

























# Tissue-specific study across the stem reveals the chemistry and transcriptome dynamics of birch bark

Juan Alonso-Serra<sup>1,2,3\*</sup> , Omid Safronov<sup>1,2\*</sup> , Kean-Jin Lim<sup>2,4,5\*</sup> , Sara J. Fraser-Miller<sup>6,7</sup> ,  
Olga B. Blokhina<sup>1,2</sup> , Ana Campilho<sup>8</sup> , Sun-Li Chong<sup>9,5</sup> , Kurt Fagerstedt<sup>1,2</sup> , Raisa Haavikko<sup>6</sup> ,  
Ykä Helariutta<sup>1,2,3,10</sup> , Juha Immanen<sup>1,2,3,11</sup> , Jaakko Kangasjärvi<sup>1,2</sup> , Tiina J. Kauppila<sup>6</sup> ,  
Mari Lehtonen<sup>12</sup> , Laura Ragni<sup>13</sup> , Sitaram Rajaraman<sup>1,2</sup> , Riikka-Marjaana Räsänen<sup>6</sup> ,  
Pezhman Safdari<sup>1,2</sup> , Maija Tenkanen<sup>9</sup> , Jari T. Yli-Kauhaluoma<sup>6</sup> , Teemu H. Teeri<sup>2,4</sup> ,  
Clare J. Strachan<sup>6</sup> , Kaisa Nieminen<sup>11</sup>  and Jarkko Salojärvi<sup>14,1,2,15</sup> 

<sup>1</sup>Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, 00014 Helsinki, Finland; <sup>2</sup>Viikki Plant Science Centre, University of Helsinki, Helsinki, 00014 Helsinki, Finland; <sup>3</sup>Institute of Biotechnology, University of Helsinki, Helsinki, 00014 Helsinki, Finland; <sup>4</sup>Department of Agricultural Sciences, University of Helsinki, Helsinki, 00014 Helsinki, Finland; <sup>5</sup>State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Lin'an, 311300 Hangzhou, China; <sup>6</sup>Drug Research Program, Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, 00014 Helsinki, Finland; <sup>7</sup>The Dodd-Walls Centre for Photonic and Quantum Technologies, Department of Chemistry, University of Otago, 9054 Dunedin, New Zealand; <sup>8</sup>Research Center in Biodiversity and Genetic Resources, Department of Biology, Faculty of Sciences, University of Porto, 4485-661 Porto, Portugal; <sup>9</sup>Department of Food and Nutrition, University of Helsinki, 00014 Helsinki, Finland; <sup>10</sup>Sainsbury Laboratory, University of Cambridge, Cambridge, CB2 1LR, UK; <sup>11</sup>Natural Resources Institute Finland (Luke), 00710 Helsinki, Finland; <sup>12</sup>Laboratory Center, Finnish Environment Institute (SYKE), 00790 Helsinki, Finland; <sup>13</sup>ZMBP-Center for Plant Molecular Biology, University of Tübingen, D-72076 Tübingen, Germany; <sup>14</sup>School of Biological Sciences, Nanyang Technological University, 637551 Singapore, Singapore; <sup>15</sup>Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, 637551 Singapore, Singapore

## Summary

Authors for correspondence:

Jarkko Salojärvi

Tel: +65 69047231

Email: jarkko@ntu.edu.sg

Kaisa Nieminen

Tel.: +358 295325274

Email: kaisa.p.nieminen@luke.fi

Clare J. Strachan

Tel.: +358 294159736

Email: clare.strachan@helsinki.fi

Teemu H. Teeri

Tel.: +358 294158380

Email: teemu.teeri@helsinki.fi

Received: 26 October 2018

Accepted: 25 January 2019

*New Phytologist* (2019) **222**: 1816–1831

doi: 10.1111/nph.15725

**Key words:** bark, *Betula pendula* (silver birch), cambium, genome evolution, metabolic pathways, periderm, phellem, phellogen.

- Tree bark is a highly specialized array of tissues that plays important roles in plant protection and development. Bark tissues develop from two lateral meristems; the phellogen (cork cambium) produces the outermost stem–environment barrier called the periderm, while the vascular cambium contributes with phloem tissues. Although bark is diverse in terms of tissues, functions and species, it remains understudied at higher resolution.
- We dissected the stem of silver birch (*Betula pendula*) into eight major tissue types, and characterized these by a combined transcriptomics and metabolomics approach. We further analyzed the varying bark types within the Betulaceae family.
- The two meristems had a distinct contribution to the stem transcriptomic landscape. Furthermore, inter- and intraspecies analyses illustrated the unique molecular profile of the phellem. We identified multiple tissue-specific metabolic pathways, such as the mevalonate/betulin biosynthesis pathway, that displayed differential evolution within the Betulaceae. A detailed analysis of suberin and betulin biosynthesis pathways identified a set of underlying regulators and highlighted the important role of local, small-scale gene duplication events in the evolution of metabolic pathways.
- This work reveals the transcriptome and metabolic diversity among bark tissues and provides insights to its development and evolution, as well as its biotechnological applications.

## Introduction

Tree bark displays a broad morphological diversity. Most angiosperm and gymnosperm tree species typically have thick and fissured barks (for example *Quercus* spp., *Picea* spp.), while

\*These authors contributed equally to this work.

smooth textures and various colors are found, for example, in cherry trees (*Prunus* spp.) and birches (*Betula* spp.). Anatomically, bark consists of tissues outwards of the vascular cambium: the phloem and periderm, the latter comprised of phelloderm, phellogen and phellem. Bark tissues originate from stem secondary development, which in woody plants takes place in two distinct lateral meristems. The first meristem, vascular cambium, produces the phloem and xylem tissues. The second meristem, phellogen (cork cambium), produces phellem towards outside of the meristem, and phelloderm towards the inside. Phellem constitutes the outermost barrier between the stem and the environment, while phelloderm is typically limited to a few parenchymatic cell layers. This developmental progression remains poorly dissected in trees, but a recent study in *Arabidopsis* described how differentiation steps lead to periderm development starting from pericycle cells (Wunderling *et al.*, 2018).

The fundamental role of bark is to harbor phloem transport and provide protection against environmental factors (Paine *et al.*, 2010), but also to carry out stem photosynthesis (Wittmann & Pfanz, 2007; Vandegehuchte *et al.*, 2015). Bark tissues are rich in secondary metabolites; as such, bark has been a source of medicinal compounds such as salicylic acid and paclitaxel. Recently betulin and related triterpenes extracted from birches have been assessed as multifunctional drugs to treat cancerous and infectious diseases (Pisha *et al.*, 1995; Sun *et al.*, 1998; Liby *et al.*, 2007).

Birches (family Betulaceae, order Fagales) have a broad geographic distribution across the temperate and boreal forest regions (EUFORGEN, 2009; Bret-Harte *et al.*, 2001), and they display a vast diversity of bark phenotypes. White and smooth phellem is a predominant feature but, among the Betulaceae, phellem colors can range from red to yellow. The reason for this color diversity is not known, but it has been suggested that white color reduces the risk of winter sunscald injury (Karels & Boonstra, 2003). Within Betulaceae, the closest related genus to *Betula* (Schenk *et al.*, 2008) is alder (*Alnus*). Unlike birch, alder develops a dark-grey and rough phellem that sheds very slowly, thereby forming a thick and fissured bark, a rhytidome, a feature common to many tree species. The two genera thereby provide a striking contrast of bark phenotypes within the same family, and thus a good comparative system to study bark evolution.

The molecular regulation that takes place in the vascular cambium is well studied because of its direct effect on lignocellulosic biomass production (Etchells *et al.*, 2015; Immanen *et al.*, 2016; Sundell *et al.*, 2017). By contrast, although bark anatomy and composition are well documented, and even some transcriptomes have been analyzed (Park *et al.*, 2008; Mantello *et al.*, 2014; Rosell *et al.*, 2014; Celedon *et al.*, 2017; Rains *et al.*, 2017; Boher *et al.*, 2018), bark development is still poorly understood at a molecular level. This is partially because not all tissues have been dissected and studied in the context of the whole stem, thus making it difficult to address tissue-specificity. Here we explore the molecular fingerprints of eight main tissue types in the stem of *B. pendula*: phellem, combined phellogen and phelloderm, nonconductive secondary phloem, conductive phloem, cambium, developing xylem, xylem, and last year's xylem tissue. We carry out an

integrative study by characterizing the chemical composition and transcriptional profiles from RNA sequencing (RNA-Seq) for each tissue. The approach reveals, for the first time, the ontogeny-dependent transcriptional profile of bark fractions, and identifies both common and specific regulatory components for the phellogen and vascular cambium. Finally, we show that the phellem is exceptionally different within the stem transcriptome, revealing both conserved and diversified expression patterns for metabolic pathways between two Betulaceae species, birch and alder.

The molecular evolution of the genetic mechanisms leading to bark formation is largely unknown. Here we study the pattern of molecular evolution in the gene families expressed in bark tissues. Overall, gene families evolve through gene duplication and loss events. Duplications can be divided into two overall categories: whole-genome multiplications (WGM) and small-scale duplications (SSD), including segmental, tandem, and transposon-induced duplication events (Panchy *et al.*, 2016; Tasdighian *et al.*, 2017). Individual gene families evolve by either one of these processes. After WGMs, transcriptional and developmental regulators and signal transduction components are preferentially retained (Freeling, 2009; Carretero-Paulet & Fares, 2012), whereas genes evolving by SSDs are enriched for environmental responses and secondary metabolism (Panchy *et al.*, 2016); we recently observed similar patterns in silver birch (Salojärvi *et al.*, 2017). A prevailing hypothesis for the retention bias is dosage balance; complex pathways and protein complexes may require that the stoichiometric balance of the interacting components is maintained. Therefore, selection would act against duplications in SSDs, as maintaining dosage balance requires the duplication of all components, whereas after WGMs the losses need to be coordinated (Birchler *et al.*, 2001). As bark is the main protection against environmental factors and a rich source of secondary metabolites we would therefore expect gene families evolving through SSDs to contribute significantly to the transcriptomics profile in bark tissues.

## Materials and Methods

### Plant material

Three 13-yr-old *B. pendula* (v5834) clones were obtained from the experimental field of Viikki Campus (University of Helsinki), and three individuals of *B. davurica* (1990-0384), *B. alleghaniensis* (1990-0384), *B. ermanii* (1993-0483), and *B. papyrifera* (1992-0177) from Kumpula Botanic Garden, Helsinki. Three *Alnus glutinosa* (v7090) clones were provided by LUKE (Natural Resources Institute Finland). Trees were sampled at a 1.5 m height.

### Tissue-specific sampling

Sequential tangential cryosections were collected with three biological replicates per fraction (F) and processed independently. Once a fraction was removed, a cross-section was observed under the microscope to localize the anatomical position. A similar strategy has been used before (Tuominen *et al.*, 1997; Hellgren

*et al.*, 2004; Fagerstedt *et al.*, 2015; Immanen *et al.*, 2016; Sundell *et al.*, 2017). The fractions were divided into eight tissue types (Supporting information Fig. S1). Phellem (F1) layers (L1–L4) from *Betula* species were collected by peeling, whereas *Alnus* was dissected with a microtome blade by tangential sections. Phellogen and phelloderm (F2) were dissected through cryotome sections (10  $\mu\text{m}$ ,  $-27^\circ\text{C}$ ). Phellogen includes the first layer of stem cells and the phelloderm, corresponding to green photosynthetic tissue (max. 15 cell files); this fraction may have minimum traces of F3. Nonconductive secondary phloem (F3) was collected after complete removal of phelloderm uncovering a dark-brown tissue, and thus there are no traces of F2. The conductive phloem (F4) contains the developing and actively transporting phloem, tissue between phloem fibers and cambium. The cambium fraction (F5) consists of gelatinous-like meristematic tissue. The developing xylem (F6) contained soft fibrous xylem tissues. After F6 was removed, xylem (F7) was collected until the annual ring. Finally, xylem tissue (F8) was collected from the previous annual ring. Two series of samples were generated for RNA extraction and chemical analyses. Post RNA-Seq, the expression profiles of tissue-specific marker genes (WOX4, ANT, APL, VND7) were checked to monitor sample purity.

#### Noncellulosic carbohydrate analysis

The tissue samples were milled with a bead mill (30 Hz, 90 s), followed by washing three times in 50% ethanol to remove soluble sugars. The alcohol insoluble residues (AIR) were dried in a freeze dryer and treated by *Bacillus licheniformis*  $\alpha$ -amylase (Megazyme, 5 U  $\text{mg}^{-1}$  AIR) to remove starch. The destarched AIR was washed in water and dried in a freeze dryer. The noncellulosic sugar composition was determined through a small-scale acid methanolysis (Chong *et al.*, 2015). The methyl ester methyl glycosides/methyl glycosides were trimethylsilylated and separated by gas chromatography according to (Chong *et al.*, 2013).

#### Histological sections and suberin staining

Histological sections were processed as described before (Idänheimo *et al.*, 2014) and imaged using a Leica 2500 microscope. Fluorol yellow 088 (Sigma) was used according to (<http://wp.unil.ch/geldnerlab/files/2013/07/Fluorol-Yellow-staining.pdf>). Staining was applied on top of the resin-embedded 5  $\mu\text{m}$  microtome sections.

#### Desorption atmospheric pressure photoionization-mass spectrometry measurements

The DAPPI method allows a direct desorption/ionization of compounds from bark tissue surface, and is described in (Haapala *et al.*, 2007). Briefly, the sample surface was sprayed with hot solvent jet for 8 s, producing gaseous analytes which were ionized with a vacuum ultraviolet lamp and a suitable dopant, here the spray solvent toluene. The resulting ions were detected with MS using positive ion mode, and the mass range was  $m/z$  50–700. A more detailed description is given in Method S1.

#### Quantitative analysis of triterpenes

The bark/wood fractions were pre-cut with scissors, and ground with a Mixer Mill MM 400 for  $3 \times 30$  s at 30 Hz. The milled fractions were freeze dried overnight and stored at  $-20^\circ\text{C}$ ; 10 mg of dried material was weighed, and depending on the expected content of triterpenes, 1–8000  $\mu\text{g}$  of cholesterol (50–10000  $\mu\text{g ml}^{-1}$  in ethanol) was added. Triterpenes were extracted with 500  $\mu\text{l}$  of 80% ethanol at  $80^\circ\text{C}$  for 30 min, followed by extraction with 500  $\mu\text{l}$  of 80% ethanol at room temperature. The sample was centrifuged at 8944  $g$  for 5 min between the extraction steps, and the collected supernatants were combined. The triterpenes were recovered from the supernatant by adding 500  $\mu\text{l}$  of saturated sodium chloride solution, followed by extraction with 1 ml of heptane:diethyl ether (1 : 1, v/v). The extraction was repeated twice. The details of GC–MS analysis are described in Method S2.

#### Infrared spectroscopy

Infrared (IR) absorbance spectra were recorded with a Vertex 70 Fourier Transform Infrared (FTIR) spectrometer (Bruker Optik, Ettlingen, Germany) fitted with a MIRacle<sup>®</sup> single-bounce diamond attenuated total reflection (ATR) accessory (Pike Technologies, Fitchburg, WI, USA). Each sample was measured in triplicate in the bulk region of each phellem layer. The spectra were collected with OPUS 5.5 software (Bruker Optik, Ettlingen, Germany) over the spectral range of 650–4000  $\text{cm}^{-1}$  with 4  $\text{cm}^{-1}$  resolution. Each spectrum was the average of 64 scans. Principal component analysis (PCA) of the IR spectra was carried out using THE UNSCRAMBLER X 10.3 (CAMO, Oslo, Norway) after standard normal variate (SNV) normalization over the spectral region 650–1800  $\text{cm}^{-1}$ . Informative principal components were detected using leave-one-out cross-validation.

#### Total RNA extraction and RNA sequencing

Samples from *B. pendula* and *Alnus glutinosa* cryosections (three biological replicates per fraction) were ground in an Oscillating Mill MM400 by cooling with liquid nitrogen in 30-s intervals. The sample from F1 was cut into small pieces in a mortar in liquid nitrogen prior grinding. Total RNA isolation was carried out as previously described (Lim *et al.*, 2016). Single-end RNA sequencing was conducted for birch at Biomedicum Functional Genomics Unit (FuGU, University of Helsinki) using an Illumina HiSeq2500 platform. *Alnus glutinosa* libraries were made using the Ovation<sup>®</sup> Universal RNA-Seq System from NuGen and custom ribosomal removal oligos. Following control with RNA Nano Chip (Agilent Technologies, Santa Clara, CA, USA) the samples were sequenced with NextSeq500 using 75 bp sequencing kit.

#### RNA-Seq analysis

After quality control with FASTQC, adapter removal and read trimming was carried out with TRIMMOMATIC-0.36 (Bolger *et al.*, 2014). KALLISTO v.0.43.0 (Bray *et al.*, 2016) with 4000 bootstrap

replicates was used to map reads to *B. pendula* and *A. glutinosa* gene models, followed by removal of ribosomal gene models. The k-mer was determined with KMERGENIE-1.6982 (Chikhi & Medvedev, 2014). The final count tables were obtained as the mean of bootstrap replicates. A full description of RNA-Seq data processing and analysis is given in Method S3.

Data accessibility

All sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession code PRJEB29260.

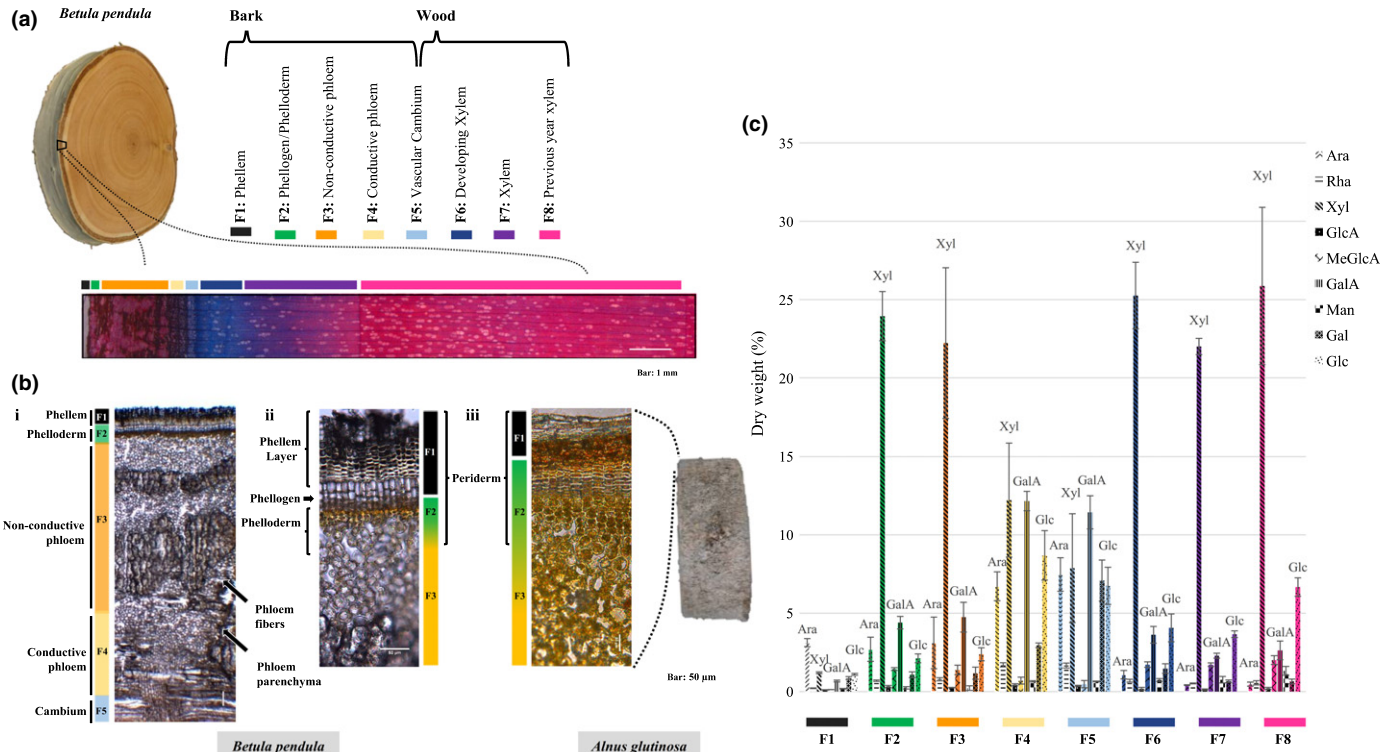
Results

Spatial tissue-specific profiling of tree stem, dissecting bark and wood

Tangential cryosectioning provided a stemwide perspective of bark tissues in *B. pendula*. The stem was dissected into eight anatomically differentiated tissues: phellem (F1), phellogen (cork cambium) and phelloderm (F2), nonconductive phloem (F3), conductive and developing phloem (F4), vascular cambium (F5), developing xylem (F6), mature xylem (F7) and the previous year's xylem tissues, or annual ring (F8); (Fig. 1a). The tissues clearly differed in their anatomy (Fig. 1b, i,ii). Among the periderm tissue samples, the phellem (F1) was composed of a white,

hydrophobic and multilayered tissue, and fraction F2 consisted of phellogen (3–4 meristematic cell layers) and phelloderm, a soft green tissue with parenchymatic cells rich in chlorophyll. The nonconductive phloem (F3) contained phloem fibers, sclereids, and crushed sieve element cells. Radial sections showed that about 88% of the bark width was composed of nonconductive (F3) and conductive phloem (F4), produced by the vascular cambium (F5), whereas the remaining 12%, the periderm, originates from the phellogen (F2). Analogous tissues and cell types were found in the periderm of *Alnus glutinosa*, (Fig. 1b, iii). In alder, the fractions F2 and F3 were rich in living cells whereas the phellem (F1) consisted mostly of dead sclereids merged with the older rhytidome.

The lignin composition in the same bark and xylem fractions has been analyzed earlier (Fagerstedt *et al.*, 2015), showing that bark tissues have a low S/G ratio, increasing towards the xylem. We continued by comparing the relative amounts of noncellulosic sugars. Interestingly, the bark fraction F2 and F3 profiles resembled mature xylem fractions F6 to F8, all being rich in xylose (22–26% dry weight) (Fig. 1c). By contrast, the developing phloem (F4) and vascular cambium (F5) contained less xylose (8–12% dry weight), whereas the amounts of glucose, galacturonic acid, and arabinose were significantly elevated compared to other fractions. Additionally, galactose was abundant in the vascular cambium (F5). The sugar profiles detected in F4 and F5 are rich in xylose and galacturonic acid, the constituents of



**Fig. 1** Comparative dissection of stem cryosections. (a) The stem of *Betula pendula* was dissected into eight anatomically distinct fractions according to tissue type. (b) Magnification of bark radial sections. (bi, bii) Anatomy of *Betula pendula* bark tissues and periderm fractions (F1–F3) including the first layer of phellem (L1). (biii) Radial section from *Alnus glutinosa* periderm. (c) Noncellulosic sugar content across eight stem fractions of *B. pendula*. Ara, arabinose; Rha, rhamnose; Xyl, xylose; GlcA, glucuronic acid; meGlcA, 4-O-methyl glucuronic acid; GalA, galacturonic acid; Man, mannose; Gal, galactose; Glc, glucose. Error bars indicate the SD from three biological replicates.

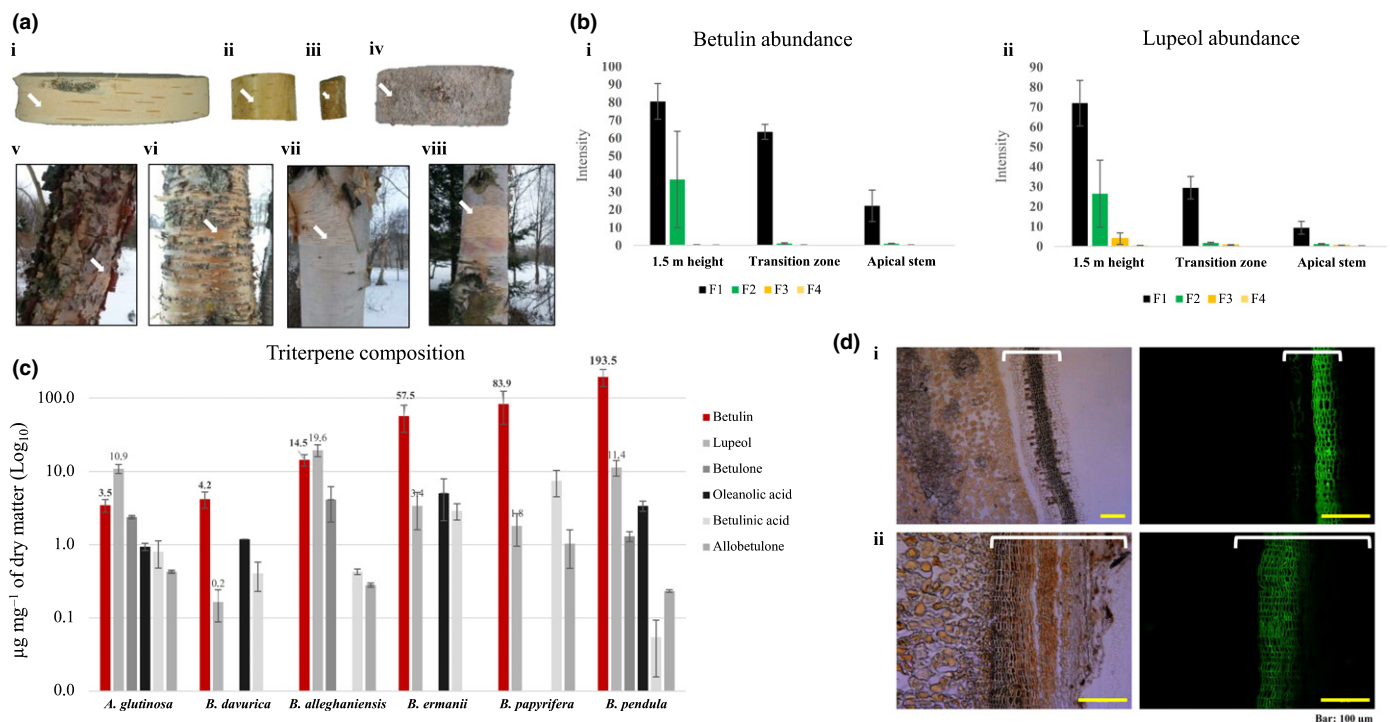
xyloglucan and pectin, typical components of primary cell walls (McNeil *et al.*, 1984) abundant in meristematic and parenchymatic tissues. The chemical profile of phellem (F1) was exceptional with less noncellulosic sugars than any other fraction; therefore, we further studied its chemical composition.

### Triterpene and suberin composition of periderm across species, developmental stages and tissues

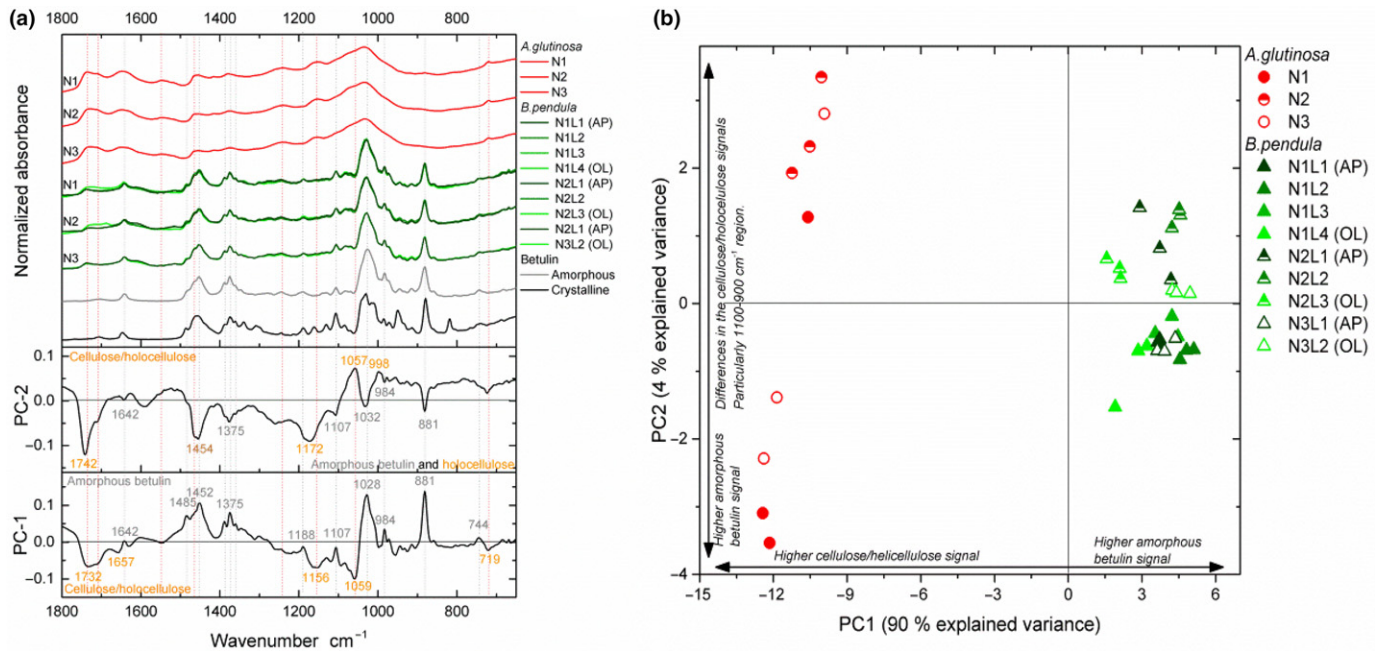
The characteristic white color of many birch species has been associated with triterpenes, with betulin being the most abundant (Krasutsky, 2006; Yin *et al.*, 2012). We analyzed the relative intensities of betulin and its precursor, lupeol, in three developmental stages of silver birch stems (Fig. 2a, i–iii): mature stem at 1.5 m height (white color), the transition zone (light-brown color), and young stems (brown color). We found increasing betulin and lupeol concentrations from young brown stems towards white mature samples. We studied their tissue-specific localization in four bark fractions (F1–F4), observing high concentrations in F1 and F2 and low traces in F3 and F4 (Fig. 2b, i, ii; Table S1). Since the metabolites are predominantly accumulated in the phellem (F1), we compared the concentration of six triterpenes among different Betulaceae species (Fig. 2c; Table S1). To capture the variation in color and texture we

examined *A. glutinosa* (rough grey phellem), *B. davurica* (red), *B. alleghaniensis* (yellow), and three white phellem species: *B. ermanii*, *B. papyrifera* and *B. pendula* (Fig. 2a, iv–vii). The highest concentration of betulin was found in *B. pendula* and the lowest in *A. glutinosa*, with the phellem of (brown-colored) apical stems in *B. pendula* containing as much betulin as other white birches. Interestingly, the concentration of the lupeol precursor was very similar between *B. pendula* and *A. glutinosa*. These results suggest that high betulin concentration has been under selective pressure in silver birch and is strongly associated with the white color. Despite the close phylogenetic relationship between *A. glutinosa* and *Betula* species, a higher accumulation of betulin and its derivatives was observed in the birches, suggesting that the betulin biosynthesis pathway is more active in birches. In addition to betulin, we studied the suberin localization using a staining method. Birch and alder shared a phellem-specific localization (Fig. 2d), similar to other angiosperm tree species *Quercus* spp. and *Populus* spp. (Soler *et al.*, 2007; Rains *et al.*, 2017).

Anatomically, birch phellem is a multilayered tissue in which the innermost layer (L1) is in direct contact with the phellogen, the meristem from which it develops. To further characterize the composition and structure across phellem layers, we analyzed the infrared absorbance spectra of these tissues in *B. pendula* and *A. glutinosa* at 1.5 m height. In all *B. pendula* phellem layers the



**Fig. 2** Inter- and intra-species chemical analysis of bark tissues in the Betulaceae family. (a) Examples of bark color and texture in six *Betula* and *Alnus* species. (ai) *Betula pendula* mature stem at 1.5 m height. (a(ii)) *Betula pendula* transition zone stem. (a(iii)) *Betula pendula* young apical stem. (a(iv)) *Alnus glutinosa* mature stem. (av) *Betula davurica* mature stem. (avi) *Betula alleghaniensis* mature stem. (avii) *Betula ermanii* mature stem. (aviii) *Betula papyrifera* mature stem (white arrows indicate the phellem tissue). (b) *Betula pendula* betulin (bi) and lupeol (bii) intensities measured by desorption atmospheric pressure photoionization–mass spectrometry (DAPPI–MS) showing the tissue-specific metabolite localization in bark cryosections at different developmental stages. (c) Quantitative analysis by gas chromatography–mass spectrometry (GC–MS) of triterpenes in the phellem (F1) of six species. In (b, c), bar plot illustrates the means and the SEs calculated from three biological replicates in each fraction, developmental stage and species. (d) Suberin localization by fluorol yellow O88 staining in *B. pendula* (di) and *A. glutinosa* (dii), light (left) and ultraviolet (UV) light (right) microscopy images (white brackets, F1).



**Fig. 3** Spectral analysis of phellem layers comparing *Betula pendula* and *Alnus glutinosa*. (a) The normalized absorbance spectra from attenuated total reflection infrared spectroscopy (ATR-IR; upper panel) and the loadings from principal component analysis (PCA; lower panels) illustrate that the main difference between *B. pendula* and *A. glutinosa* is associated with relative betulin (Cinta Pinzaru *et al.*, 2002; Falamas *et al.*, 2011) content to typical cellulose (Djikanovic *et al.*, 2016) and holocellulose (Schwanninger *et al.*, 2004) signals. (b) A plot of PCA scores displays distinct separation between *B. pendula* (green triangles) and *A. glutinosa* (red circles) bark samples. Samples are identified by: N, biological replicate number; L, layer number. AP, the layer adjacent to the phellogen (latest developed); OL, the outermost layer.

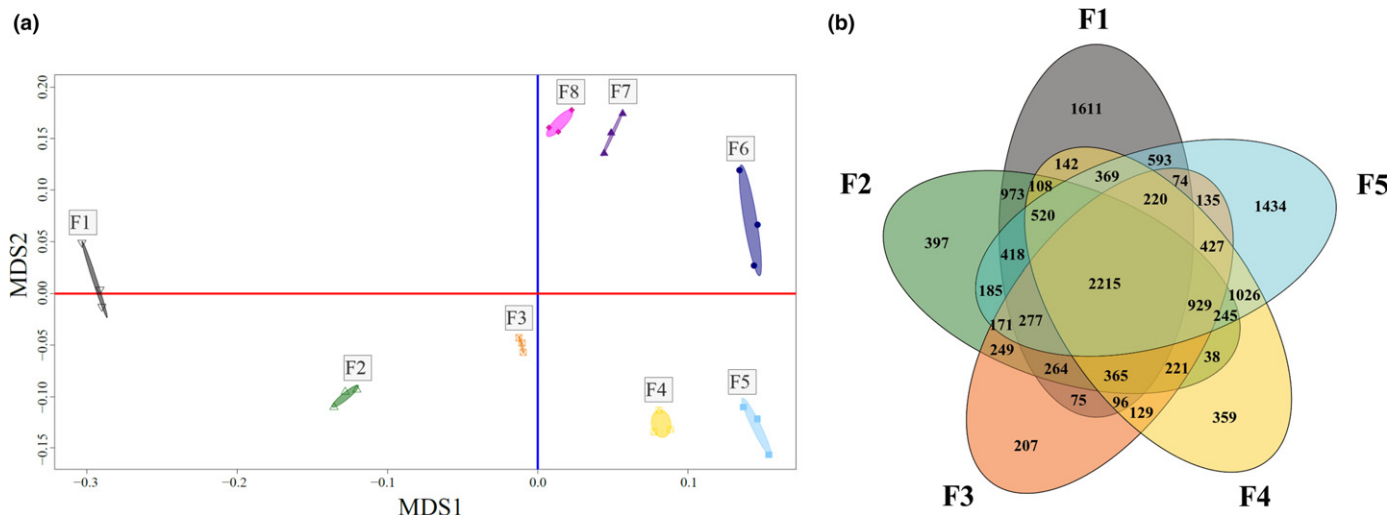
spectra were dominated by signal corresponding to betulin in an amorphous state (Fig. 3a; Cinta Pinzaru *et al.*, 2002; Falamas *et al.*, 2011). By contrast, the spectra of *A. glutinosa* were devoid of betulin and were instead dominated by cellulose and holocellulose features (Schwanninger *et al.*, 2004; Djikanovic *et al.*, 2016). These differences are captured in the loadings (Fig. 3a) and scores (Fig. 3b) from the PCA, with the first principal component (PC1) explaining 90% of the total spectral variation and clearly differentiating *B. pendula* and *A. glutinosa*. The PCA plot also demonstrates that there was no systematic variation between the biological replicates, nor were there differences between the phellem layers for *B. pendula*. The birch stem analyzed at different heights shows variation in the spectra according to the chemical composition of the phellem across different developmental stages (Fig. S2). Altogether, these analyses highlight that betulin abundance largely contributes to phellem characteristics in *Betula*, and that chemical composition showed a clear distinction of the periderm.

### Stem-wide gene expression profiling in *Betula*

We next carried out transcriptional profiling with RNA-Seq across the eight stem tissues to link the chemical characteristics with gene expression. In total, 57.1% (16790/29439) of all transcripts were expressed in the stem (Table S2). A multidimensional scaling (MDS) plot of the data showed clustering according to tissue type (Fig. 4a). Interestingly, the phellem (F1) stood out as the most transcriptionally differentiated tissue. MDS1 and MDS2 together separated developing and mature

wood tissues (F6–F8) from bark fractions (F1–F4) and vascular cambium (F5). The diversity of transcriptional profiles across bark fractions is highlighted in the MDS1 axis, where tissues are ordered according to their position in the stem and following their developmental progression: on one side F1–F2, which develop from the phellogen (F2), and on the other side F4–F5, which originate from the vascular cambium (F5), with F3 in the middle.

To compare differentially expressed genes (DEGs) between bark fractions and identify tissue-specific genes, we calculated fold changes against F8, the previous year's annual ring (Table S2). As indicated by the MDS plot, F1 had the largest number of tissue-specific DEGs (8320 genes; Fig. 4b). The pairs F1–F2 (periderm), followed by F4–F5 (phloem and vascular cambium) had considerable numbers of overlapping DEGs (973 and 1026 genes, respectively) (Fig. 4b). Gene ontology (GO) and biochemical pathway enrichment analysis was then carried out to summarize the tissue-specific transcriptional profiles (Tables S3, S4). F2 was enriched for GO terms associated with photosynthesis and suberin biosynthesis, whereas F3 (the largest bark fraction) was enriched for diverse mechanisms of secondary cell wall biogenesis and modification, as well as response to physical stimulus. One interesting example of F3 peaking genes is PIN3 (Bpev01.c1162.g0001), which is involved in stem gravitropism (Gerttula *et al.*, 2015). Fraction F4 was enriched for GO terms related to phloem development and function, including key phloem identity genes such as APL (Bpev01.c0189.g0073) (Bonke *et al.*, 2003). The previously reported difference in lignin compositions in F4 vs F6 (phloem vs xylem) was reflected in



**Fig. 4** Tissue-specific transcriptome analysis across the stem of *Betula pendula*. (a) Multidimensional scaling (MDS) plot of expression counts of bark tissues from *B. pendula*. The three biological replicates from each fraction are grouped with an ellipse (95% confidence interval), and colored according to the fraction. (b) Venn diagram of differentially expressed genes in *B. pendula* bark cryosections, including both meristems (F2 and F5).

comparative transcriptomics, revealing different laccases and class III peroxidases expressed in each tissue (Table S5).

### Genome evolution has shaped birch transcriptome profile

Considering the role of bark as the main stem protectant against environmental factors, we expected to see genes evolving through SSDs to contribute significantly to gene expression in bark tissues. Surprisingly, the differential expression profiles across all stem fractions were significantly enriched for genes originating from the *gamma* paleohexaploid WGM event (Jaillon *et al.*, 2007), such as zinc finger, bHLH and WRKY transcription factors many of which are differentially expressed in the periderm (F1 and F2), but also in the remaining fractions. Conversely, tandemly duplicated genes showed depletion (Table S6). This suggests that gene expression profiles of the differentiated tissue types and possibly the tissue types themselves have evolved from gene regulatory functional diversification following the WGM events, given that WGM-duplicated genes are enriched for transcriptional functions (for example Salojärvi *et al.*, 2017).

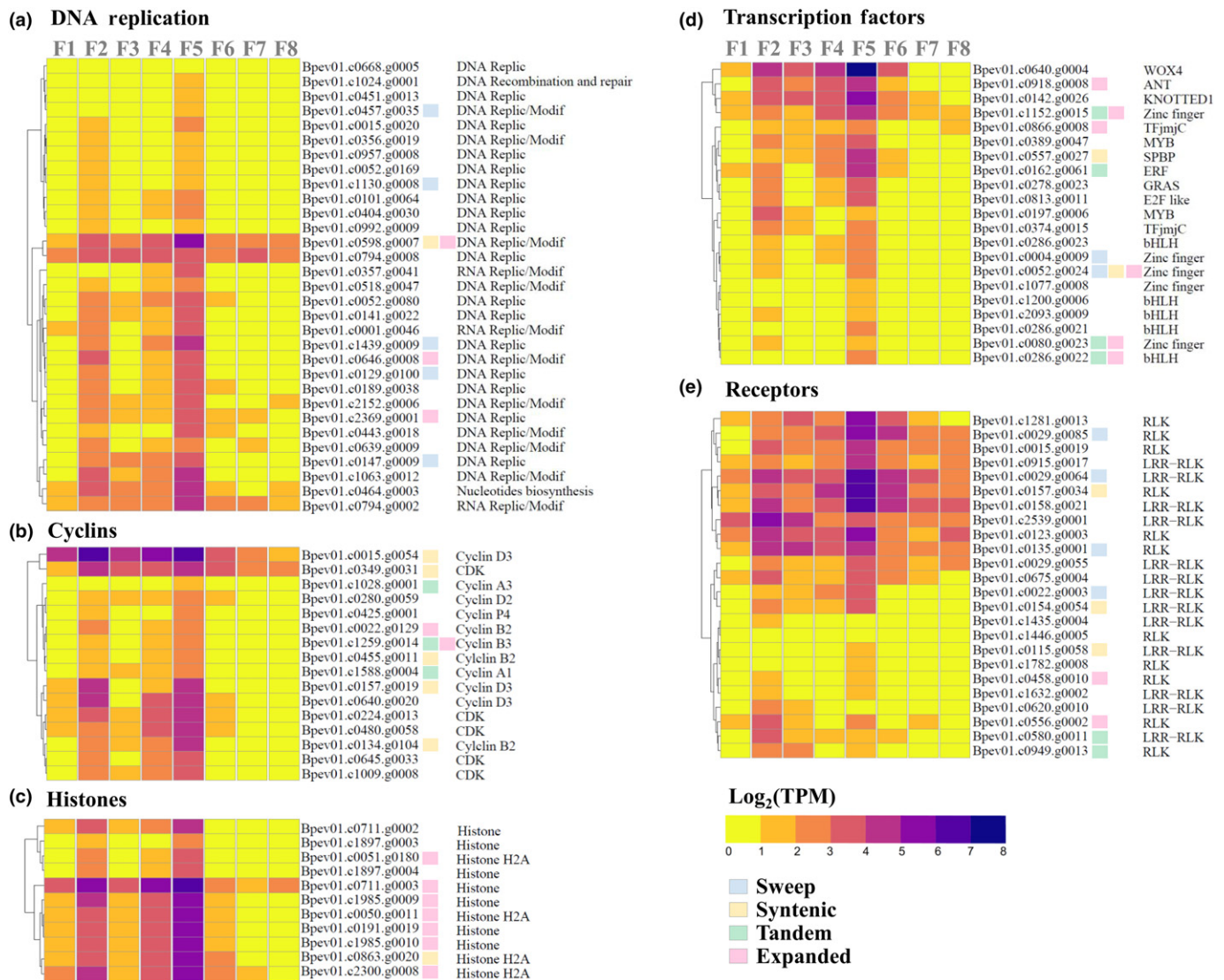
We further carried out enrichment analysis of the previously detected set of 913 genes under putative selective sweeps in silver birch (Salojärvi *et al.*, 2017) to explore whether these genes have a role in these tissues. All stem fractions were enriched for this set (Table S6), suggesting that local adaptation and thereby environmental conditions have a strong influence on tissue differentiation and thus wood composition in birch.

### F2 and F5: Bark stem cell niches-potential meristematic regulators of phellogen and vascular cambium

Gene expression profiling at a high resolution makes it possible for the first time to compare the transcriptomes of the two lateral meristems, phellogen and vascular cambium, to the transcriptomes in the rest of the stem. In order to identify the genes involved in meristematic activity of both F2 and F5, these

fractions were pooled and compared against the rest of the pooled fractions (F1, F3, F4, F6, F7 and F8), resulting in 526 significant DEGs (Fig. S3; Table S7). Many of these genes peak in both fractions, and this subset includes central regulators of cell proliferation (Fig. 5). These genes were enriched for GO categories related to meristematic activity, DNA replication and cell division (Figs 5, S4), but also photosynthesis (Table S7). The photosynthesis activity was significantly biased towards F2, likely due to the presence of phellogen cells (Table S7). F5 on the other hand expressed slightly more genes related to plant-type cell wall and regulation of meristem growth, although the bias was not significant after false discovery rate (FDR) correction. We further curated the list of 526 genes to reveal genes that predominantly peaked in F2 or F5, and additionally compared the fractions individually against the rest of the stem to identify candidate cambium- or phellogen-specific genes (Fig. S3).

Altogether, 17 transcription factors (TFs) showed an F2-specific peak in expression. One example is the ortholog of *Arabidopsis* KANADI1, a known regulator of organ polarity (Kerstetter *et al.*, 2001). Out of 32 TFs with significantly elevated transcript levels, 22 peaked in both fractions, for example birch orthologs of WUSCHEL RELATED HOMEODOMAIN 4 (WOX4) and AINTEGUMENTA (ANT). In *Arabidopsis*, WOX4 controls cell divisions in the vascular cambium (Hirakawa *et al.*, 2010) and ANT is required for lateral organ formation (Elliott *et al.*, 1996) and cell proliferation in the vascular cambium in a synergistic genetic interaction with CYCD3.1 (Randall *et al.*, 2015). Similarly in birch, the expression patterns of putative CYCD3.1 and ANT orthologs showed significant correlation (Table S8). We also identified an additional 15 CYC and CDK genes with high expression levels in F2 and F5. Interestingly, only three cyclins were specific to F5, and one type-D to F2, indicating that only a few cyclins contribute to the meristem-specific regulation of the cell cycle. Another large group of transcripts that reflects the mitotic activity in both meristems are enzymes and histones associated with the mechanisms of DNA replication. Finally, 36



**Fig. 5** Normalized abundances of common meristematic regulators of phellogen and vascular cambium in *Betula pendula*. The heat maps show  $\log_2$  of transcript per million (TPM) values. The color is proportional to the transcript abundance, and scale is displayed in the color key. The genes are highlighted according to the presence of nearby selective sweep patterns and the duplication origin of the genes; whole-genome duplication (syntenic), tandemly duplicated gene (tandem), or whether the gene family has experienced a birch-specific expansion. Columns represent bark cryosections (F1–F8) and rows the homologous genes classified to five categories: (a) DNA replication related proteins; (b) cyclins; (c) histones; (d) transcription factors; and (e) receptor-like kinases (RLK). Expression values are given in Supporting Information Table S2.

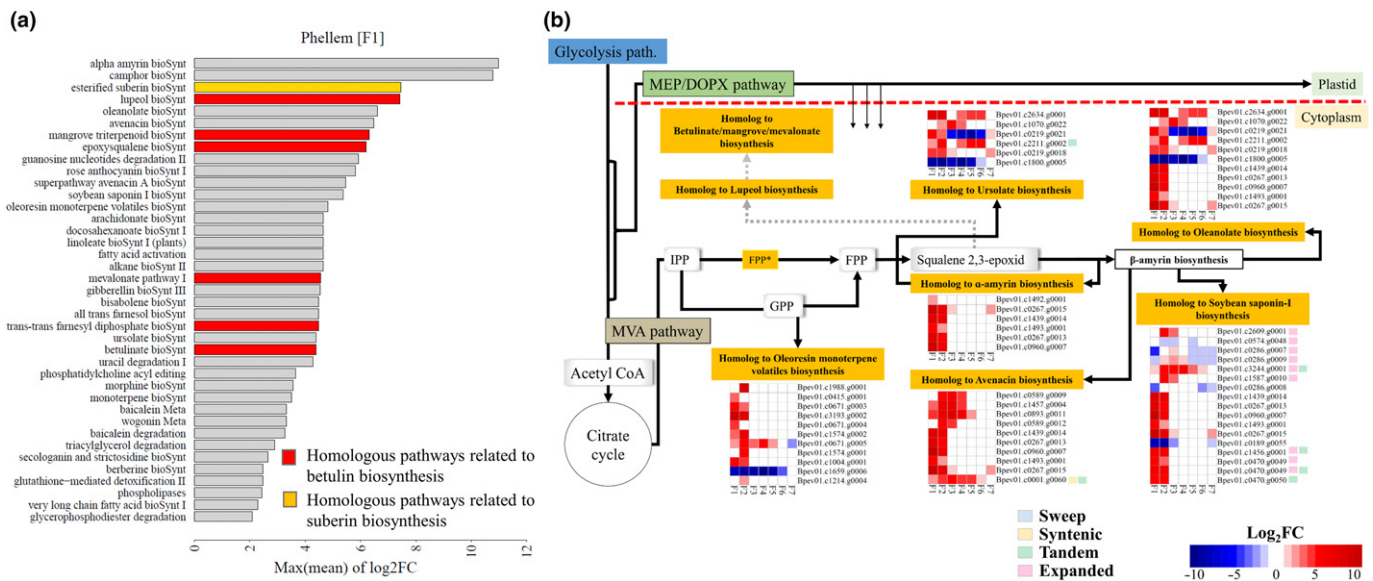
putative receptor-like kinases (RLKs) potentially involved in signaling were identified in F2 and/or F5 (Table S7). Taken together, the gene expression profiles of the phellogen and vascular cambium demonstrate gene groups with largely overlapping components, but also meristem-specific elements, opening new questions on the evolution and development of lateral meristems in trees.

### F1: Transcriptional profiling of phellem reveals diversification of triterpenoid biosynthetic pathways in *Betula*

The birch phellem (F1) has a unique metabolic and transcriptional fingerprint, and this was reflected by the largest number of fraction-specific DEGs. We found a significant enrichment for

44 PlantCyc metabolic pathways related to secondary metabolism, such as sesquiterpenes, and mono- and triterpenoids (Fig. 6a; Table S4). The majority of these compounds are known phytotoxins with roles in plant defence (Chen *et al.*, 2016). Many of the enriched pathways share common reactions or joint intermediate compounds (Fig. 6b), making it difficult to identify the correct pathway. Therefore curation taking into account the biochemical composition of birch bark was required. For example, both ursolate and  $\alpha$ -amyrin biosynthesis pathways were significantly enriched in F1. Ursolate is the product of oxidation reactions at the C-28 of  $\alpha$ -amyrin, and birch bark is known to contain ursolic acid (Mishra *et al.*, 2016; Godia *et al.*, 2018). Similarly, through further curation, we identified enrichment of the saponin, oleanolic acid and oleoresin pathways.





**Fig. 6** Schematic representation of phellem (F1) enriched pathways in *Betula pendula*. (a) Homologous metabolic pathways enriched in phellem of *B. pendula* (reference database: PlantCyc v10.0). Y-axis lists the significantly enriched metabolic pathways, x-axis shows the max(mean)  $\log_2$  of fold change ( $\log_2FC$ ) summary statistic used for assessing the enrichment of the pathway. Fold changes were calculated from phellem (F1) vs the old xylem (F8) comparison. The pathways that contribute to betulin or suberin biosynthesis are highlighted in red and orange, respectively. (b) The enriched homologous pathways of secondary metabolite biosynthesis in the periderm (F1 and F2) of *B. pendula*. The color intensity corresponds to the  $\log_2$  fold change, the differential expression value of the homologs in each fraction (F1–F7) vs  $F_8$  comparison, and the color key illustrates the scale. The genes are highlighted according to the presence of nearby selective sweep patterns, and their duplication origins; either whole-genome duplication (syntenic), tandem duplication (tandem), or whether the gene family has experienced a birch-specific expansion. An asterisk indicates an enzyme. IPP, isopentenyl diphosphate; GPP, geranylgeranyl diphosphate; FPP, farnesyl pyrophosphate (synthetase). Mevalonate pathway (MVA) is located in the cytoplasm, and 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway (MEP/DOXP) in plastid. Both pathways are exchanging intermediate compounds (shown by arrows). Metacyc IDs: oleoresin (PWY-5423),  $\beta$ -amyrin (PWY-5377), avenacin (PWY-7476), saponin (PWY-5203) and oleanolate (PWY-7069), ursolate (PWY-7068), lupeol (PWY-112), betulinic acid (PWY-7067) biosynthesis.

Saponins accumulate in epidermal cells and saponin has been identified in birch leaves (Rickling & Glombitza, 1993; Tava & Odoardi, 1996; Deepak *et al.*, 2018), whereas oleanolic acid was quantified in our chemical characterization (Table S1). Finally, the constituents of oleoresin have been identified in *B. pendula* buds (Demirci *et al.*, 2004; Maja *et al.*, 2015). These metabolites have important roles in plant defence. Oleanolic acid has antimicrobial effects (Fontanay *et al.*, 2008), whereas oleoresin is known to accumulate at the site of injury to seal off the open wounds in plants and to act as an insecticide (Trapp & Croteau, 2001). Altogether, the set of enriched metabolic pathways transcriptionally active in F1 emphasizes the importance of phellem in plant defence against biotic and abiotic stresses.

## F2 vs F3: Towards understanding the transcriptomics landscape that contribute to differences in bark characteristics between *Betula* and *Alnus*

By contrast with birch, alders develop a phellem layer that merges with a dark-grey and rough rhytidome. In order to understand the gene regulation underlying this difference we compared transcriptional profiles in anatomically comparable fractions. While birch phellem (F1) contains layers of living tissue and dead cells, F1 in alder is mainly composed of dead phellem (Fig. 1b, III), which yields only small traces of highly degraded RNA. This does

not exclude the presence of an active phellem in alder, but given its anatomy and structure, the phellem seems to consist of only a few cell files merged with older rhytidome, which in our cryosectioning approach was indistinguishable from the phellogen (F2). As in birch the main periderm pathways were already upregulated in F2 compared to F3, we dissected the corresponding tissues from alder to identify transcriptional differences. From the pre-processed RNA-Seq reads, 41.65% mapped to *A. glutinosa* gene models (Salojärvi *et al.*, 2017; Method S3), and altogether 69% (29 436/42 653) of all gene models were expressed (Table S9).

To explore the genome evolution underpinning the phenotypic differences between the two species, we clustered the proteomes of *Arabidopsis thaliana*, *B. pendula*, *A. glutinosa*, *Populus trichocarpa* and *Vitis vinifera* into computationally derived gene families, orthogroups (OGs), using Orthofinder. The analyses yielded 15154 orthogroups (Table S10), with 3863 OGs containing at least one significantly differentially expressed gene ( $FDR \leq 0.05$ ) in F2. We identified birch-specific expansions by comparing the gene counts to alder and grapevine, resulting in 935 orthogroups with birch-specific expansions, comprising 4042 genes. Altogether 29 GO categories were enriched in this set, related to biotic stresses, nutrient transport and plant defences but also to wood development (Table S11). Tandemly duplicated genes were enriched among the birch-specific expansions ( $P = 4.6e-29$ ; Fisher's test), but depleted in transcript profiles. By

contrast, birch-specific expansions were enriched in the expression profiles across all tissues (Table S6), suggesting an important role for localized, and possibly coregulated, gene duplicates.

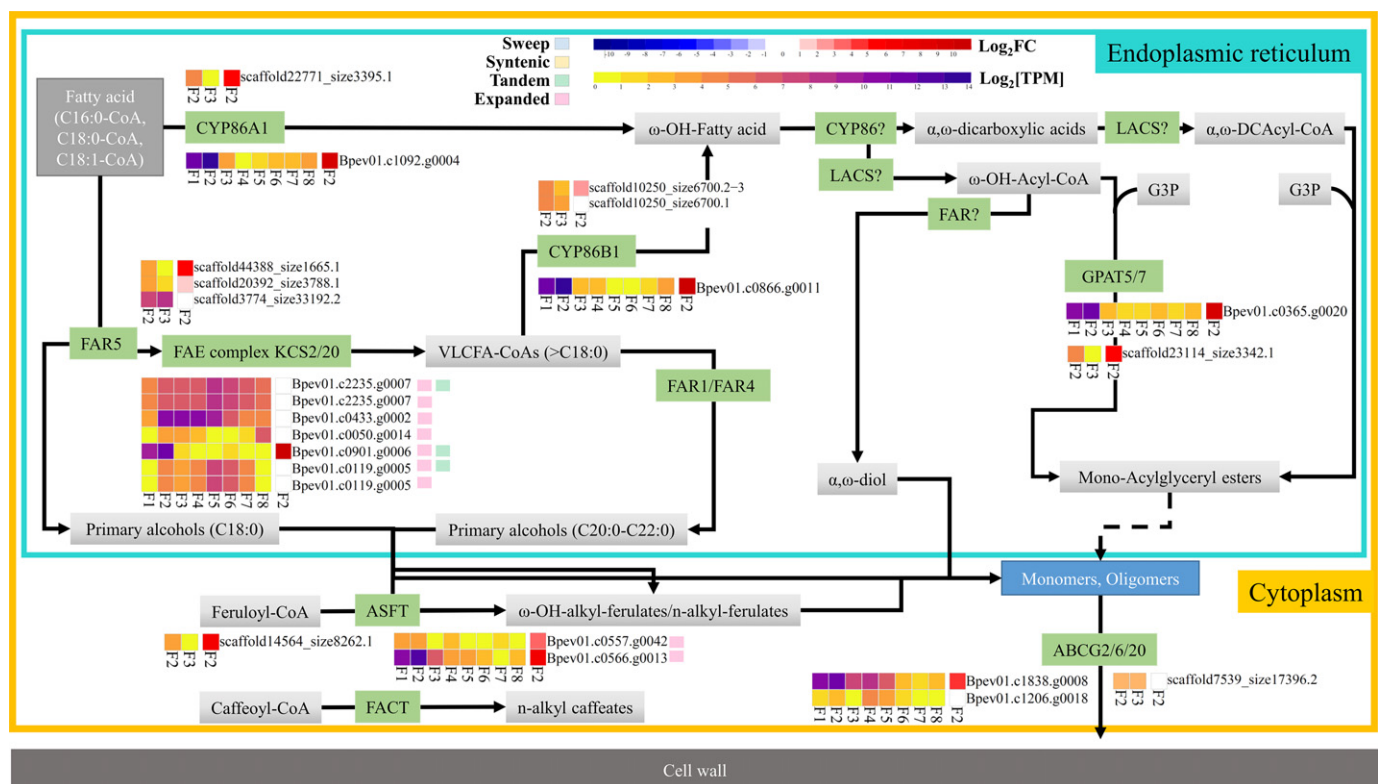
### Suberin biosynthetic genes share similar bark expression pattern in *Betula* and *Alnus*

Suberin is a glycerolipid-phenolic heteropolymer constituting a hydrophobic barrier in various tissues. We identified the putative birch and alder orthologs involved in suberin biosynthesis (Vishwanath *et al.*, 2015) and studied their expression patterns across all tissue samples (Fig. 7). In birch, the suberization pathway was predominantly active in the periderm tissues F1 and F2. However, while most of the genes were expressed in both tissues, suberin staining was specific to F1, indicating that suberin is deposited in the differentiated phellem layers. In alder, the expression levels were significantly correlated (Spearman  $\rho = -0.94$ ,  $P = 0.005$ ) with the putative birch orthologs (Fig. S5a; Table S8), albeit with a different sign, suggesting

decreasing activity in alder, and suberin staining was similar to birch in terms of localization (Fig. 2d). This periderm specificity has been shown in poplars and oaks (Rains *et al.*, 2017) as well, suggesting conservation between several tree species. Birch-specific gene expansions ( $P = 0.009$ ; Fisher's test) were enriched in the suberin pathway, more specifically the fatty acid elongase-like genes (Fig. 7). Most of these genes were not highly expressed in periderm tissues, suggesting possible functional specializations in other tissues or for other substrates.

### Mevalonate/betulin biosynthetic genes are highly expressed in *Betula* periderm

Isopentenyl pyrophosphate and dimethylallyl pyrophosphate are produced by the cytosolic mevalonate (MVA; PlantCyc: mangrove triterpenoid pathway) and the plastid nonmevalonate MEX/DOXP pathways. They are a substrate for an immense diversity of secondary metabolites such as sterols, carotenoids, sesquiterpenes and triterpenes. Betulin is produced by the MVA

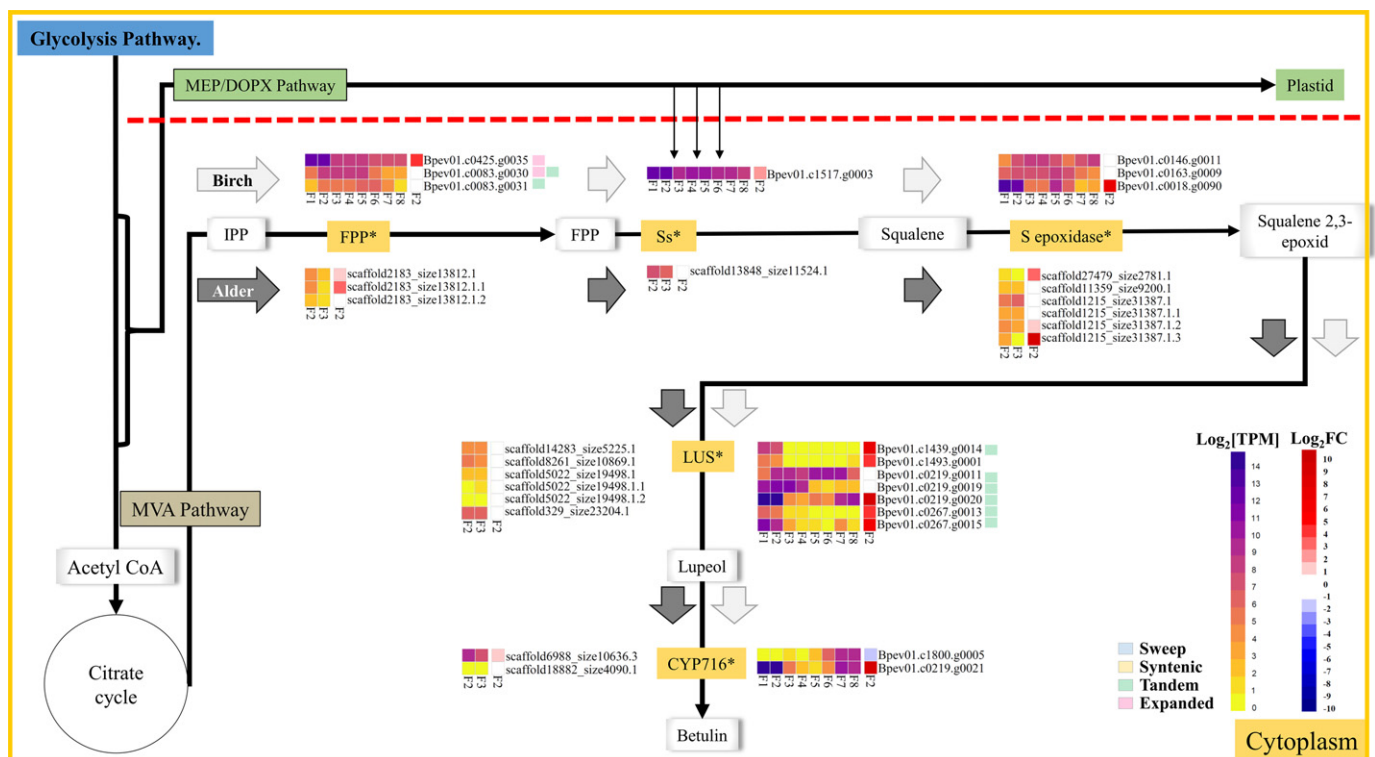


**Fig. 7** Expression pattern of putative suberin biosynthesis genes in fractions F<sub>2</sub> and F<sub>3</sub> of *Betula pendula* and *Alnus glutinosa*. Suberin pathway genes are highly expressed in the periderm of both species. Suberin pathway orthologs were identified by first assessing homology to known genes from *Arabidopsis* (Vishwanath *et al.*, 2015), and then estimating a phylogenetic tree for the homologs. The gene IDs 'Bpev01' identify *B. pendula* gene models, and the gene IDs with 'scaffold' identify *A. glutinosa* gene models. For each metabolic step, the gene expression levels from both species are shown side by side. The color keys show the scale of log<sub>2</sub> of fold change (log<sub>2</sub>FC) and log<sub>2</sub> of transcript per million (log<sub>2</sub>TPM) values of the homologous genes. The blue-red palette illustrates the differential expression value of the genes in F<sub>2</sub> vs F<sub>3</sub> comparison, and the yellow-blue palette the normalized abundances of the genes in each fraction (F<sub>1</sub>–F<sub>8</sub>). The genes are highlighted according to the presence of nearby selective sweep patterns and the duplication origin of the genes; either whole-genome duplication (syntenic), tandem duplication (tandem), or whether the gene family has experienced a birch-specific expansion. FAE, fatty acid elongation; FAR, fatty acyl reductases; CYP, cytochrome P450; ω-OHs, ω-hydroxy fatty acids; DCA, α,ω-dicarboxylic acids; G3P, glycerol-3-phosphate; GPAT, glycerol 3-phosphate acyltransferases; LACS, long-chain acyl-CoA synthetases; ABC, ATP-binding-cassette; ASFT, aliphatic suberin feruloyl transferase; FACT, fatty alcohol:caffeoyl-CoA caffeoyl transferase. The expression values of F<sub>2</sub> vs F<sub>3</sub> are presented in Supporting Information Table S2.

pathway; it is synthesized from 2,3-oxidosqualene through conversion to lupeol by lupeol synthase, and then to betulinic acid by lupeol 28-monooxygenase. Betulin and its derivatives have been studied intensively for their potential pharmaceutical and industrial applications (Sun *et al.*, 1998; Kashiwada *et al.*, 2001; Yamashita *et al.*, 2002; Tubek *et al.*, 2013; Haavikko *et al.*, 2014; Harma *et al.*, 2015; Laavola *et al.*, 2016).

An orthology-based approach identified homologs of the five key enzymes in squalene and betulinic acid biosynthesis in *B. pendula* and *A. glutinosa*. The last step in betulin biosynthesis, conversion of lupeol to betulinic acid, is catalyzed by lupeol 28-monooxygenase, an enzyme belonging to the multifunctional CYP450 family. Phylogenetic analysis including experimentally verified enzymes from *Medicago truncatula* and *Vitis vinifera* (Fukushima *et al.*, 2011) identified Bpev01.c0219.g0021 and EVM.TU.scaffold6988\_size10636.3 as the likely orthologs in birch and alder, respectively (Figs S6a, 8). The birch ortholog demonstrated high expression levels in periderm tissues, whereas the expression level of the alder ortholog was only slightly elevated.

Tandemly duplicated genes were significantly enriched ( $P=0.006$ ; Fisher test) among birch homologs for betulin biosynthesis, with duplications in farnesyl diphosphate synthases (FPP) and lupeol synthases (LUS). There were altogether nine LUS homologs in the *B. pendula* genome. The two most likely LUSs (Fig. S6b) appeared in tandem within 28 kbp of the lupeol monooxygenase. Microsyntenic comparison with alder, birch, *M. truncatula* and *V. vinifera* (grape) showed a similar orientation in grape, except that in *Vitis* the untranslated regions (UTRs) were separated by 5 bp, and the LUS duplication was birch-specific (Fig. S7). Comparison against *Amborella trichopoda* (Amborella Genome Project, 2013) suggests that the common ancestor contained a tandem of two CYP450 (Fig. S8). Further comparison with *Coffea canephora* identified an arrangement similar to grape, suggesting that the genome rearrangements before the *gamma* event placed LUS in their proximity, and in birch the LUSs experienced a tandem duplication. Similar tandem expansions have occurred in cycloartenol cyclases, as well as for other CYP450 with possibly Betulaceae-specific functions (Fig. S6b). In addition to ubiquitous AT-rich motifs, analysis of the LUS



**Fig. 8** Expression pattern of putative genes of mevalonate pathway (betulin biosynthesis) in fractions F2 and F3 of *Betula pendula* and *Alnus glutinosa*. Betulin biosynthesis is transcriptionally more active in the periderm of *B. pendula* compared to *A. glutinosa*, as observed in F2 vs F3 comparison. The gene IDs 'Bpev01' identify *B. pendula* gene models, and the gene IDs with 'evm.model' identify *A. glutinosa* gene models. For each metabolic step, the gene expression levels from both species are shown side by side. The color keys show the scale of  $\log_2$  of fold change ( $\log_2$ FC) and  $\log_2$  of transcript per million ( $\log_2$ TPM) values of the homologous genes. The blue-red palette illustrates the differential expression value of the genes in F2 vs F3 comparison, and the yellow-blue palette the normalized abundances of the genes in each fraction (F1–F8). The genes are highlighted according to the presence of nearby selective sweep patterns and the duplication origin of the genes; either whole-genome duplication (syntenic), tandem duplication (tandem), or whether the gene family has experienced a birch-specific expansion. Mevalonate pathway (Kegg ID: map00900) is located in the cytoplasm and 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway (MEP/DOXP) in plastid, both pathways are exchanging intermediate compounds (shown by arrows). An asterisk indicates an enzyme. IPP, isopentenyl diphosphate; GPP, geranylgeranyl diphosphate; FPP, farnesyl pyrophosphate (synthetase); Ss, squalene synthase; S epoxidase, squalene epoxidase; LUS, lupeol synthase; CYP716, cytochrome P716. The expression values of F2 vs F3 are presented in Supporting Information Table S2.

and CYP450 promoters identified several MYB, and WRKY as well as circadian CCA1 and LHY1 motifs (Table S12). Corresponding regions in grape and alder contained only a single WRKY motif in LUS, and overall considerably less MYB/WRKY motifs in CYP450 and no circadian motifs.

Both birch and alder demonstrated elevated expression levels in the MVA pathway in periderm. However, in birch the last two steps of betulin biosynthesis were transcriptionally considerably more active than in alder (Fig. S5b; Table S8). Additionally, five out of nine birch LUS paralogs were highly expressed in F2. The expression pattern was similar also in TPM-normalized count values (Fig. 8). The chemical quantification showed similar concentrations of lupeol, but significantly different concentrations of betulin between birch and alder. Taken together, these observations suggest that in birch, a considerably larger amount of squalene gets converted into lupeol due to the birch-specific expansion of the LUS gene family. Similar lupeol concentrations further suggest that the lupeol monooxygenase in birch efficiently converts most of the lupeol to betulinate, altogether suggesting that the conversion of squalene to lupeol is the pathway bottleneck that has been optimized in birch.

#### Identification of species specific vs common genes coding for periderm-abundant TFs in *Betula* vs *Alnus*

Taken together, our tissue-specific analyses strongly suggest that the periderm, and more precisely, the phellem is the final destination for many secondary metabolites. Furthermore, the periderm seems to partially have its own developmental regulation. In order to identify TFs potentially involved in the regulation of periderm-specific processes, we analyzed all predicted TFs from *B. pendula* (Table S13) and identified 82 TFs with periderm-specific expression patterns. Significant correlation with suberin pathway genes was observed for 37 TFs, among them seven orthologs of putative regulators of suberization in *Populus* (Rains *et al.*, 2017). Similar analysis yielded 33 TFs with significant correlation with betulin biosynthesis pathway genes, among them 9 WRKYs, 5 NACs and 5 MYB family TFs (Table S8).

The search for alder orthologs of birch periderm TFs returned 62 TFs, of which 41 were also upregulated in F2 compared to F3 (Table S7). The expression of 11 and 17 TFs were found to be species specific in birch and alder periderms, respectively. The set of birch-specific genes contained two WRKYs, both of which correlated significantly with birch betulin pathway enzymes (putative WRKY6, Bpev01.c0127.g0039; and putative WRKY23, Bpev01.c0932.g0014), whereas the set of common TFs contained 5 MYBs, all of which correlated with betulin pathway enzymes.

## Discussion

Trees have evolved long-term adaptive traits to withstand environmental factors native to their habitat, and their bark constitutes the first barrier to environmental stresses. While studies have characterized some aspects of bark, periderm or wood development, the division into bark-specific vs wood-specific processes has remained

elusive due to the lack of further tissue-specific dissection. Here we dissected birch bark tissues into five different fractions (from phellem to vascular cambium) and wood into three stages of xylem development. The chemical characterization showed that the non-conductive phloem and xylem were very similar in their composition of structural noncellulosic sugars, whereas phellem and phloem-cambium tissues were different, highlighting the difference between primary and secondary cell walls.

Within bark, the phellem (F1) had less monosaccharides, instead having high concentrations of amorphous betulin (approx. 20% of dry weight). Betulin concentration increased with the age of the stem, and spatially the highest amount was encountered in the phellem (F1), indicating the importance of betulin to the outermost stem–environment barrier. Compared to other Betulaceae species, *B. pendula* had the highest concentration of betulin, whereas the concentration was lowest in *A. glutinosa*. Both contained similar amounts of the metabolic precursor lupeol, indicating that betulin accumulation in the phellem has diversified within the Betulaceae family. Differing betulin concentration was the distinctive feature in subsequent spectral analyses as well. By contrast, suberin was equally localized in the phellem of both species.

Transcriptomic analysis of the eight *Betula pendula* stem tissues revealed the activated metabolic pathways and developmental processes. The expression patterns across all tissues were enriched for gene duplicates from the *gamma* WGM event, suggesting that tissue differentiation may have its origin in the transcriptional regulatory functional diversification following the WGM. Additionally, the expression profiles were depleted for tandem duplications.

Unexpectedly, the largest set of tissue-specific genes was observed in the phellem. This result indicates that this tissue, far from being unimportant, harbors a plethora of specific mechanisms required for development, protection and stress responses. In the transcriptomics data, many of the pathways enriched in phellem shared common reactions or joint intermediate compounds, making it difficult to identify the correct pathway. One such example was an enrichment of avenacin biosynthesis in F1. Exclusively present in oats (genus *Avena*), avenacin is produced by hydroxylation of  $\beta$ -amyrin by  $\beta$ -amyrin monooxygenase. However, in grape (*Vitis vinifera*) the enzyme with highest sequence similarity has a confirmed role in betulin production, and therefore the avenacin pathway is likely an annotation error due to high sequence similarity. Therefore a careful curation of the enriched pathways was needed by comparing against the known birch metabolites.

One well conserved periderm pathway across tree species is suberin biosynthesis. Accordingly, in birch and alder suberin deposition was observed in the phellem (F1). The pathway was already active in F2 and showed similar expression patterns in both species, suggesting common transcriptional regulation. By contrast, the betulin biosynthesis pathway showed periderm specification in birch, but not in alder. The important contribution of tandem gene expansions to secondary metabolism was highlighted here as well, since the analysis suggested that the enhanced betulin biosynthesis was at least partially due to a birch-specific tandem duplication of the LUS genes. Additionally, a genomic

organization with similar TF binding sites makes coexpression of LUS and the lupeol 28-monooxygenase highly likely. The expression data showed that the enzymes involved in the last two steps of betulin biosynthesis were more highly expressed in *B. pendula* than in *A. glutinosa*, supporting *Betula*-specific amplification of the betulin biosynthesis pathway.

The ontogeny of bark tissues has been so far characterized from an anatomical perspective (Evert & Esau, 2007). Earlier transcriptome comparison between cork and wood tissues suggested a shared set of genes for wood and bark development (Boher *et al.*, 2018). The higher resolution in our analyses showed that the transcriptional profile of developing bark tissues preferentially groups with the meristems from which they originate. This was visible from the MDS analysis, but also in the numbers of overlapping DEGs among bark tissues. Overlapping transcriptional and hormonal gradients across the stem tissues have been reported during wood formation (Immanen *et al.*, 2016), and current results indicate that analogous profiles may be active also during bark development. Our analysis identified many common regulatory components (TFs, cyclins, and RLK), but also pointed out meristem-specific regulators. Altogether 22 TFs peaked in both meristems, including ANT, KNOTTED1 and WOX4, the last also reported in oak (Boher *et al.*, 2018). Interestingly, the birch orthologs of the oak genes reported in similar context do not have meristem-specific profiles. The regulation of these genes may have diverged between the two species or, alternatively, the higher spatial resolution of our study could explain the difference. Finally, the analyses revealed 82 TFs that predominantly peaked in the periderm, including members of the MYB, NAC, WRKY, GARP and LOB families. Among them were orthologs of organ development regulators, including a MYB66 ortholog involved in epidermal cell fate determination in *Arabidopsis* (Lee & Schiefelbein, 1999), and KANADI1, involved in organ polarity (Kerstetter *et al.*, 2001). Furthermore, the set contained seven TFs that correlated significantly with suberin biosynthesis genes and which have also been proposed as regulators of periderm suberization in poplar. In the case of betulin biosynthesis, promoter analysis narrowed down the putative regulators to the MYB and WRKY families, and coexpression analysis yielded 14 candidates from these families. Altogether, the integration of transcriptional and chemical profiles across the stem provides a molecular framework to answer questions about bark function and development. Due to its unique secondary metabolite composition, bark has ample potential for biotechnological applications: salicylic acid, paclitaxel and quinine are classical examples of bark derived drugs, but modern pharmaceutical biotechnology will provide completely new opportunities for bark product development.

## Acknowledgements

This study was carried out by Genochem Consortium supported by the Academy of Finland (decision 26421), Finnish Centre of Excellence in Molecular Biology of Primary Producers, Academy of Finland CoE 2014–2019 (271832) and project (286404), the Gatsby Foundation (GAT3395/PR3), the National Science

Foundation Biotechnology and Biological Sciences Research Council grant (BB/N013158/1), University of Helsinki (79992091), and the European Research Council Advanced Investigator Grant SYMDEV (323052). The authors are grateful for the assistance of Dr Marie Barberon for the fluorol yellow 088 staining for suberin, the Kumpula Botanical Garden for information and access to birch species, and Airi Lamminmäki and Eija Rinne for excellent technical support.

## Author contributions

YH, JK, MT, JTY-K, THT, KF, JA-S, OS, JI, TJK, R-MR, CJS, KN and JS designed the research; K-JL and JA-S performed the RNA extractions. JA-S, KN and JI did the sampling for GCMS and FTIR and produced all anatomical pictures. CJS supervised and planned the FTIR spectroscopic analyses, associated multivariate analyses and figures, data interpretation and performed language checking. SJF-M carried out the FTIR data collection, analysis, interpretation of these data and created the figures. S-LC performed the carbohydrate analyses. RH did the synthesis of triterpene reference compounds. ML performed the quantitative triterpene analysis, and R-MR the DAPPI analyses. JTY-K interpreted the organic chemistry of biological pathways, planned and supervised the synthesis of triterpene reference compounds, and was the person-in-charge of the Genochem pharmacy subproject. OS, JS carried out the analysis of transcriptomic data, and OS, JA-S, PS, OBB, LR, AC, SR and JS participated in the interpretation and identification of candidate genes. SR annotated alder genome assembly. OS did functional annotation. JA-S, OS, KN, CJS and JS wrote the manuscript with input from SJM, KJL, OBB, YH and feedback from all the authors. JA-S, OS and K-JL contributed equally to this work.

## ORCID

Juan Alonso-Serra  <https://orcid.org/0000-0002-2810-4691>  
 Olga B. Blokhina  <https://orcid.org/0000-0003-0808-9663>  
 Ana Campilho  <https://orcid.org/0000-0002-3470-1943>  
 Sun-Li Chong  <https://orcid.org/0000-0003-4183-774X>  
 Kurt Fagerstedt  <https://orcid.org/0000-0002-6839-2958>  
 Sara J. Fraser-Miller  <https://orcid.org/0000-0003-1061-2441>  
 Raisa Haavikko  <https://orcid.org/0000-0002-2445-503X>  
 Ykä Helariutta  <https://orcid.org/0000-0002-7287-8459>  
 Juha Immanen  <https://orcid.org/0000-0003-1098-4843>  
 Jaakko Kangasjärvi  <https://orcid.org/0000-0002-8959-1809>  
 Tiina J. Kauppila  <https://orcid.org/0000-0002-3796-5573>  
 Mari Lehtonen  <https://orcid.org/0000-0002-3169-1030>  
 Kean-Jin Lim  <https://orcid.org/0000-0003-2147-0215>  
 Kaisa Nieminen  <https://orcid.org/0000-0001-7004-9422>  
 Laura Ragni  <https://orcid.org/0000-0002-3651-8966>  
 Sitaram Rajaraman  <https://orcid.org/0000-0001-5171-3578>  
 Riikka-Marjaana Räsänen  <https://orcid.org/0000-0003-3073-9861>  
 Pezhman Safdari  <https://orcid.org/0000-0001-9620-2106>  
 Omid Safronov  <https://orcid.org/0000-0003-2044-3258>  
 Jarkko Salojärvi  <https://orcid.org/0000-0002-4096-6278>

Clare J. Strachan  <https://orcid.org/0000-0003-3134-8918>  
Teemu H. Teeri  <https://orcid.org/0000-0002-3812-7213>  
Maija Tenkanen  <https://orcid.org/0000-0003-2883-2717>  
Jari T. Yli-Kauhaluoma  <https://orcid.org/0000-0003-0370-7653>

## References

- Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. *Science* 342: 1241089.
- Birchler JA, Bhadra U, Bhadra MP, Auger DL. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental Biology* 234: 275–288.
- Boher P, Soler M, Sanchez A, Hoede C, Noirot C, Paiva JAP, Serra O, Figueras M. 2018. A comparative transcriptomic approach to understanding the formation of cork. *Plant Molecular Biology* 96: 103–118.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bonke M, Thitamadee S, Mähönen AP, Hauser MT, Helariutta Y. 2003. APL regulates vascular tissue identity in *Arabidopsis*. *Nature* 426: 181–186.
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-Seq quantification. *Nature Biotechnology* 34: 525–527.
- Bret-Harte MS, Shaver GR, Zoerner JP, Johnstone JF, Wagner JL, Chavez AS, Gunkelman RF, Lippert SC, Laundre JA. 2001. Developmental plasticity allows *Betula nana* to dominate tundra subjected to an altered environment. *Ecology* 82: 18–32.
- Carretero-Paulet L, Fares MA. 2012. Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Molecular Biology and Evolution* 29: 3541–3551.
- Celedon JM, Yuen MMS, Chiang A, Henderson H, Reid KE, Bohlmann J. 2017. Cell-type- and tissue-specific transcriptomes of the white spruce (*Picea glauca*) bark unmask fine-scale spatial patterns of constitutive and induced conifer defense. *The Plant Journal* 92: 710–726.
- Chen J, Zheng G, Zhang Y, Aisa HA, Hao XJ. 2016. Phytotoxic terpenoids from *Ligularia cymbulifera* roots. *Frontiers in Plant Science* 7: 2033.
- Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* 30: 31–37.
- Chong SL, Derba-Maceluch M, Koutaniemi S, Gomez LD, McQueen-Mason SJ, Tenkanen M, Mellerowicz EJ. 2015. Active fungal GH115 alpha-glucuronidase produced in *Arabidopsis thaliana* affects only the UX1-reactive glucuronate decorations on native glucuronoxylans. *BMC Biotechnology* 15: 56.
- Chong SL, Koutaniemi S, Virkki L, Pynnonen H, Tuomainen P, Tenkanen M. 2013. Quantitation of 4-O-methylglucuronic acid from plant cell walls. *Carbohydrate Polymers* 91: 626–630.
- Cinta Pinzaru S, Leopold N, Kiefer W. 2002. Vibrational spectroscopy of betulinic acid HIV inhibitor and of its birch bark natural source. *Talanta* 57: 625–631.
- Deepak M, Lihavainen J, Keski-Saari S, Kontunen-Soppela S, Salojärvi J, Tenkanen A, Heimonen K, Oksanen E, Keinänen M. 2018. Genotype- and provenance-related variation in the leaf surface secondary metabolites of silver birch. *Canadian Journal of Forest Research* 48: 494–505.
- Demirci B, Paper DH, Demirci F, Can Baser KH, Franz G. 2004. Essential oil of *Betula pendula* Roth. buds. *Evidence-based Complementary and Alternative Medicine* 1: 301–303.
- Djikanovic D, Devecerski A, Steinbach G, Simonovic J, Matovic B, Garab G, Kalauzi A, Radotic K. 2016. Comparison of macromolecular interactions in the cell walls of hardwood, softwood and maize by fluorescence and FTIR spectroscopy, differential polarization laser scanning microscopy and X-ray diffraction. *Wood Science and Technology* 50: 547–566.
- Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker WQJ, Gerentes D, Perez P, Smyth DR. 1996. AINTEGUMENTA, an APETALA2-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* 8: 155–168.
- Etchells JP, Mishra LS, Kumar M, Campbell L, Turner SR. 2015. Wood formation in trees is increased by manipulating PXY-regulated cell division. *Current Biology* 25: 1050–1055.
- EUFORGEN. 2009. *Distribution map of silver birch (Betula pendula)*. [WWW document] URL <http://www.euforgen.org> [accessed 12 December 2016].
- Evert RF, Esau K. 2007. *Esau's plant anatomy: meristems, cells, and tissues of the plant body: their structure, function, and development*. Hoboken, NJ, USA: Wiley.
- Fagerstedt VK, Saranpää P, Tapanila T, Immanen J, Serra AJ, Nieminen K. 2015. Determining the composition of lignins in different tissues of silver birch. *Plants* 4: 183–195.
- Falamas A, Pinzaru SC, Dehelean CA, Peev CI, Soica C. 2011. Betulin and its natural resource as potential anticancer drug candidate seen by FT-Raman and FT-IR spectroscopy. *Journal of Raman Spectroscopy* 42: 97–107.
- Fontanay S, Grare M, Mayer J, Finance C, Duval RE. 2008. Ursolic, oleanolic and betulinic acids: antibacterial spectra and selectivity indexes. *Journal of Ethnopharmacology* 120: 272–276.
- Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* 60: 433–453.
- Fukushima EO, Seki H, Ohshima K, Ono E, Umemoto N, Mizutani M, Saito K, Muranaka T. 2011. CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant and Cell Physiology* 52: 2050–2061.
- Gerttula S, Zinkgraf M, Muday GK, Lewis DR, Ibatullin FM, Brumer H, Hart F, Mansfield SD, Filkov V, Groover A. 2015. Transcriptional and hormonal regulation of gravitropism of woody stems in *Populus*. *Plant Cell* 27: 2800–2813.
- Godia D, Paze A, Rizhikovs J, Stankus K, Virsis I, Nakurte I. 2018. Stability studies of bioactive compounds from birch outer bark ethanolic extracts. *Key Engineering Materials* 762: 152–157.
- Haapala M, Pol J, Saarela V, Arvola V, Kotiaho T, Ketola RA, Franssila S, Kauppila TJ, Kostiaainen R. 2007. Desorption atmospheric pressure photoionization. *Analytical Chemistry* 79: 7867–7872.
- Haavikko R, Nasereddin A, Sacerdoti-Sierra N, Kopelyanskiy D, Alakurtti S, Tikka M, Jaffe CL, Yli-Kauhaluoma J. 2014. Heterocycle-fused lupane triterpenoids inhibit *Leishmania donovani* amastigotes. *Medical Chemical Communications* 5: 445–451.
- Harma V, Haavikko R, Virtanen J, Ahonen I, Schukov HP, Alakurtti S, Purev E, Rischer H, Yli-Kauhaluoma J, Moreira VM *et al.* 2015. Optimization of invasion-specific effects of betulin derivatives on prostate cancer cells through lead development. *PLoS ONE* 10: e0126111.
- Hellgren JM, Olofsson K, Sundberg B. 2004. Patterns of auxin distribution during gravitational induction of reaction wood in poplar and pine. *Plant Physiology* 135: 212–220.
- Hirakawa Y, Kondo Y, Fukuda H. 2010. TDIF peptide signaling regulates vascular stem cell proliferation via the WOX4 homeobox gene in *Arabidopsis*. *Plant Cell* 22: 2618–2629.
- Idänheimo N, Gauthier A, Salojärvi J, Siligato R, Brosche M, Kollist H, Mähönen AP, Kangasjärvi J, Wrzaczek M. 2014. The *Arabidopsis thaliana* cysteine-rich receptor-like kinases CRK6 and CRK7 protect against apoplastic oxidative stress. *Biochemical and Biophysical Research Communications* 445: 457–462.
- Immanen J, Nieminen K, Smolander OP, Kojima M, Alonso-Serra J, Koskinen P, Zhang J, Elo A, Mähönen AP, Street N *et al.* 2016. Cytokinin and auxin display distinct but interconnected distribution and signaling profiles to stimulate cambial activity. *Current Biology* 26: 1990–1997.
- Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C *et al.* 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.
- Karels TJ, Boonstra R. 2003. Reducing solar heat gain during winter: the role of white bark in northern deciduous trees. *Arctic* 56: 168–174.
- Kashiwada Y, Chiyo J, Ikeshiro Y, Nagao T, Okabe H, Cosentino LM, Fowke K, Lee KH. 2001. 3,28-Di-O-(dimethylsuccinyl)-betulin isomers as anti-HIV agents. *Bioorganic & Medicinal Chemistry Letters* 11: 183–185.

- Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS. 2001. KANADI regulates organ polarity in *Arabidopsis*. *Nature* 411: 706–709.
- Krasutsky PA. 2006. Birch bark research and development. *Natural Production Reports* 23: 919–942.
- Laavola M, Haavikko R, Hämäläinen M, Leppänen T, Nieminen R, Alakurtti S, Moreira VM, Yli-Kauhahuoma J, Moilanen E. 2016. Betulin derivatives effectively suppress inflammation *in vitro* and *in vivo*. *Journal of Natural Products* 79: 274–280.
- Lee MM, Schiefelbein J. 1999. WEREWOLF, a MYB-related protein in *Arabidopsis*, is a position-dependent regulator of epidermal cell patterning. *Cell* 99: 473–483.
- Liby KT, Yore MM, Sporn MB. 2007. Triterpenoids and rexinoids as multifunctional agents for the prevention and treatment of cancer. *Nature Reviews Cancer* 7: 357–369.
- Lim KJ, Paasela T, Harju A, Venäläinen M, Paulin L, Auvinen P, Kärkkäinen K, Teeri TH. 2016. Developmental changes in scots pine transcriptome during heartwood formation. *Plant Physiology* 172: 1403–1417.
- Maja MM, Kasurinen A, Holopainen T, Kontunen-Soppela S, Oksanen E, Holopainen JK. 2015. Volatile organic compounds emitted from silver birch of different provenances across a latitudinal gradient in Finland. *Tree Physiology* 35: 975–986.
- Mantello CC, Cardoso-Silva CB, da Silva CC, de Souza LM, Scaloppi EJ, Gonçalves PD, Vicentini R, de Souza AP. 2014. De novo assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. *PLoS ONE* 9: e102665.
- Mcneil M, Darvill AG, Fry SC, Albersheim P. 1984. Structure and function of the primary-cell walls of plants. *Annual Review of Biochemistry* 53: 625–663.
- Mishra T, Arya RK, Meena S, Joshi P, Pal M, Meena B, Upreti DK, Rana TS, Datta D. 2016. Isolation, characterization and anticancer potential of cytotoxic triterpenes from *Betula utilis* bark. *PLoS ONE* 11: e0159430.
- Paine CET, Stahl C, Courtois EA, Patino S, Sarmiento C, Baraloto C. 2010. Functional explanations for variation in bark thickness in tropical rain forest trees. *Functional Ecology* 24: 1202–1210.
- Panchy N, Lehti-Shiu M, Shiu S-H. 2016. Evolution of gene duplication in plants. *Plant Physiology* 171: 2294–2316.
- Park S, Keathley DE, Han KH. 2008. Transcriptional profiles of the annual growth cycle in *Populus deltoides*. *Tree Physiology* 28: 321–329.
- Pisha E, Chai H, Lee IS, Chagwedera TE, Farnsworth NR, Cordell GA, Beecher CWW, Fong HHS, Kinghorn AD, Brown DM *et al.* 1995. Discovery of betulinic acid as a selective inhibitor of human-melanoma that functions by induction of apoptosis. *Nature Medicine* 1: 1046–1051.
- Rains MK, Gardiyehewa de Silva ND, Molina I. 2017. Reconstructing the suberin pathway in poplar by chemical and transcriptomic analysis of bark tissues. *Tree Physiology* 38: 340–361.
- Randall RS, Miyashima S, Blomster T, Zhang J, Elo A, Karlberg A, Immanen J, Nieminen K, Lee JY, Kakimoto T *et al.* 2015. AINTEGUMENTA and the D-type cyclin CYCD3;1 regulate root secondary growth and respond to cytokinins. *Biology Open* 4: 1229–1236.
- Rickling B, Glombitza KW. 1993. Saponins in the leaves of birch? Hemolytic dammarane triterpenoid esters of *Betula pendula*. *Planta Medica* 59: 76–79.
- Rosell JA, Gleason S, Mendez-Alonzo R, Chang Y, Westoby M. 2014. Bark functional ecology: evidence for tradeoffs, functional coordination, and environment producing bark diversity. *New Phytologist* 201: 486–497.
- Salojärvi J, Smolander O-P, Nieminen K, Rajaraman S, Safronov O, Safdari P, Lamminmäki A, Immanen J, Lan T, Tanskanen J *et al.* 2017. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nature Genetics* 49: 904–912.
- Schenk MF, Thienpont CN, Koopman WJM, Gilissen LJWJ, Smulders MJM. 2008. Phylogenetic relationships in *Betula* (Betulaceae) based on AFLP markers. *Tree Genetics & Genomes* 4: 911–924.
- Schwanninger M, Rodrigues JC, Pereira H, Hinterstoisser B. 2004. Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose. *Vibrational Spectroscopy* 36: 23–40.
- Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M. 2007. A genomic approach to suberin biosynthesis and cork differentiation. *Plant Physiology* 144: 419–431.
- Sun IC, Wang HK, Kashiwada Y, Shen JK, Cosentino LM, Chen CH, Yang LM, Lee KH. 1998. Anti-AIDS agents. 34. Synthesis and structure-activity relationships of betulin derivatives as anti-HIV agents. *Journal of Medicinal Chemistry* 41: 4648–4657.
- Sundell D, Street NR, Kumar M, Mellerowicz EJ, Kucukoglu M, Johnsson C, Kumar V, Mannapperuma C, Delhomme N, Nilsson O *et al.* 2017. AspWood: high-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in *Populus tremula*. *Plant Cell* 29: 1585–1604.
- Tasdighian S, Van Bel M, Li Z, Van de Peer Y, Carretero-Paulet L, Maere S. 2017. Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *Plant Cell* 29: 2766–2785.
- Tava A, Odoardi M. 1996. Saponins from *Medicago* spp.: chemical characterization and biological activity against insects. *Tissue Renin-Angiotensin Systems* 405: 97–109.
- Trapp S, Croteau R. 2001. Defensive resin biosynthesis in conifers. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 689–724.
- Tubek B, Mitula P, Niezgodna N, Kempinska K, Wietrzyk J, Wawrzenczyk C. 2013. Synthesis and cytotoxic activity of new betulin and betulinic acid esters with conjugated linoleic acid (CLA). *Natural Product Communications* 8: 435–438.
- Tuominen H, Puech L, Fink S, Sundberg B. 1997. A radial concentration gradient of indole-3-acetic acid is related to secondary xylem development in hybrid aspen. *Plant Physiology* 115: 577–585.
- Vandegheuchte MW, Bloemen J, Vergeynst LL, Steppe K. 2015. Woody tissue photosynthesis in trees: salve on the wounds of drought? *New Phytologist* 208: 998–1002.
- Vishwanath SJ, Delude C, Domergue F, Rowland O. 2015. Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. *Plant Cell Reports* 34: 573–586.
- Wittmann C, Pfanz H. 2007. Temperature dependency of bark photosynthesis in beech (*Fagus sylvatica* L.) and birch (*Betula pendula* Roth.) trees. *Journal of Experimental Botany* 58: 4293–4306.
- Wunderling A, Ripper D, Barra-Jimenez A, Mahn S, Sajak K, Targem MB, Ragni L. 2018. A molecular framework to study periderm formation in *Arabidopsis*. *New Phytologist* 219: 216–229.
- Yamashita K, Lu HW, Lu JC, Chen G, Yokoyama T, Sagara Y, Manabe M, Kodama H. 2002. Effect of three triterpenoids, lupeol, betulin, and betulinic acid on the stimulus-induced superoxide generation and tyrosyl phosphorylation of proteins in human neutrophils. *Clinica Chimica Acta* 325: 91–96.
- Yin J, Ren CL, Zhan YG, Li CX, Xiao JL, Qiu W, Li XY, Peng HM. 2012. Distribution and expression characteristics of triterpenoids and OSC genes in white birch (*Betula platyphylla* suk.). *Molecular Biology Reports* 39: 2321–2328.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1** Stem cryosectioning approach.

**Fig. S2** Phellem chemical spectrum along the stem in *B. pendula*.

**Fig. S3** Shared and meristem-specific genes.

**Fig. S4** Meristematic regulators predominantly expressed in the phellogen or vascular cambium.

**Fig. S5** Scatterplot of differentially expressed genes from suberin biosynthesis and mevalonate pathways in F2 vs F3 of *B. pendula* and *A. glutinosa*.

**Fig. S6** Gene trees of homologs of lupeol 28-monooxygenase (CYP716) and lupeol synthase (LUS) enzymes identify the putative orthologs in *B. pendula* and *A. glutinosa*.

**Fig. S7** Microsynteny plot of lupeol 28-monooxygenase (CYP716) and lupeol synthase (LUS) orthologs from *B. pendula* and *Alnus glutinosa* (two genome assembly versions), *Medicago truncatula* and *Vitis vinifera*.

**Fig. S8** Microsynteny plot of *Amborella trichopoda* (top), *Betula pendula* (middle), and *Vitis vinifera* (bottom) illustrates the genome evolution around lupeol 28-monooxygenase (CYP716) gene.

**Methods S1** DAPPI–MS.

**Methods S2** Triterpenes GC–MS.

**Methods S3** RNA-Seq data analysis.

**Table S1** Triterpene quantification in Betulaceae species by GC–MS and DAPPI.

**Table S2** The statistics of RNA-Seq data preprocessing and gene expression analysis of *B. pendula* cryosections F1–F8.

**Table S3** Enriched gene ontology (GO) categories among differentially expressed genes throughout stem fractions.

**Table S4** Enriched biochemical pathways (PlantCyc) among differentially expressed genes throughout stem fractions.

**Table S5** List of developing xylem and phloem genes involved in lignification.

**Table S6** Enrichment analysis of genes under selective sweeps, genes originating from whole-genome duplication events, tandem expansions, or birch-specific expansions.

**Table S7** Analysis of meristem-specific and periderm-specific differentially expressed genes.

**Table S8** Correlation analysis of AINTEGUMENTA and CYC3.1, as well as betulin and suberin biosynthesis genes with candidate TFs.

**Table S9** Gene expression analysis of *A. glutinosa* cryosections F2 and F3.

**Table S10** Overall statistics of ortholog analysis.

**Table S11** GO enrichment analysis of birch-specific expanded gene families expressed in *B. pendula* phellogen/phelloderm tissue (F2).

**Table S12** Identification and characterization of *cis*-motifs upstream of lupeol 28-monooxygenase and lupeol synthase genes in *B. pendula*, *A. glutinosa* and *V. vinifera*.

**Table S13** Atlas of all transcription factors in birch, selected periderm-specific transcription factors, and crossreference comparison.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



## About New Phytologist

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**