Ito, K., and Fukuda, H.** (2003). HD-Zip III homeobox genes that include a novel member, ZeHB-13 (*Zinnia*)/ATHB-15 (*Arabidopsis*), are involved in procambium and xylem cell differentiation. *Plant Cell Physiol.* **44**: 1350–1358.
- Ohashi-Ito, K., Kubo, M., Demura, T., and Fukuda, H.** (2005). Class III homeodomain leucine-zipper proteins regulate xylem cell differentiation. *Plant Cell Physiol.* **46**: 1646–1656.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L.** (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**: 290–295.
- Petzold, H.E., Rigoulot, S.B., Zhao, C., Chanda, B., Sheng, X., Zhao, M., Jia, X., Dickerman, A.W., Beers, E.P., and Brunner, A.M.** (2018). Identification of new protein-protein and protein-DNA interactions linked with wood formation in *Populus trichocarpa*. *Tree Physiol.* **38**: 362–377.
- Plomion, C., Leprovost, G., and Stokes, A.** (2001). Wood formation in trees. *Plant Physiol.* **127**: 1513–1523.
- Qiu, D., Bai, S., Ma, J., Zhang, L., Shao, F., Zhang, K., Yang, Y., Sun, T., Huang, J., Zhou, Y., Galbraith, D.W., Wang, Z., and Sun, G.** (2019). The genome of *Populus alba* × *Populus tremula* var. *glandulosa* clone 84K. *DNA Res.* **26**: 423–431.
- Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C.** (2017a). Single-cell mRNA quantification and differential analysis with census. *Nat. Methods* **14**: 309–315.
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C.** (2017b). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**: 979–982.
- Rennie, E.A., and Scheller, H.V.** (2014). Xylan biosynthesis. *Curr. Opin. Biotechnol.* **26**: 100–107.
- Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L.** (2011). Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinformatics* **27**: 2325–2329.
- Ryu, H.K., Huang, L., Kang, H.M., and Schiefelbein, J.** (2019). Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* **179**: 1444–1456.
- Saadatpour, A., Lai, S., Guo, G., and Yuan, G.C.** (2015). Single-cell analysis in cancer genomics. *Trends Genet.* **31**: 576–586.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A.** (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**: 495–502.
- Satterlee, J.W., Strable, J., and Scanlon, M.J.** (2020). Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. USA* **117**: 33689–33699.
- Schuetz, M., Benske, A., Smith, R.A., Watanabe, Y., Tobimatsu, Y., Ralph, J., Demura, T., Ellis, B., and Samuels, A.L.** (2014). Laccases direct lignification in the discrete secondary cell wall domains of protoxylem. *Plant Physiol.* **166**: 798–807.
- Serrano-Ron, L., Perez-Garcia, P., Sanchez-Corriónero, A., Gude, I., Cabrera, J., Ip, P.L., Birnbaum, K.D., and Moreno-Risueno, M. A.** (2021). Reconstruction of lateral root formation through single-cell RNA-seq reveals order of tissue initiation. *Mol. Plant* **14**: 1362–1378.
- Shi, R., Sun, Y.H., Li, Q., Heber, S., Sederoff, R., and Chiang, V.L.** (2010). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: Transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol.* **51**: 144–163.
- Shi, R., Wang, J.P., Lin, Y.-C., Li, Q., Sun, Y.-H., Chen, H., Sederoff, R.R., and Chiang, V.L.** (2017). Tissue and cell-type co-expression networks of transcription factors and wood component genes in *Populus trichocarpa*. *Planta* **245**, 927–938.
- Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., and Dickel, D.E.** (2019). High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**: 2241–2247.e2244.
- Song, Q., Ando, A., Jiang, N., Ikeda, Y., and Chen, Z.J.** (2020). Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in *Arabidopsis* female gametophytes. *Genome Biol.* **21**: 178.
- Sundell, D., Street, N.R., Kumar, M., Mellerowicz, E.J., Kucukoglu, M., Johnsson, C., Kumar, V., Mannapperuma, C., Delhomme, N., and Nilsson, O.** (2017). AspWood: High-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in *Populus tremula*. *Plant Cell* **29**: 1585–1604.
- Suzuki, S., Li, L., Sun, Y.H., and Chiang, V.L.** (2006). The cellulose synthase gene superfamily and biochemical functions of xylem-specific

- cellulose synthase-like genes in *Populus trichocarpa*. *Plant Physiol.* **142**: 1233–1245.
- Takata, N., Awano, T., Nakata, M.T., Sano, Y., Sakamoto, S., Mitsuda, N., and Taniguchi, T.** (2019). *Populus* NST/SND orthologs are key regulators of secondary cell wall formation in wood fibers, phloem fibers and xylem ray parenchyma cells. *Tree Physiol.* **39**: 514–525.
- Taylor-Teeple, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., and Lamotte, R.** (2015). An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature* **517**: 571–575.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L.** (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**: 381–386.
- Turco, G.M., Rodriguez-Medina, J., Siebert, S., Han, D., Valderrama-Gomez, M.A., Vahldick, H., Shulse, C.N., Cole, B.J., Juliano, C.E., Dickel, D.E., Savageau, M.A., and Brady, S.M.** (2019). Molecular mechanisms driving switch behavior in xylem cell differentiation. *Cell Rep.* **28**: 342–351.
- Vanholme, R., Meester, B.D., Ralph, J., and Boerjan, W.** (2019). Lignin biosynthesis and its integration into metabolism. *Curr. Opin. Biotechnol.* **56**: 230–239.
- Xi, W., Song, D., Sun, J., Shen, J., and Li, L.** (2017). Formation of wood secondary cell wall may involve two type cellulose synthase complexes in *Populus*. *Plant Mol. Biol.* **93**: 419–429.
- Xu, X., Crow, M., Rice, B.R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E., Lu, Z., Wang, L., Fox, N., Wang, X., Drenkow, J., Luo, A., Char, S. N., Yang, B., Sylvester, A.W., Gingeras, T.R., Schmitz, R.J., Ware, D., Lipka, A.E., Gillis, J., and Jackson, D.** (2021). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell* **56**, 557–568.
- Yamaguchi, M., Mitsuda, N., Ohtani, M., Ohme-Takagi, M., Kato, K., and Demura, T.** (2011). VASCULAR-RELATED NAC-DOMAIN7 directly regulates the expression of a broad range of genes for xylem vessel formation. *Plant J.* **66**: 579–590.
- Yeh, C.S., Wang, Z., Miao, F., Ma, H., Kao, C.T., Hsu, T.S., Yu, J.H., Hung, E.T., Lin, C.C., Kuan, C.Y., Tsai, N.C., Zhou, C., Qu, G.Z., Jiang, J., Liu, G., Wang, J.P., Li, W., Chiang, V.L., Chang, T.H., and Lin, Y.J.** (2019). A novel synthetic-genetic-array-based yeast one-hybrid system for high discovery rate and short processing time. *Genome Res.* **29**: 1343–1351.
- Young, M., Wakefield, M.J., Smyth, G.K., and Oshlack, A.** (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **11**: R14.
- Zhang, T.Q., Chen, Y., Liu, Y., Lin, W.H., and Wang, J.W.** (2021a). Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat. Commun.* **12**: 2053.
- Zhang, T.Q., Chen, Y., and Wang, J.-W.** (2021b). A single-cell analysis of the *Arabidopsis* vegetative shoot apex. *Dev. Cell* **56**, 1056–1074.
- Zhang, T.Q., Xu, Z.G., Shang, G.D., and Wang, J.W.** (2019). A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant* **12**: 648–660.
- Zhao, Q., Nakashima, J., Chen, F., Yin, Y., Fu, C., Yun, J., Shao, H., Wang, X., Wang, Z.Y., and Dixon, R.A.** (2013). Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in *Arabidopsis*. *Plant Cell* **25**: 3976–3987.
- Zhao, Y., Sun, J., Xu, P., Zhang, R., and Li, L.** (2014). Intron-mediated alternative splicing of WOOD-ASSOCIATED NAC TRANSCRIPTION FACTOR1B regulates cell wall thickening during fiber development in *Populus* species. *Plant Physiol.* **164**: 765–776.
- Zheng, G., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., and Zhu, J.** (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**: 14049.
- Zhong, R., Demura, T., and Ye, Z.H.** (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell* **18**: 3158–3170.
- Zhong, R., and Ye, Z.H.** (2007). Regulation of cell wall biosynthesis. *Curr. Opin. Plant Biol.* **10**: 564–572.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: <http://onlinelibrary.wiley.com/doi/10.1111/jipb.13159/supinfo>

Dataset S1. GO analysis of 12 clusters

Dataset S2. KEGG analysis of 12 clusters

Dataset S3. DEGs in 12 clusters

Dataset S4. Family members in cellulose, lignin, and xylan biosynthesis

Dataset S5. Comparison of cell type genes between single-cell RNA sequencing (scRNA-seq) and laser microdissection (LMD) method

Dataset S6. Gene expression levels in 12 clusters

Dataset S7. Marker gene selection

Dataset S8. Transcription factors (TFs) involved in the differentiation of three cell types

Dataset S9. Transcription factor (TFs) identified in branch analysis

Dataset S10. Laccase family in *P. alba* × *P. glandulosa*

Dataset S11. Comparison between two cluster analyses

Dataset S12. Primers used in RNA *in situ* hybridization (RISH)

Figure S1. Quality control of single-cell RNA sequencing (scRNA-seq)

(A) Cells with gene numbers from 500 to 10,000 were retained. The cells with fewer than 500 genes are considered low quality, and the cells with more than 10,000 genes are likely to be two or more cells in one drop. (B) Cells with 30,000 unique molecular identifiers (UMIs) were retained. The cells with UMI numbers more than 30,000 are likely two or more cells in one drop. (C) Venn diagram of gene comparison between scRNA-seq and RNA-seq. 85.84% of the transcripts obtained from single cell sequencing are found in differentiating xylem. (D) Pearson correlation analysis. Each point represents a unique gene. The bulk xylem RNA-seq was from three biological replicates.

Figure S2. Distribution of related index in each cluster

(A) Cell numbers. (B) Median genes per cell. (C) Median unique molecular identifier (UMI) counts per cell.

Figure S3. WGCNA analysis of 12 clusters

(A) Selection process of a soft threshold. (B) Classification process of modules. (C) Gene expression heatmap in different modules. (D) Correlation heat map between different modules and clusters. Green color means low correlations, and red color means high correlations.

Figure S4. Gene expression heatmap in different modules

The heatmaps show gene expression patterns in different clusters. Green color in heatmaps means downregulation, and red color in heatmaps means upregulation. Histograms represent eigenvalue of modules in different clusters.

Figure S5. Gene co-expression network within 12 clusters

(A) Rectangles with different colors represent different modules. Big points represent hub genes in the network. Red color means wood formation related-genes. (B) OliveDrab and MidnightBlue color represent low expression level and high expression level of corresponding genes respectively. Point size represent expressed proportion of corresponding gene in all of cells. The top 1,000 coexpression relationships are shown in each module.

Figure S6. Enriched lignin biosynthesis by KEGG

analysis In the phenylpropanoid biosynthetic pathway (ko00940), the enriched genes (highlighted in red) encode all the enzymes necessary for lignin biosynthesis.

Figure S7. Transcript abundances of secondary cell wall (SCW) pathway genes in differentiating xylem of *P. alba* × *P. glandulosa*

(A) Five cellulose synthase A (CesA) genes for cellulose biosynthesis. (B) Seven genes in xylan biosynthesis, including *IRX8-1, -2, PARVUS-1, -2, -L-2, GT43A, B, C, D, FRA8, IRX10-1, -2, -L-A1, GUX1a, 1b, and GUX2*. (C) Twenty genes for monolignol biosynthesis, including phenylalanine ammonia-lyase genes *PAL1, PAL2, PAL3* and *PAL4/5*, cinnamyl alcohol dehydrogenase genes *CAD1* and *CAD2*, p-coumarate CoA ligase gene *4CL3*, hydroxyl cinnamoyl transferase genes *HCT1* and *HCT6*, caffeoyl

shikimate esterase genes *CSE1* and *CSE2*, cinnamoyl-CoA reductase gene *CCR2*, caffeic acid O-methyltransferase gene *COMT2*, caffeoyl-CoA O-methyl transferase genes *CCoAOMT1* and *CCoAOMT2*, *p*-coumaroyl-CoA 3-hydroxylase gene *C3H3*, cinnamate 4-hydroxylase genes *C4H1* and *C4H2*, ferulate/coniferaldehyde 5-hydroxylase genes *CAld5H1* and *CAld5H2*.

Figure S8. Pseudotime trajectory analysis for 12 clusters

(**A**) Distribution of 12 clusters along differential trajectory. Each dot represents a single cell, and different color depicts different clusters. (**B**) Trajectory was divided in four states. Each dot represents a single cell, and different color depicts different states.

Figure S9. DEGs at branch point along vessel differentiation trajectory

(**A**) Three states were generated along the trajectory using the cells from Xpc, Vel, and Vell. Different colors represent different differentiation states of the cells. (**B**) Branch heatmap for wood formation-related genes along trajectory. Pseudotime increases from middle to two sides along X-axis. Blue and red colors represent low and high gene expression levels, respectively.

Figure S10. DEGs at branch point along fiber cell differentiation trajectory

(**A**) Five states were generated along trajectory using the cells of Xpc, FRpc, Fcl, and Fcll. Different colors represent different differentiation states of the cells. (**B, C**) Branch heatmaps for wood formation-related genes along

Xylem cell differentiation by scRNA-seq

trajectory. Pseudotime increases from middle to two sides along X-axis. Blue and red colors represent low and high gene expression levels, respectively.

Figure S11. DEGs at branch point along ray cell differentiation trajectory

(**A**) Seven states were generated along the trajectory using the cells from Xpc, FRpc, Rcl, Rcll, and Rclll. Different colors represent different differentiation states of the cells. (**B–D**) Branch heatmaps for wood formation-related genes along trajectory. Pseudotime increases from middle to two sides along X-axis. Blue and red colors represent low and high gene expression levels, respectively.

Figure S12. Cluster analysis using the cells from the nine clusters that do not have secondary cell wall (SCW) thickening

(**A**) t-SNE plot of all cells, generating 12 Clusters 0–11 (also shown in Figure 1). (**B**) t-SNE plot of the cells from nine clusters 0, 1, 3, 4, 5, 8, 9, 10, and 11, generating five clusters 0'–4'. Clusters 2, 6, and 7 are also included in the figure. Clusters are shown in different colors, and each dot represents a single cell. Comparison A and B shows that Cluster 0' is the sum of Clusters 1, 3, 8, 11, and partial 5 (38.43%), Cluster 1' is matching Cluster 0, Cluster 2' is sum of Clusters 4 and 9, Cluster 4' is matching Cluster 10, Cluster 3' matching half of Cluster 5 (Dataset S10).

Table S1. Basic statistics information of single-cell RNA sequencing (scRNA-seq) results



Scan using WeChat with your smartphone to view JIPB online



Scan with iPhone or iPad to view JIPB online